

An Experimental System of Machine Translation

by the Programmed Sentence Scan Method

Syuzo ASAKURA* and Shinpei NAKANO*

(Received September 25, 1974)

This paper reports on machine translation by the syntactical method, but a noun is subclassified on the basis of considering remarkable syntactical function of the noun or semantics.

To improve the quality and speed of translation, we propose the processing method of machine translation by the programmed sentence scan method with classifying a part of speech into the stratified subclasses, without increase of the kind of a part of speech.

1. Introduction

The problems on machine translation have been discussed on the determination of one part of speech in multi-parts of speech¹⁾, the method of storing and searching the index of machine dictionary, and the analysis of sentence based on syntax and/or semantics. The theory and experimental results which are based on Phrase Structure Grammar have been also published^{2), 3)}.

There are many papers about the syntactical methods which process a part of speech pattern by the table search^{2), 3)}, but there are a few reports on the result of analyzing the pattern by the programmed sentence scan method⁴⁾.

To improve the quality of translation, we propose, in this paper, the method of machine translation by the programmed sentence scan method with classifying a part of speech into the stratified subclasses without the increase of the number of a part of speech. The experimental results and the merits of this method are reported.

This paper is the development of a master's thesis⁵⁾ of one of the writers.

2. Formal Relation between English and Japanese

Analyzing English sentence with P. S. G., the order of Japanese versions of analyzed English words and adding inflection of Japanese and *joshi* in case of need must be considered.

An example sentence is given,

A cat catches a rat.

This sentence is rewritten into the following form with P. S. G..

$S \longrightarrow NP_1 + VP$. (1)

$NP_1 \longrightarrow Det. + N_1$. (2)

$VP \longrightarrow V + NP_2$. (3)

* Department of Electronic Engineering

$NP_2 \longrightarrow \text{Det.} + N_2 . \quad (4)$

where, S: Sentence, NP: Noun Phrase, VP: Verb Phrase, Det.: Determinant, N: Noun, V: Verb.

The symbol $\longrightarrow$ means production or rewriting.

The symbol $+$ means concatenation.

This sentence is translated into Japanese in the following order.

$NP_1 \text{ } wa \text{ } VP . \quad (1')$

$\text{Det.} + N_1 . \quad (2')$

$NP_2 + V . \quad (3')$

$\text{Det.} + N_2 . \quad (4')$

There is an inversion between English sequence and Japanese sequence at (3'). But at the other sequences there are no inversions. As English sequence and Japanese sequence are the same, Japanese versions are put according to English sequence adding inflection of Japanese and *joshi* in case of need.

3. Classification of a part of speech

In machine translation, the classification of a part of speech is very important, because speed of pattern processing and the method depend on it.

The classifications based on almost common use or considering the classification by C. C. Fries or P. Roberts and Japanese²⁾ have been published.

In this paper, the classification is almost based on common use. But a word which is used as another part of speech besides its part of speech is joined to multi-parts of speech word. A part of speech is stratified into subclasses based on a part of speech or function of Japanese version, remarkable function in English, or semantics. The representation of a symbol of a part of speech in computer is shown in Fig. 1. One word* is allocated to one symbol of one part of speech. In a word, at 0-7 bit a part of speech, at 8-11 bit the subclass, at 12-13 bit inflection of Japanese version and *joshi*, and at 14-15 bit the space where the result will be written through the Inflection Processing Routine are represented.

3.1 Noun

Subclass

- | | |
|--------------------------|--------------------------------|
| (1) pronoun | (2) common noun |
| (3) proper noun | (4) collective noun |
| (5) material noun | (6) abstract noun |
| (7) numeral noun | (8) season or time noun |

A noun is subclassified on the basis of the remarkable function in English or semantics. A few nouns are classified into other part of speech.

