OTS: 60-31,780

JPRS: 3948

22 September 1960

FOREIGN DEVELOPMENTS IN INFORMATION PROCESSING AND MACHINE TRANSLATION

Distributed by:

OFFICE OF TECHNICAL SERVICES
U. S. DEPARTMENT OF COMMERCE
WASHINGTON 25, D. C.

U. S. JOINT PUBLICATIONS RESEARCH SERVICE
205 EAST 42nd STREET, SUITE 300
NEW YORK 17, N. Y.

F O R E W O R D

This publication was prepared under contract by the UNITED STATES JOINT PUBLICATIONS RESEARCH SERVICE, a federal government organization established to service the translation and research needs of the various government departments.

JPRS: 3948

CSO: 3901-D/14

FOREIGN DEVELOPMENTS IN INFORMATION PROCESSING AND MACHINE TRANSLATION

FOREWORD

This translation series presents information from foreign-language literature on developments in the following fields of language data processing: Machine translation studies; questions on structural linguistics, phonological theory, investigation of morphological models, development of syntactic structures and transform analysis; theory of language communication; logical and linguistic problems of constructing information machines and information languages; logical semantics; mathematical and applied linguistics; automatic programming; organization, storage and retrieval of information; automatic indexing and abstracting; character and pattern recognition; automatic speech input; documentation, etc. The series is published as an aid to U. S. Government research.

Previously issued reports on this subject include:

JPRS: 68, 241, 319, 355, 379, 387, 487, 621, 646, 662, 705, 729, 863, 893, 925, 991, 992, 1006, 1029, 1130, 1131, 1132, 1133, 3225, 3300, 3356, 3433, 3502, 3532, 3570, 3597, 3598, 3599, 3613, 3629, 3731, 3758, 3796 and 3797.

AUTOMATIC TRANSLATION BY THE "YAMATO" COMPUTER

- JAPAN -

[Following is a translation of a Japanese-language article by R. Tadenuma and J. Kita of the Electrical Testing Laboratory which was reprinted from Proceedings of the Symposium on Automation, October 1959. Place of origin or publication not given.]

1. Preface

Automatic translation of languages with the aid of an electronic computer is being attempted in several places. However, in nearly all of these cases, a universal electronic computer with adequate memory is being used. When the Electrical Testing Laboratory first planned automatic translation, such a universal electronic computer was not available in Japan, and therefore, we constructed a simple computer for translation purposes. This is the Yamato computer and by February of this year, we were able to translate fairly simple English sentences of about first-or second-year middle school grade level. The following is a description of the method of translation and programming.

2. General Description of the Method of Translation

From the standpoint of programming, in order that an electronic computer can solve a problem, first the solution must be given in terms of human words, and then this solution in terms of human words must be translated into terms of words of the computer. Therefore, the matter of whether the computer can solve a problem or not depends on whether we can first express the solution in human words and also on whether this can be translated into computer logic. These same problems exist in the automatic translation of a language, the first of these problems being a method of determining the order of sentence structure in the translation.

The method of translation adopted with the Yamato computer was to first take one sentence, strip this sentence down to its bare essential structure by removing all modifying words, determining the sentence pattern, translating the English sentence pattern into Japanese sentence pattern, and then finally, adding the modifiers.

This method can be generally described as follows:

(1) Words are read one at a time, the translation being recorded on a magnetic memory drum, until a sentence is completed.

(2) When one sentence has been translated and recorded word by word, we next inspect the sentence for idioms. If there are such, these word groups are transformed into idioms.

(3) In order to determine the structure of a sentence, the sentence is next stripped of all modifying words. For instance, if we have a form "adjective-noun," the adjective is removed.

(4) The stripped sentence is then compared with a slate of basic sentence patterns, and the type of sentence pattern is determined.

(5) Depending on the type of sentence pattern, the order of the words are changed in order to conform to the corresponding Japanese sentence pattern, at the same time, the "joshi" or auxiliary verbs being determined.

(6) The modifiers are then inserted into their proper locations in the sentence, and the entire sentence is obtained as a sequence of words.

(7) The translation is printed out sentence by sentence. Cases and tenses are taken care of at this point.

3. Code for Letters and Symbols

In order that computer automatic translation be possible, human words must be coded into words that the computer can understand. In automatic translation with the Yamato computer, the following letters and symbols were coded on 8-unit tape.

