22

1965 International Conference on Computational Linguistics

SENTENCE GENERATION BY SEMANTIC CONCORDANCE

Toshiyuki Sakai and Makoto Nagao

Department of Electrical Engineering, Kyoto University

Kyoto, Japan

ABSTRACT

Generation of English sentence is realized in the following three steps. First, the generation of kernel sentence by phrase structure rules; second, the application of transformational rules to the kernel sentence; and finally the completion of a sentence by the morphophonemic modifications.

At the first stage of generating kernel sentence, the semantics of words are fully utilized. The method is such that a pair of words in the generation process (subject noun and predicate verb, verb and object or complement, adjective and modified noun etc.) is selected in accordance with the semantic categories which are attached to each word in the word dictionary. The semantic categories are determined by considering both the meaning of words themselves and also the functioning of words in sentences.

At the stage of transformational rules, sentence is considered not as a simple string but as the one having the internal tree structure, and the transformational rules are applied to this tree structure. For these two stages the generation process is formalized strictly and is realized in a computer programming. We have presented in relation to the transformational rules a method of sentence generation not from the axiom (from the top of the tree) but from any point, from which the whole tree is constructed.

We have also proposed that the morphophonemic rules can be presented as a kind of operators operating on words in the neighbourhood of a generated string.

1. INTRODUCTION

At present on behalf of the experience obtained from the research already done, we have fairly exact knowledge about the real difficulties in the MT research which are to be solved in the near future. Among these problems the most important ones might be how to construct the syntax of a language and how to grasp the semantics of sentences of the language.

There have been many excellent contributions to the problems of syntax, but there are still few to the problems of semantics and the interrelationship between syntax and semantics. We have tried an investigation in this area by the method of generation of English sentence.

The first paper ever published concerning the generation of sentences might have been that by Prof. V.H. Yngve of MIT in 1961. We have adopted the method once again from the following points of view.

(i) it acts as a powerful test to the study of sentence structure,
(ii) " to the study of semantics,
(iii) " to the study of the relationship between syntax and semantics in natural language.

Generally speaking the sentence generation method, contrary to the analysis of a given sentence (which is guaranteed to be a correct one), tends to demand the severe construction of syntactic rules and word selection rules. It may seem at the present level of mechanical translation that the treatment of the sentence property in its entirety is too difficult to realize. But if we hope to have the translation as perfect as possible, we are necessarily to confront with this problem.

The quality of a linguistic theory of a language may be best examined by the generation of sentences according to the specified linguistic theory. Especially the effect of the interrelationship between syntax and semantics seems to be clearly exemplified by this so to speak "crude" test. Thus the sentence generation method surely responds to this overall treatment of the sentence property. This may be considered as a step towards the general theory of natural language.

2. EVALUATION OF THE GENERATED SENTENCE

2.1 Evaluation of the generated sentence -syntax-

Several methods are developed for the description of sentence structure. We represent the syntax of English by a phrase structure grammar, transformational grammar, and mophophonemic rules. The kernel sentence is generated by the phrase structure grammar, then some proper transformational rules are applied to it, and then the modification of the sentence by the morphophonemic rules produces the final output.

These rules should generate "conceivable sentence structures", although the actually used sentence structures have several constraints. These are for example,

(i) the depth of the sentence structure
(ii) the coordination structure
(iii) the intrinsic unsymmetry of sentence structure --- progressive structure, top heavy structure etc.

In general the rules which are suitable for analysis of a given sentence seem to differ from the rules which generate good sentences. The difference between these two is the difference between the actual spoken sentences and the conceivable sentences. Here we can see man's tendency to the language structure. Therefore it will not be worthless to know the frequency ratios of the phrases used in the actual sentences.

2.2 Evaluation of the generated sentence -semantics-

The next question, and the more difficult one than the former, is the determination of what is the proper meaningful sentence. The test for the semantic sentence anomaly is far more difficult than the test for the grammaticality of the sentence. Here we can think of the following three levels of criteria for the right sentence.