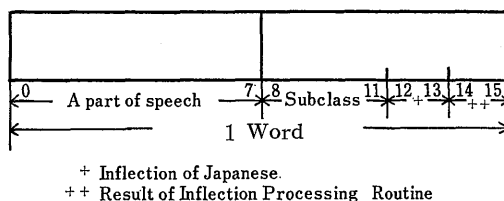


Fig. 1 Bit pattern of the memory for the part of speech

* FACOM 270-30, word-machine, 1 word=16 bits.

Two concrete examples are given.

i) pronoun

Possessive case of a pronoun is used as an adjective, so it is classified into adjective. A demonstrative pronoun is often used as a demonstrative adjective, so it is classified into adjective. A demonstrative pronoun is often used as a demonstrative adjective, so it is classified into both demonstrative pronoun and demonstrative adjective.

ii) adverb use of noun

A noun which is possible to be used as adverb use (e.g. today, tomorrow, etc.) is classified into both noun and adverb.

3.2 Adjective

Subclass

(1) adjective 1

An adjective of which Japanese version has no inflection is subclassified into adjective 1.

(2) adjective 2

An adjective of which Japanese version is an adjective is subclassified into adjective 2.

(3) adjective 3

An adjective of which Japanese version is *keiyodoshi* is subclassified into adjective 3.

(4) adjective 4

An adjective which precedes an article and of which Japanese version has no inflection is subclassified into adjective 4.

3.4 Article

The function of an article and the function of an adjective are the same, but an article is not modified by an adverb. An article is treated independently. For example, in the following pattern such as <verb>+<adverb>+<article>+<noun>, <adverb> doesn't modify <article>. But if <article> is replaced with <adjective>, it is difficult to determine which a part of speech <adverb> modifies, <verb> or <adjective>.

3.5 Not

A not needs conjugation of *mizenkei* in Japanese version.

3.6 Adverb

Subclass

(1) adverb 1

An adverb which is possible to be used as an adjective use of adverb is subclassified into adverb 1. Adjective use of adverb is to modify noun.

(2) adverb 2

(except adverb 1)

An adverb modifies ordinarily a verb, an adjective, an adverb or sometimes a noun. Adverb 1 is ordinarily used as an adverb and the number of it is a few, so adverb 1 is not joined to multi-parts of speech word but subclassified. An adverb often modifies the whole sentence. This adverb is independent of a part of speech pattern of sen-

tence, so it is processed with no relation to sentence pattern.

3.6 Conjunction

A coordinate conjunction and a subordinate conjunction are not subclassified in this experiment.

3.7 Preposition

Subclass

(1) preposition 1

(except "to")

(2) preposition 2

(only "to")

A preposition "to" is often used as a regular preposition, but it has more important function to make *to-infinitive*.

3.8 Verb

Subclass

(1) be verb

(2) have verb

(3) transitive verb

(4) intransitive verb

(5) dative verb

(6) incomplete transitive verb

A verb must be strictly subclassified because, the phrase following it is determined as an object, a complement, an indirect object, or an objective complement phrase, and the inflection of Japanese version and *joshi* are determined.

3.9 Auxiliary verb

Subclass

(1) auxiliary verb 1

An auxiliary verb of which Japanese version has a conjugation is subclassified into auxiliary verb 1.

(2) auxiliary verb 2

An auxiliary verb of which Japanese version has no conjugation is subclassified into auxiliary verb 2.

3.10 Relative

Subclass

(1) relative pronoun

(2) relative adjective

(3) relative adverb

A relative pronoun is possible to be used as a relative adjective.

3.11 Interrogative

Subclass

(1) interrogative pronoun

(2) interrogative adjective

(3) interrogative adverb

An interrogative pronoun is possible to be used as an interrogative adjective.

3.12 Ing

This part of speech is ordinarily made through the Inflection Processing Routine.

An Ing-from is used as either a present participle or a gerund. Their function is very important, so this part of speech is treated independently.

4. Machine Dictionary

As the classification of a part of speech has been decided, the concrete specification of dictionary can be made. The specification of machine dictionary includes not only the lemma and its content but also the searching method of it.

There are reports of the method in which a word being separated and check-summed for the index of machine dictionary and of the searching method³⁾ of table by the longest matching.