"katakana" letters	75	ka, ki, ku, ke, ko, ga, gi, gu, ge, go, etc.
English letters	52	Upper and lower cases.
Numbers	10	0 through 9
Symbols	28	.,?! etc.
Others	4	Space, line, back-space, deletion.
Total	169	

The following points were considered in determining the above method of coding:

(1) The upper case letters in the English alphabet are followed by the small letters.

(2) There is a simple relation between the upper and lower case letters.

(3) The "katakana" is a convenient 5-row alphabet.

(4) The regular "katakana" sounds and the sonant and semi-voiced "katakana" sounds are simply related.

(5) The integers are in a 1248 code.

4. Types of Dictionaries

In order to perform translation, a dictionary is required. The types of dictionaries used for automatic translation with the Yamato computer were as follows:

4.1 English Dictionary

Since our object was to translate English of about middle school grade level, the English words were selected from textbooks for the first to third years of middle school. There were altogether about 2,500 words, and since the Yamato could handle about 3000 words, there was adequate capacity. We first took about 900 words from first-year level textbooks.

In order to minimize the number of words to be used, only root forms were adopted with the exception of a few very common words which variational forms were also adopted.

With regard to the order of the words, we had at first planned on ordering words according to their frequency of use. However, from the standpoint of sentence pattern, the words were categorized by their part of speech, the order of the particular part of speech category depending on the frequency of use of words in the category. The order of words in each category were in a,b,c order. An example of our computer English dictionary is shown in Table 1.

Table 1

Part of Speech	Word
Pronoun	he, her, him, ...
Adjective, adverb	a, about, absent, ...
Numbers, exclamation	one, two, three, ...
Noun	accordion, afternoon, ...
Verb	am, be, being, ...
Mothers	Mr., Mrs., No., ...

4.2 Japanese Dictionary

The Japanese dictionary used had about the same number of words as the English, the type of words used was about middle school level, and special expressions or words were avoided. Since the translation would be typed out in "katakana" or in phonetic syllables without the use of any "kanji" or Chinese characters, the "kun"-pronunciation of Chinese characters was adopted as much as possible in order to avoid misinterpretation of "katakana" grouping.

The ordering of words in the Japanese dictionary was about the same as that for the English dictionary, but with some slight variations since there was apt to be more than one translation for a particular English word.

4.3 Construction of Machine Words

From the time an English sentence is read until the translated Japanese sentence is typed out, all operations are undertaken with machine or coded words. These machine words must necessarily include all of the necessary information concerning an English word.

A machine word was composed of 8 letters, each letter being represented by 7 bits, so that each word was constructed of 56 bits. The letters were divided into 4 groups of 2 letters each, each pair of letters being used in the following manner:

Machine Word: A B C D

A:

English part of speech	Ending	Disposition of ending

B:

Conjugation, yes or no	Conjunction	Japanese part of speech	Type of connection

C:

		Number of Japanese translation

D:

Idiom, yes or no	Other translation, yes or no	Number of other translation

The bits which are not identified above were used for various intermediate operations in the mechanical translation.

Classification of English parts of speech are as shown in Table 3.

Table 2

Verbs	Auxiliary, intransitive complete, intransitive incomplete, transitive complete, transitive incomplete, passive voice, "be" verb.
Nouns	Noun, possessive case.
Pronouns	Personal, demonstrative, interrogative, relative, "something".
Adjective	Numerical, gender, demonstrative
Adverb	Simple, demonstrative, interrogative, relative

Preposition	To, of, etc.
Conjunction	Coordinate, subordinate, "than".
Exclamation	--

The Japanese parts of speech were classified in general according to Japanese rules of grammar. However, in the actual programming, the words were classified as to conjugation yes-or-no, type of conjugation, and type of connection as shown in Table 3.

Table 3

Type of conjugation: Ge-dan, kami-ichi, shimo-ichi, ka-hen, adjective, adjective verb.
Type of connection: Ren-yo, shushi, rentai, "u"-type, chushi.

[Translator's note: See book on Japanese grammar for details concerning type of conjugation and connection.]