(i) The grammatical sentence which is spoken or written by the average person (and the sentence which conveys a concrete concept without knowing the circumstances the sentence is spoken). Here "grammatical" covers the phonology, phonemics, morphology, syntax etc. These sentence which are grammatical but which are contradictory in meaning and which we do not speak are to be rejected.
(ii) The sentences which are incomplete in the word usages, inflexions and so on but which convey clear understandable concepts. These are the so-called corrigible sentences.
(iii) The sentences which are grammatical but carry no concrete meaning

if they are not supplemented by tediously long explanations about the righteousness of the expression.

We have here adopted the second criterion for our generation of English sentences. That is because we can transform the corrigible sentences into the complete ones comparatively easily by checking the concordance of gender, number, case etc. Hereafter we are mainly concerned with the sentence which carries very definite concept, that is to say, the sentence of complete semantic consistency.

3. SENTENCE GENERATION BY SEMANTIC CONCORDANCE

We concentrate our efforts on the generation of the affirmative active declarative sentences. We try to generate this kernel sentence by the expansion rules. By the generation of the kernel sentence the attention is on the structural balance of the whole sentence, the influence of the choice of a word to the other part of the phrase, and their relation to the unified concept of the sentence. An expansion rule has a main constituent and the other non-main constituents in the expanded part. The latter symbols may contain optional elements.

When an expansion rule is applied to a non-terminal symbol, to which there is already given a concrete word, the word is assigned to the main constituent of the expanded part. The words to the non-main constituent symbols are selected in relation to the main constituent word. A verb or a noun is taken as the main constituent of the non-terminal symbol "sentence" (initial symbol).

3.1 Terminology

A set of syntactic word classes (abbr. SWC)

$$S = (s_1, s_2, \ldots\ldots)$$

A set of words

$$W = (w_1, w_2, \ldots\ldots)$$

A set of semantic categories

$$P = (p_1, p_2, \ldots\ldots)$$

A set of non-terminal symbols

$$Z = (z_0, z_1, z_2, \ldots\ldots), \qquad z_0 : \text{axiom}$$

$$M = S \cup Z = (\rho_1, \rho_2, \ldots\ldots)$$

A set of P's belonging to a word w

$$P(w) = (p_1(w), p_2(w), \ldots\ldots)$$

$$P_i(w) = (p_{w_{i_1}}, p_{w_{i_2}}, \ldots\ldots)$$

A set of words belonging to a syntactic word class s

$$W(s) = (w_{s_1}, w_{s_2}, \ldots\ldots)$$

The type of expansion rules:

$$z \rightarrow \mathcal{X}, \quad z \in Z$$

$\mathcal{X}$: string of symbols in M. ($\mathcal{X}$ is called a syntactic unit)

$$\mathcal{X} = \rho_{\mathcal{X}_1} \cdot \rho_{\mathcal{X}_2} \cdot \cdots \rho_{\mathcal{X}_n}$$

$M(\mathcal{X}) = P_{\mathcal{X}m}$: main constituent of string $\mathcal{X}$.
$NM_j(\mathcal{X}) = P_{\mathcal{X}j}$: non-main constituent of $\mathcal{X}$. j is attached 1,2,3,.. from left to right to the non-main constituents of string $\mathcal{X}$. If $\mathcal{X}$ is composed of only one symbol, there is no non-main constituent.

Optional elements in the expansion rules are indicated by a pair of brackets attached to the symbols. An optional element in $\mathcal{X}$ can not be the main constituent.

The type of selection rules:

$s \rightarrow w$ or $s \leftrightarrow w$

Semi-terminal derivation:

Expansion rules are applied on non-terminal symbols, to the stage where there is no symbol to be expanded. The final string is composed of SWC.

A set of the derived main constituents for a symbol z:

A set of all the SWC which can be the main constituent of a non-terminal symbol z or the main constituents of a phrase which is generated by successive expansions of the main constituents of the original z.

$$S(z) = (s_{z1} , s_{z2} , \ldots..)$$

$S(s_i) \equiv (s_i)$ is assumed.

3.2 The process of generation of a kernel sentence (I)

We suppose that a sentence has one central thing or concept to be mentioned first of all. This is the main constituent of a sentence. Then a second important concept is determined with its grammatical position, referring to the central concept already selected. Next a third important one is determined likewise, and so on. This process is formally represented in the following.