In this paper, the method that directly separates the stem and the ending of a word by searching the dictionary constructed with tree structure is applied. The machine dictionary is divided in the same manner of ordinary dictionary on the first letter of a word. By the restriction reflected from the construction of computer system*, each part of dictionary divided on the first letter is independently programmed and stored in the drum memory. An input sentence is divided into words through the Dividing into Words Routine. After a word is classified according to its first letter, necessary part of the dictionary is loaded on the core memory and is now ready available. The searching method conforms to the well-known method for tree structure. The machine dictionary consists of the part of searching, and the part of lemma and its content. The former part is programmed and the latter part is a part to be ordinarily called a dictionary. The latter part consists of the lemma that is the label which indicates the address, the root, the symbol of a part of speech, and Japanese version.

5. Analysis of a part of speech pattern

An English sentence that is read through card is transformed into a string of the symbols of a part of speech by consulting the dictionary. The string is analyzed by the syntax rules and this analysis decides on the order of Japanese versions and inflection and *joshi* if necessary.

Immediate Constitute method⁶⁾ is often used for analysis of sentence. In this paper, this method is almost applied. But introducing Predictive Analysis⁶⁾ for a part of speech pattern without scanning the invalid pattern, speed of pattern processing is improved.

In case of the multi-parts of speech word, a part of speech in multi-parts of speech is selected according to the order written in the dictionary, and the pattern is analyzed. If the pattern is found to be impossible through the analysis, the next part of speech is selected and the new pattern is analyzed.

5.1 Syntax Rule

A part of speech pattern is analyzed by means of the programs that perform the Syntax Rule 1: the set of Table 1. If a concatenation of some symbols of a part of speech listed in the right is found in the pattern, it is rewritten with a part of speech in the left. After the pattern is analyzed by these rules, declarative sentence is reduced to the Sentence Pattern 1 (shown in Table 2). These pattern of declarative sentence

* FACOM 270-30, core memory 32KW, drum memory 256KW, Computer Center, Chubu Institute of Technology.

is constructed with no interrogative. As the pattern applied to the first case shown in the Interrogative Sentence Pattern 2 (shown in Table 7), masking the subject of the pattern, the Syntax Rule 3 is applied. As the pattern applied to the last case is shown in the Interrogative Sentence Pattern 3 (shown in Table 8), the Syntax Rule 5 (shown in Table 9) and the Syntax Rule 3 are applied in this order.

Table 4 Syntax Rule 3

<noun> ::=<verb> <noun>

Table 5 Interrogative Sentence Pattern 1

<auxiliary verb> <declarative sentence> <?>

<interrogative> <auxiliary verb> <declarative sentence> <?>

Table 6 Syntax Rule 4

<interrogative>::=<interrogative pronoun> | <interrogative adverb> |
 <interrogative adjective> <noun>

<noun> <verb>::=<auxiliary verb> <noun> <verb> |
 <interrogative> <auxiliary verb> <noun> <verb>

Table 7 Interrogative Sentence Pattern 2

<be verb> <noun> <noun> <?>

<be verb> <noun> <adjective> <?>

Table 8 Interrogative Sentence Pattern 3

<interrogative> <be verb> <noun> <?>

Table 9 Syntax Rule 5

<be verb> ::=<interrogative> <be verb>

6. Processing Program

Two methods are ordinarily applied to pattern processing. One is that pattern is analyzed with table of grammar by keeping the analysis form of pattern uniform. The other is that pattern is analyzed with programs that describe the syntax rules.

In this paper, the latter method is applied. The form of program is independent of each other for every part of speech.

List processing program is applied to indicate the order of Japanese versions and the information of inflection of Japanese and *joshi* if necessary.

6.1 Conjunction Processing Routine (CONJ)

Processing a concatenation of three symbols of a part of speech. Scanning from the head to the tail of sentence.

(1) Conjunction Processing Routine 1 (CONJ 1)

Processing the conjunction of adverbs and adjectives by a conjunction.