As an example, the machine word for the word "spring" ("tobihaneru" in Japanese) is as follows:

A: 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0

Regular verb	Root form	Type of ending

B: 1 0 0 0 1 1 0 0 0 0 0 0 0 0

Conjugation, yes Conjugation, no	Shimo-ichi-dan conjugation	Type of connection

C: 0 0 0

Number of Japanese translation

D: 0 1 0

Other translation, yes	Number of other translation

A second meaning for the same word "spring" ("haru") is coded as follows:

A: 0 0 1 1 1 1 0 0 0 0 0 0 0

Regular noun	Singular

B:	0	0	00000	00000000
	Conjugation, no	Connection, no	Regular noun	

C:	0 0 0

Number of Japanese translation

D:	0 0 0

4.4 Dictionary for Compound Words and Idioms

Idioms can be classified into two types, namely, a first type which consists of a series of consecutive words, and the second type also of a number of words but not in succession. Here, we have considered only the first type, and only those idioms which would be misinterpreted by a word-for-word translation. One example is "good morning," another type of example is "to look at" (transitive verb) which is derived from the intransitive verb "to look."

Since an idiom is constructed from a combination of the Japanese translations of English words, if there is more than one English word with the same Japanese translation, it is possible that an idiom may be constructed where none exists. In order to avoid the possibility of such a confusion, words which form idioms were purposely marked. Even this could have resulted in error; however, no such case occurred, no doubt due to our use of only a limited vocabulary.

4.5 Sentence Pattern

In order to translate English sentence pattern into Japanese pattern, transposition of words and the selection of the proper auxiliary verbs are necessary. For this purpose, about 50 forms of sentence patterns were prepared. Of course, there are many more forms; however, for simplicity in programming, only these were used. For example, the affirmative English sentence forms and corresponding Japanese sentence forms are as follows:

1. N V_{ci}	N ga V_{ci}
2. N V_{ii} N	N ga N_{ii}
3. N V_{ii} A_i	N ga A_i V_{ii}
4. N V_{ci} N	N ga N wo V_{ct}
5. N V_{it} N N	N ga N wo N V_{it}
6. N V_{it} N A_i	N ga N wo A_i V_{it}
7. N V_d N N	N ga N N wo V_d

The abbreviations used here mean:

N	:	Noun
A_i	:	Adjective
V_{ci}	:	Verb, intransitive, complete
V_{ii}	:	Verb, intransitive, incomplete
V_{ct}	:	Verb, transitive, complete
V_{it}	:	Verb, transitive, incomplete
V_d	:	Dative verb

4.6 Miscellaneous

In addition to the above routines, tables were also prepared for inflection, transformation of part of speech, declension, conjugation, use of auxiliary verbs, and other purposes. The routines for inflection alone contained about 15 types, numbering 50 in all.

The following is a table showing transformation of part of speech corresponding to different English word endings:

Table 4

Ending	Verb	Noun	Adjective	Adverb
ing	Verb, noun			
or	Noun			
ers	Noun, plural			
er's	Noun, possessive			
ers'	Noun, possessive, plural			
er	Noun		Adj. comparative	Adv. comparative
ed	Verb			
en	Verb			
es	Verb	Noun, plural		
es'		Noun, possessive		
's		Noun, possessive		
th			Number, ordinal	
ly			Adverb	
est			Adjective, superlative	Adverb, superlative
'		Noun, possessive		

5. Program

Reading of English

The operations followed from the reading of English words to the construction of the corresponding machine words were as follows:

When an English word is first read, the prepared English word dictionary is consulted. If the word is in the dictionary, the corresponding machine word is extracted from the machine word dictionary.

If the word is not in the dictionary, we check to see if the word starts with a capital letter. If so, a search is made for the same word in small letters.

If the word is already in small letters, but still not in the dictionary, the ending of the word is modified to give the basic or root form of the word and the dictionary is again consulted. If the word is found, the machine word is extracted, with a notation indicating the type of ending. If the part of speech also changes, this correction is also noted. For instance, if a verb ends with "er," the part of speech is changed to a noun.

If the word is still not available in the prepared dictionary, the word is considered as a noun, the machine word extracted, and the English word is typed out in its original form in the translation.

When the operations just described have been completed for a full sentence, we next scan the sentence for idioms.

At the end of a sentence, we generally have ",", "?", or ".". Since the period is also used as an abbreviation as well as the termination of a sentence, we entered words such as "Mr." and "No." at the end of the dictionary so that in consulting the dictionary, the location of the machine word would automatically indicate that these were abbreviated words and not sentence endings. Of course, this system had its disadvantage with respect to a negative answer such as "No," which could have been read as "number".

Determination of Idioms

Idioms are constructed by a combination of Part C of the word code. We first check Part D to see whether or not this word can form an idiom. If the answer is yes, we use Part C, and then proceed to an examination of the next word in the sentence. Since the longest idiom used was "how do you do", consisting of 4 words, we limited the sequence of Part C to four. This sequence of Part C of words is then compared with an idiom dictionary. If the idiom is not available, we drop the last "Part C", and try with a sequence of three. If this process is repeated but with no success, we

conclude that the first word does not form part of an idiom, and we therefore repeat the entire process of idiom determination, starting with the second word.

