

## Mechanical Syntactic Analysis

SAE YAMADA\*

This proposed system of mechanical syntactic analysis of Japanese is performed at the same time as the dictionary oriented segmentation of the sentence. Because of the non-existence of word boundary blanks in Japanese, one must distinguish each lexemic unit in the sentence.

Almost all necessary information, needed for this segmentation, is also necessary for the syntactic analysis. So, the simultaneous execution of these two operations is economically convenient.

A sentence is scanned from left to right. As the lexemic units composing the sentence are found, the analysis advanced in this direction.

In this summary, the principal idea and procedure of proposed syntactic analysis will be discussed in particular.

Naturally, the feature of syntactic analysis and syntactic description is dependent upon the grammatical theory that is chosen. The grammar rules utilized in this method are phrase structure rules. But they are not context free in the strict sense. A non-terminal rule in this grammar has the form  $X + Y \rightarrow Z/i$  ( $1 < i < 5$ ).  $i$  indicates the strength of connection between two symbols  $X$  and  $Y$ , where  $X$ ,  $Y$  and  $Z$  are the names of non terminal symbols. This rule means that two symbols on the left hand are rewritten by the symbol of the right hand, if and only if there is no another symbol before  $X$  or after  $Y$  with which  $X$  or  $Y$  can be connected with the superior strength. Thus the relationship between the symbols is considered in the rules of this type. But the relationship doesn't mean that of the dependency type. We do not regard the direction of dependency.

The notion of "subject" in Japanese grammar is adopted from the grammatical notion in European languages. In Japanese a nominative noun is only an optional complement of the predicate verb as nouns in accusative, genitive, dative etc. The former doesn't take priority of the other. So we can describe, conventionally the complements-predicate verb relation in the sentence as a progressive construction. On the contrary, the structures of noun modification or of verb composition take the tree form in which the left recursive character is dominant. A verb phrase can come directly before noun and modify the noun without any subordinate conjunction. And we can repeat this structure several times. It is evident that the scanning from left to right is more convenient than the inverse direction. The

---

This paper first appeared in Japanese in *Joho Shori* (the Journal of the Information Processing Society of Japan), Vol. 5, No. 3 (1964), pp. 134-141.

\* Electrotechnical Laboratory, Agency of Industrial Science and Technology, Ministry of International Trade and Industry, Tokyo.

operation of sentence analysis proceeds in the same manner as the calculation of a mathematical expression which has several operators (i.e.  $+$ ,  $-$ ,  $\times$ ,  $\div$ , etc.). The push down stores are used efficiently for this purpose. In this case, the strength of connection between two symbols is considered as a mathematical symbol. We show an example of a sentence which assigns the degrees of connection strength between the symbols.

koko<sup>5</sup> ni<sup>3</sup> motii<sup>5</sup> te<sup>5</sup> i<sup>5</sup> ru<sup>1</sup> kaihou<sup>5</sup> ha<sup>2</sup> iwayuru<sup>4</sup> bakuhatu<sup>5</sup> hou<sup>3</sup> dca<sup>3</sup> ru<sup>1</sup>.

At present we distinguish 1~5 values of strength and use the special symbol \$. If the strength of connection is equal to 5, two symbols are concatenated and rewritten to another symbol immediately. If the strength of connection is equal to 2, 3 or 4, the concatenation of the two symbols is reserved until a mark of lesser degree appears. The sign \$ indicates a special case. Where we find \$ sign, all possible concatenations with sign 3 or 4 should be executed. It means that in this case the \$ sign takes a value weaker than 3.

After all concatenation before it are done, it takes the same value as 4. In the above example, the application of all grammar rules before that \$ sign gives us the name of the verb phrase "u.c. ft" (transitive verb phrase in terminal form) which is assigned to "kokonimotiiirareteiru". After that the \$ sign takes the value of 4.