(i) $z_0 \leftrightarrow w(s_i(z_0))$

$s_i(z_0)$ is an element of $S(z_0)$ which is the set of the derived main constituents for z_0. $w(s_i(z_0))$ is a word belonging to the set $W(s_i(z_0))$. This shows the process starts from the selection of a word w for the axiom z_0, and the sentence is to be constructed with the core word w.

(ii) $z_0 \rightarrow \mathcal{X}$, if $s_i(z_0) \in S(M(\mathcal{X}))$

The axiom z_0 can be expanded into the syntactic unit $\mathcal{X}$ if and only if the already selected $s_i(z_0)$ at the stage (i) is contained in the set of the derived main constituents for $M(\mathcal{X})$.

(iii) $M(\mathcal{X}) \leftrightarrow w(s_i(z_0))$

The already selected word $w(s_i(z_0))$ is assigned to the main constituent $M(\mathcal{X})$ of the expanded syntactic unit $\mathcal{X}$.

(iv) $NM_k(\mathcal{X}) \leftrightarrow w_{m\ell k}(s_{\ell k}(NM_k(\mathcal{X})))$, for all k.

if a certain condition

$$f_{\mathcal{X}}(P(w), P(w_{m\ell 1}), P(w_{m\ell 2}), \cdots)$$

is satisfied.

To each non-main constituents $NM_k(\mathcal{X})$ is corresponded each word $w_{m\ell k}$ if the semantic categories for the words have a certain relation $f_{\mathcal{X}}$ with that of the word w assigned to $M(\mathcal{X})$.

At the n-th stage of the generation:
It is supposed that a word is already assigned to the symbol.

$$z \leftrightarrow w(s(z))$$

then:

(i) $z \rightarrow \mathcal{X}$, if $s(z) \in S(M(\mathcal{X}))$

(ii) $M(\mathcal{X}) \leftrightarrow w(s(z))$

(iii) $NM_k(\mathcal{X}) \leftrightarrow w_{m\ell k}(s_{\ell k}(NM_k(\mathcal{X})))$ if a certain condition

$$f_{\mathcal{X}}(P(w), P(w_{m\ell_1}), P(w_{m'\ell'_2}), \cdots\cdots)$$

is satisfied.

3.3 Condition $f_{\mathcal{X}}$

To all the elements of the semantic categories $P = (p_1, p_2, \ldots)$, the semantic distances are supposed to be defined.

$$d_{ij} = d(p_i, p_j) \qquad i \neq j$$
$$d_{ii} = d(p_i, p_i) = 0$$

The condition f may be the following.

$$z \rightarrow \mathcal{X} \quad (\equiv \rho_{\mathcal{X}_1} \cdot \rho_{\mathcal{X}_2} \cdot \cdots \cdot \rho_{\mathcal{X}_n})$$
$$M(\mathcal{X}) \leftrightarrow w(s(z))$$
$$NM_k(\mathcal{X}) \leftrightarrow w(s(NM_k(\mathcal{X})))$$
$$\beta_{\mathcal{X}} \leq \sum a_i d(p, p_k) + \sum b_{ij} d(p_i, p_j) \leq \alpha_{\mathcal{X}}, \qquad a_i, b_{ij}, \alpha_{\mathcal{X}}, \beta_{\mathcal{X}}: \text{constants.}$$

i, j, k over all non-main constituent symbols of $\mathcal{X}$.

$$p_k \in P(w(s(NM_k(\mathcal{X}))))$$
$$p \in P(w(s(z)))$$

An example of this process is illustrated in Fig. 1. The double line indicates the main constituent of a phrase symbol which is written one line above. Certain semantic conditions are imposed on the pair of phrase names in a phrase, which are underlined.

3.4 The process of generation of a kernel sentence (II)

The generation process explained in 3.2 is from the axiom. But there are the cases where we want to construct a sentence from arbitrary grammatical positions and a given word. For example when we write a complex sentence like "The old gentleman whom we saw at the theatre was his father.", the main constituent of the subordinate clause is not "gentleman", but the verb "saw". So we must generate a sentence from a noun "gentleman" and its grammatical position: objective case.