Auxiliary Verbs

With the determination of idioms in a sentence, we next turn to the auxiliary verbs. The auxiliary verbs are important even if they do not necessarily determine sentence structure, and they are determined by the following main verb. Following this reasoning, we were able to omit the auxiliary verb only in the case when it was followed by the root form of a verb. We also omitted the "be" verb when followed by a verb ending in "ing" and also the "be" and "have" verbs when followed by a past participle. These rules were memorized as a type of sentence form, and only used in the last stage of the operation when the Japanese translation was being typed out.

If we admit adverbs as the only part of speech coming between the main and the auxiliary verbs, the special rules governing disposition of the auxiliary verb apply in such a case only and no other.

Elimination of Modifiers

In order to simplify a sentence, it is often expedient to eliminate all but the last word of a particular word group. For instance, in the case "article-adjective-noun," this word group in a sentence can be identified with the noun. The other words in the group can be associated with this noun and stored in the computer's memory. For ease of handling, groups were limited to not more than four words, larger word groups being treated by repetition of the procedure.

Word groups which were considered were the following two: Demonstrative adjective-simple adverb-adjective-noun. Demonstrative adjective-adjective-adjective-noun.

Following the above procedure, the sentence was then scanned for the preposition "of" and all adjective phrases of the type "noun-of-noun" were eliminated.

The sentence was next scanned for other prepositions, particularly those forming adverb phrases. These phrases were retained in the sentence as adverbs at this point, but later to be excluded in the determination of sentence pattern.

Dictionary of Sentence Patterns

We have now reached the point where it becomes possible to compare a given sentence with our dictionary of fundamental sentence patterns. Since a fundamental sentence pattern consists

basically of nouns, adjectives, and verbs, the sentence is stripped of all other words. Here, pronouns should be retained as nouns.

In comparing a given sentence with our dictionary, if a corresponding sentence pattern is identifiable, we must check to see whether or not the given verb has another meaning. If the verb has another meaning, but its function is no longer a verb, we must then search for another meaning for the noun or nouns in the sentence, and a corresponding identifiable sentence pattern. If no sentence pattern can be determined, we must then conclude that this is not a sentence, and resort to direct word-for-word translation.

Re-Introduction of Modifiers

All modifiers which had previously been eliminated from the sentence but which had been identified with a word in the stripped sentence can now be re-introduced into the transformed sentence pattern.

We start by taking the first machine word and seeing whether this had any modifiers. If it did, we substitute a word group, retrieved from the computor memory, for this machine word. If the word had no modifiers, we proceed to the next machine word. This procedure is followed to completion of a sentence, and then repeated for the same sentence in order to extract from the memory modifiers of words in a word group. This repetition of procedure ends with exhaustion of all modifiers available to the sentence.

It should be remembered that when modifiers are re-introduced into the sentence, that the order of words should conform to Japanese grammar or manner of speech. For instance, an adverb phrase "preposition-noun" becomes "noun-preposition" in Japanese.

Typing Out Japanese Translation

Since all of the required word translation has now been accomplished, we can type out the Japanese translation word-by-word from the beginning of a sentence. First, however, we must inspect the machine words to see whether or not word endings have been changed. For certain word endings, the appropriate Japanese word or words is added. For example, verbs ending in "er" call for the addition of "mono" in Japanese, meaning "person who".

Examples of English word endings and corresponding Japanese word additions are as follows:

Word ending	Before translated word	After translated word
ing		tsutsu aru
or		mono
ers		mono
er's		monono
er		mono
	motto	
is		no
th		bamme
est	mottomo	

6. Conclusion

Automatic translation from English to Japanese using the Yamato model computer designed at the Electrical Testing Laboratory was summarized in this report. The number of different commands was 46, the total programming being for about 2000 words. Although relatives and participial construction had not been considered sofar, we believe that these problems can be adequately handled since the Yamato computer can be programmed to store 4096 words.

Although we have made considerable progress towards a more complete translation program, there are still many problems to be solved before a fluent Japanese translation becomes possible. In connection with our program which has to do with translating current English into current Japanese, we have also entertained the thought that papers could be written in a language adaptable to translation by the Yamato computer. For instance, if the past tense is made to always end with "ed" and the past participle with "en," it would greatly simplify automatic translation even though the language differed from that in common use.