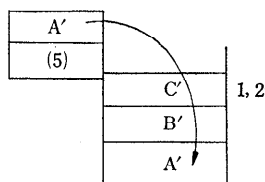
It means that the "u.c. ft" is indicated as the attribute of the noun "kaihou". The sign 1 indicate the weakest degree of connection. We find this degree sign only between a verb phrase and the period. As this degree is weakest, the grammar rule which has this sign is applied only at the end of the analysis operation and gives the symbol "S" (sentence).

For the analysis operation, the push down stores are used.

Suppose we find two successive lexemic units  $A$  and  $B$  in the beginning of the sentence. Let their function names be  $A'$  and  $B'$ . The first symbol  $A'$  is put into a pool of the push down store and the second symbol  $B'$  is put in the first entrance of this pool. Now, the grammar rule entered under " $A'+B'$ " is looked for in the grammar rule table. If we find it, we put the degree sign of the strength between  $A'$  and  $B'$  in the second entrance of the pool. Let the rule be  $A'+B'\rightarrow C'/5$ . In this case we transfer  $B'$  and  $C'$  into the pool. If the sign of strength is not equal to 5, we send only  $B'$  into the pool and we reserve the strength degree sign, until a sign of a weaker degree enters into the second entrance.

When the sign which enters the second entrance is that of a weaker strength

$A'+B'\rightarrow C'/5$



$A'+B'\rightarrow C'/4$

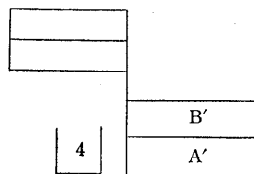


Fig. 1.

than the reserved one, two symbols in the pool will be concatenated and the new function name will be added. The history of concatenation is indicated by a pair of numerals that are added to the new function names. (C' in the above example)

By this operation of analysis, we can assign to the sentence of above example the following structure.

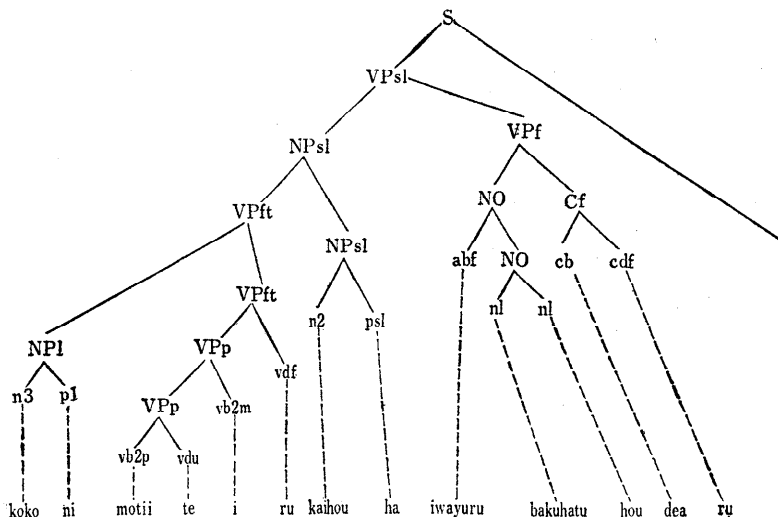


Fig. 2.

We have tried to find an efficient procedure for the application of phrase structure grammar rules for Japanese. For this purpose the form of the phrase structure rules in the strict sense has been somewhat modified and the concept of the strength of the connection between symbols has been introduced into the rules. This concept of strength will be useful in the semantic analysis.

### References

- [1] KUNO, S. AND A. G. OETTINGER, *Multiple-path Syntactic Analyser*, NSF-8, 1963, Harvard University.
- [2] VEILLON, G., *Consultation d'un Dictionnaire et Analyse syntaxique en Traduction Automatique*. la Faculté des Sciences à Grenoble, 1962.
- [3] MIKAMI, A., *Gendai Gohoo Josetzu*. Tokyo, 1959.