The process is that first the start point of generation is given by a word and its part of speech in a sentence. Next we select a proper rewriting rule which contains the part of speech of the word selected just now. Then to the remaining elements of the rewritten phrase the proper words are assigned, the semantic categories of which coincide with the one of the already selected word. This process is continued as far as there remains no element which can be rewritten by a phrase. The process is formally represented in the following.

(i) Given z, w, s, where $s \leftrightarrow w$, $s \in S(z)$

(ii) A tree structure whose top symbol is z is constructed by the method explained in 3.2.

(iii) $z' \rightarrow \mathcal{X}$, $\mathcal{X} = \rho_{\mathcal{X}_1} \cdot \rho_{\mathcal{X}_2} \cdot \cdots z \cdots \rho_{\mathcal{X}_n}$.
A phrase z' is selected which contains z as a component of the expansion rule $z' \rightarrow \mathcal{X}$.

(iv) $z \leftrightarrow w$, $\rho_{x_i} \leftrightarrow w_{x_i}$, $i = 1,2,\ldots\ldots,n$
where certain condition
$$f_X(P(w_{x_1}), P(w_{x_2}), \cdots, P(w), \cdots, P(w_{x_n}))$$
is satisfied.
A proper word is corresponded to each phrase ρ_{x_i}.
(v) Tree structures whose top symbols are ρ_{x_i} ($i = 1,2, \ldots$) are constructed for all ρ_{x_i} by the method mentioned in 3.2.
(vi) At the stage (iii), z' is newly replaced by z and the same operations from stage (iii) to (v) are performed.
(vii) When $z' = z_0$(axiom) is reached and the steps (iv) to (v) are completed, then the whole tree is accomplished under the axiom z_0.

An example of this process is illustrated in Fig. 2. The direction indicates the steps the sentence is constructed.

4. TRANSFORMATIONAL RULES

4.1 Representation of the rules

The transformational rules can explain many sentence structures which are difficult to treat in an immediate constituent method. For example in the sentence "Is he young?", which is a question form of "He is young", "is young" becomes discontinuous, separated by "he". This is difficult to treat by an immediate constituent method. It is far more natural to explain this by the application of a transformational rule concerning question to the original affirmative sentence.

The transformational rules we are now utilizing are classified to three types. The type 1 is unary transformations which may be thought of as converting a sentence from one to another. The type 2 is binary transformations which combine two sentences to form a third. And the type 3 is a transformation between two phrases. In all these cases we can formally represent the transformational rules as the following type.

$$z\ :\ X_1 \cdot X_2 \cdot \cdots\cdots \cdot X_n \rightarrow y_1 \cdot X_{i_1} \cdot y_2 \cdot X_{i_2} \cdot y \cdot \cdots\cdots\cdots \cdot y_m \cdot X_{i_m} \cdot y_{m+1} \qquad (1)$$

The symbol z which is written on the left side means that this transformational rules should be applied to the phrase z. $X_1, X_2, \cdots\cdots, X_n$ are either the elements of M or words themselves. Among them we have a special symbol $\emptyset$, which indicates that for this symbol $\emptyset$, there might or might not correspond some term in aninvestigated phrase z . That is, $\emptyset$ expresses an arbitrary term. $y_1, y_2, \cdots\cdots, y_{m+1}$ are vacant, some symbols or words which are not equal to $X_1, X_2, \cdots\cdots, X_n$. $X_{i_1}, X_{i_2}, \cdots\cdots, X_{i_m}$ are some symbols among $X_1, X_2, \cdots\cdots, X_n$.