(2) Conjunction Processing Routine 2 (CONJ 2)

Processing the conjunction of the same subclass verbs by a conjunction.

(3) Conjunction Processing Routine 3 (CONJ 3)

Processing the conjunction of the same subclass nouns by a conjunction.

(4) Conjunction Processing Routine 4 (CONJ 4)

Processing the conjunction of verbs by a conjunction.

(5) Conjunction Processing Routine 5 (CONJ 5)

Processing the conjunction of nouns by a conjunction.

6.2 Preposition Processing Routine (PREP)

Processing a concatenation of two symbols of a part of speech.

Scanning from the tail to the head of sentence.

6.3 Adverb Processing Routine (ADV)

Processing a concatenation of two symbols of a part of speech.

Scanning from the head to the tail of sentence.

6.4 Article Processing Routine (ART)

Processing a concatenation of two or three symbols of a part speech.

Scanning from the head to the tail of sentence.

6.5 Adjective Processing Routine (ADJ)

Processing a concatenation of two symbols of a part of speech.

Scanning from the head to the tail of sentence.

6.6 Verb Processing Routine (VERB)

Processing a concatenation of two symbols of a part of speech.

Scanning from the head to the tail of sentence.

6.7 Auxiliary verb Processing Routine (AUX)

Processing a concatenation of two or three symbols of a part of speech.

Scanning from the head to the tail of sentence.

(1) Auxiliary verb Processing Routine 1 (AUX 1)

Processing an auxiliary verb in declarative sentence.

(2) Auxiliary verb Processing Routine 2 (AUX 2)

Processing an auxiliary verb in interrogative sentence.

6.8 Interrogative Processing Routine (QUES)

Processing a concatenation of two or three symbols of a part of speech.

Scanning from the head to the tail of sentence.

6.9 Sentence Pattern Processing Routine (SENT)

Processing sentence pattern.

Scanning from the head to the tail of sentence.

(1) Sentence Pattern Processing Routine 1 (SENT 1)

Determination of sentence pattern.

(2) Sentence Pattern Processing Routine 2 (SENT 2)

Deleting except the subject.

(3) Sentence Pattern Processing Routine 3 (SENT 3)

Deleting the subject.

(4) Sentence Pattern Processing Routine 4 (SENT 4)

Processing transformational sentence pattern.

(5) Sentence Pattern Processing Routine 5 (SENT 5)

Processing be-verb interrogative sentence pattern.

6.10 List Processing program

One list element is allocated to each word of sentence. When a pattern is analyzed by one syntax rule, the order of Japanese versions is represented in list structure. Inflection of Japanese and *joshi* also represented in the data of list structure.

7. Experiment

Experimental system has been constructed. The general flow chart is shown in Fig. 2.

7.1 Reading Sentence

An English sentence which is punched in punch card is inputted.

7.2 Dividing Sentence into Words**Routine**

The sentence is divided into words through this routine. A word is the string of letters between a space and next space.

7.3 Inflection Processing Routine

Let the letter array $W_0(N)$ denote the stem which is written in the dictionary, and the letter array $W(N)$ denotes the used word. The letter of the array $W_0(N)$ and that of $W(N)$ are compared from head to tail every one letter. When all letters are coincide, there is no inflection in the used word. At N equal to N_1 , the letters being not coincide, if the letter $W_0(N)$ is a space, the used word has an ending, and if it is not a space, the used word has an inflection.

7.4 A part of speech Pattern and Sentence Pattern Processing

An arrangement of a part of speech Pattern Processing Routine and Sentence Pattern Processing Routine is shown in Fig. 3. The order is arranged by the consideration on the following points.

In the early processing steps, it is thought that a conjunction can connect the words when there is some relation concerning semantics or function.

The order of the routines of an adverb, an article, and an adjective is thought to be able to process the following pattern.

(article)+(adverb)+(adjective)+noun,

where the symbol () means that the element in this symbol is option.

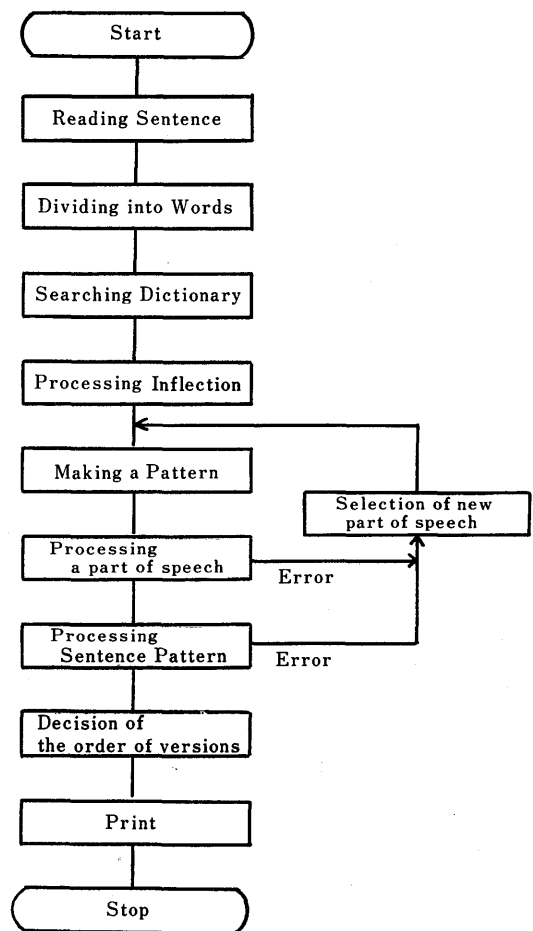


Fig. 2 General flow of the program

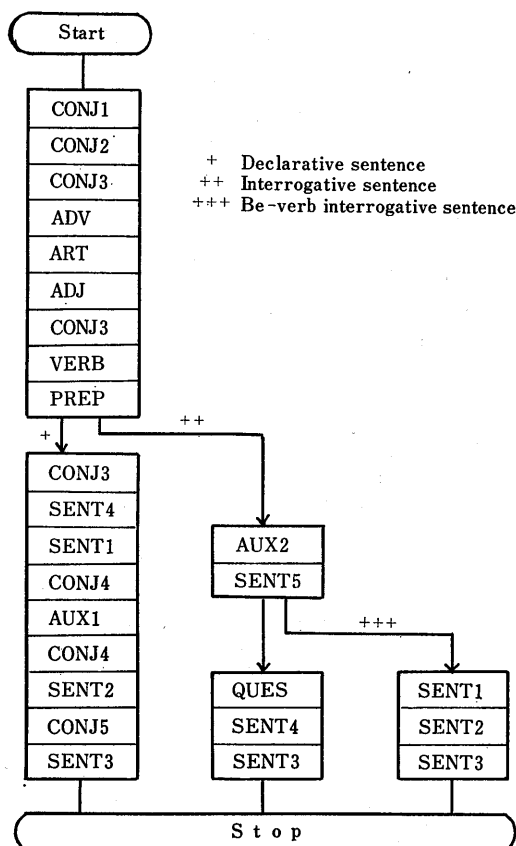


Fig. 3 Flow of the syntax analyzer

by following the list. When the Japanese version is outputted, the data in the list is referred and the inflection of Japanese and *joshi* are outputted if necessary.

8. Result of Experiment

The input sentence and the corresponding output sentence are shown below.

- (1) WE SWIM IN SUMMER.
WAREWAREWA NATU NOUTINI OYOGU
- (2) YOU AND I ARE STUDENTS.
ANATA SOSHITE WATASHIWA GAKUSEIDE ARU
- (3) I HAVE A BALL.
WATASIWA *BALL O MOTU
- (4) HE GIVES ME AN APPLE.
KAREWA WATASINI RINGOO ATAERU
- (5) WHAT IS THIS?
KOREGA NANIDE ARU KA
- (6) WHAT BOOK IS THIS?
KOREGA NANINO HONDE ARU KA
- (7) WHAT CAN YOU DO?
ANATAWA NANIO SURU KOTOGADEKIRU KA

After modifiers are processed, nouns which have been located separately by the modifier may be found connectable with a conjunction. Because the syntax rule of a conjunction says to connect the front symbol of a part of speech and the following symbol.