The phrase z may have an internal tree structure, so the transformational rule is applied to this tree. An example of this is illustrated in Fig. 3. In this figure a noun phrase "the red books" is transformed to another noun phrase "the books which are red". This transformation is done by the rule,

NP1: $\emptyset \cdot$AD$\cdot$NQ $\longrightarrow$ $\emptyset \cdot$NQ$\cdot$WHICH BE$\cdot$AD

There are problems in the transformational rules such as follows.
(i) We have no definite criteria as to what kind of sentence structure is

to be treated in the scope of phrase structure grammar, and what is in the scope of transformational rules.
(ii) We can name empirically or informally the transformational rules such as passive, that deletion, complement/object transposition, etc., but to represent these rules formally in the form of (1) without contradiction for all the sentence structures generated from the specified phrase structure grammar, is difficult.
(iii) Transformations which accompany the changes in the part of speech or the morphophonemic forms of words are difficult to treat.
(iv) A transformational rule can not be applied unconditionally to the structure satisfying the rule form, but there are many cases where the application of rules depend on the semantics of the sentence.

4.2 Application of transformational rules

For the transformational rules of the type 1, we generate a sentence by the phrase structure grammar and at the same time memorize the generation steps of the sentence by the tree structure representation. Next we apply a transformational rule of the type 1 to this tree. If the rule is found to fit to the structure, then another tree is constructed from the original tree referring to the transformational rule.

Examples of this type are:

Ø·NP·VT·NP1·Ø ⟶ 1·4·BE·3·BY·2·5 (passive form)

This book emphasized the recent development clearly.

⟶ The recent development be emphasize(d) by this book clearly.

Ø·NP·VI2·NP1·Ø ⟶ 1·WHAT·DO·2·3·5·? (question)

Last year John became a doctor of philosophy at thirty.

⟶ Last year what do John become at thirty?

The application of morphophonemic rules to these transformed sentences are explained in §6.

For the transformational rules of the type 2, we generate first a sentence by the phrase structure grammar, with its internal tree structure. Then we select a proper phrase name which is a proper branch point of the tree, with the word attached to the phrase name. Next we start the generation of another sentence starting with the phrase name and the corresponding word, which are selected just now. This generation is by the method explained in 3.4. Then the two sentences thus generated have a same word, which is the key point in the usual transformational rules of type 2.

An example of this type is illustrated in Fig. 4. This is a combination of two sentences of Fig. 1 and Fig. 2. The rule applied here is,

SS: Ø·WT1·NP1·Ø·CM·NP1·Ø ⟶ 1·2·3·WHICH·7·4

and the generated sentence is

Several most number computer already precede specialist into trend which read the paper.

This example indicates that a transformational rules can not be applied in every case, even if the structure satisfies the rule form. There are many other examples of this nature.

For the transformational rules of the type 3, we have mainly concerned with the noun phrases which are the results of the application of transformational rules to certain phrases, especially to the sentence form SE.

For example,

NP1·BE·NP1 ⟶ 1·CM·3·CM
Kennedy is the president of the U.S.
⟶ Kennedy, the president of the U.S.,
NP1·BE·PP ⟶ 1·3
Scientists are in the dome of the south pole
Scientists in the dome of the south pole.

This type of transformational rules are incorporated in the generation by the phrase structure grammar.

Another important transformational rules are those which accompany the change in the part of speech of words.

For example,

VT1·NP1·Ø ⟶ <u>nn</u>·1·PRP·2·3
apply computer to the MT research
⟶ application of computer to the MT research
VT1·NP1·Ø ⟶ <u>nn</u>·1·BE GIVEN TO·2·3
consider the problems of the theory
⟶ consideration is given to the problems of the theory

We have not investigated yet this type of transformational rules except few ones, in which the word dictionary should have information about the interchange of the parts of speech.

5. SEMANTIC CATEGORIES AND THEIR RELATIONSHIP IN SYNTACTIC UNIT

In the generation process thus defined, each word is determined by the selection rule s ⟷ w applied to SWC's. How this word selection should be done is the semantics here considered. If the selection is done randomly without any semantic restriction, completely anomalous sentence will appear. To prevent this a new word is to be selected compatible with the already selected words which are in the neighborhood. Such semantic selection of words will especially be important in the syntactic relations such as

subject noun & predicate verb
subject noun & predicate verb & complement (or object)
adjective modifier & noun
noun & noun
adjective & adjective } (coordination structure)
verb & verb
etc.

Selection of a proper word in relation to the other words will eventually require the semantic notifications to the words and their mutual relationship in a sentence. In other words the system of semantic categories is to be set up and the meanings of all the words are to be represented in the system.