In the late steps, a conjunction ordinarily connects phrases or clauses, and then the relation concerning semantics or the function is not thought very important.

When the part of speech pattern is correct as sentence pattern, all symbols of a part of speech in the sentence pattern are deleted. If there are some symbols in the pattern, the selection of one part of speech in multi-parts of speech is not correct. The next part of speech is selected and the new pattern is analyzed

7.5 Print of Japanese version

The Japanese sentence is outputted from the line printer in *Romaji* (Roman letter).

The order of Japanese versions is decided

(8) THERE IS A PEN ON THE DESK.

TUKUE NOUENO *PENGAKOKONI ARU

(9) IS THERE AN APPLE?

RINGOGAKOKONI ARUKA

(10) YES, I AM A BOY.

HAI WATASIWA SYOUNENDE ARU

The Symbol * means that the version is in English.

9. Conclusion

This system is constructed as a very simple and handy system, so the exclamatory sentence and a relative are not dealt with.

The merit of this system is the quickness of pattern processing. Even if the subclasses are increased to improve the quality of translation, the quickness of this method is not so influenced. Furthermore, in the future, the number of the multi-parts of speech words increases in proportion to the complement of the dictionary, and the number of a part of speech pattern which is made by these multi-parts of speech words increases explosively. In this case, the quickness of pattern processing is very effective.

In the subclassification of a word of manifold meanings, the following example is obtained. When the sentence which contains the word "spring" is translated into Japanese, two Japanese sentences are made. In one sentence, the Japanese version of "spring" is "BANE". In the other sentence, it is "HARU". But, in the English sentence, when the word "spring" is replaced with the words "spring and summer", only one Japanese sentence is made. The Japanese version of "spring" is "HARU" because of being restricted by the word "summer" (season or time noun). The more systematic subclassification can realize high quality translation without increase of the number of a part of speech patterns, so the more systematic subclassification shall be studied.

The order of pattern processing routine is valid for the present. By the result of experiment, the order of arrangement may be changed and the new processing routine may be also inserted.

Acknowledgements

We would like to thank Professor H. Aiiso of Keiogijuku University for continual guidance. We are also grateful to Mrs. J. Igarashi, Mr. S. Uemura and Mr. Y. Sakamoto, the staffs at Electrotechnical Laboratory, for their suggestions concerning the theoretical aspects and for their offers of research materials. In addition, we wish to express our gratitude to Dr. H. Nishimura of Electrotechnical Laboratory who criticized the construction of the syntax analyzer, the ability of it and our programming technique, and also offered research materials. Finally, we are indebted to Dr. K. Suzuki for many discussions about this work.

References

- 1) T. Kiyono, S. Mitsumori, and M. Miyamoto: "Machine Translation", Journal of the Institute of Electrical Communication Engineers of Japan, Vol. 46, No. 11, Nov. 1963.
- 2) T. Sakai and S. Sugita: "Mechanical Translation of English into Japanese", Journal of the Institute of Electrical Communication Engineers of Japan, Vol. 49, No. 2, Feb. 1966.
- 3) H. Nishimura: "The Development of a Machine Translation System", Researches of the Electrotechnical Laboratory, No. 705, March 1970.
- 4) R. Tadenuma: "English-Japanese Machine Translation I", Researches of the Electrotechnical Laboratory, No. 624, Dec. 1961.
- 5) S. Asakura: "Mechanical Translation", Master Thesis, Chubu Institute of Technology, Feb. 1973. (unpublished)
- 6) T. Tamati: "7.2-Mechanical Translation of Language, (Ⅱ)", Journal of the Institute of Electrical Communication Engineers of Japan, Vol. 46, No. 11, Nov. 1963.