The construction of a system of the semantic categories may be done best by the replaceability relation among words in sentences. For example, to the verb "walk", there is a group of words which can be the subject to the verb "walk". To the word group thus formed, there will be another word group which can be the predicate and has a verb "walk" as its member. This word classification has not been tried yet on the whole scale, and indeed this is very difficult to do. So we have done a slightly different way, although the fundamental attitude of our word categorization is the replaceability of words in sentences.

We postulate that all the words might be properly characterized by setting up a number of key concepts. For example a word "voyage" is categorized as journey with the additional images such as amusement, time duration, ocean etc. In fact when we speak we actually construct sentences fully aware of such additional meanings.

Thus our aim is to extract such word images and to know how these images are mutually connected in such and such sentence structures. So we have started the extraction of semantic categories partly taking into consideration the Roget's thesaurus and some other publications. We have assigned the following numbers to the semantic categories of the parts of speech.

100--299	verb
300--499	noun
500--699	adjective
700--799	adverb
900--	preposition

The ten's digit indicates the rough semantic categories in a part of speech and the one's digit is to the further classifications. At present the number of categories for the verbs is about 40, for the nouns about 90, for the adjectives about 50, etc. For the prepositions we have attached a number to each word. These semantic categories are shown in table 1.

Next we have to clarify the connectivity of words in a sentence. To do this we have attached several kind of semantic categories to each word in the following way.

Verb:
- P1: categories intrinsic to the verb.
- P2: categories which can modify the verb (additional images of the verb)
- P3: categories which can stand as subject to the verb.
- P4: categories which can stand as object or complement to the verb.
- P5: special prepositions following the verb if any.
- P6: grammatical indication as to the form of the verb.

Noun:
- P1: categories intrinsic to the noun.
- P2: categories which can modify the noun (additional images of the noun)
- P3: grammatical indication as to the form of the noun.

Adjective:
- P1: categories intrinsic to the adjective.
- P2: categories which can modify the adjective (additional images of the adjective)
- P3: prepositions following the predicative adjective.

Adverb:
- P1: categories intrinsic to the adverb.

This expresses, for example, that a verb can take a noun for the subject whose semantic categories P1 belong to the P3 of the verb, and can take a noun for the object or complement whose semantic categories P1 belong to the P4 of the verb, etc. These are the conditions f_x introduced in 3.3. Examples of words having these connectivity informations are shown in table 2.

The generation process is thus first the selection of a verb, and then the determination of subject, object or complement referring to the semantic concordance mentioned here. Therefore each expansion rule of the phrase structure grammar has the indication such as

SE → NP <P1,P3> VT1 · NP1 <P1,P4> · PRP <P1,P5> · NP1

NP → ADJ <P1,P2> · NQ

In this example VT1, NQ which are underlined are the main constituents to which some concrete words are selected. So for NP, NP1, PRP and ADJ, proper words are to be assigned having the semantic concordance between the pair of categories bracketed by < > . The first element in < > is the category of the phrase to the left of < , and the second element is that of the main constituent. Therefore a word is assigned to NP, whose semantic categories P1 have the same term in the P3 of VT1, and so on.

But there are many grammatical phrases where we can not tell what kind of semantic relationship are to be established. We have not attached the semantic relationship to the phrases like,

main verb : adverbial phrases preceded by preposition
main sentence : subordinate clause
noun : its adjectival clause
etc.

It is also difficult to find out the semantic relationship between the nouns of the form,

NO + NO, NO + NO + NO, NO of NO

such as,

machine translation, information processing machine
generation of sentences.

However when these phrases are given from suitable transformations of another phrases such as

solution of a problem ↔ solve a problem,
generation of sentences ↔ generate sentences,

we can establish the semantic relationship of these two nouns in the phrase before the transformation is applied. In this case we have to know the noun form of a verb or its vice versa. This information is contained in the word dictionary as P6 for the verb and as P3 for the noun.

The semantic relationship here introduced is essentially the connectivity between two words in a phrase, so there is a possibility of generating absurd sentences. To prevent this we have to know more minute mutual influence of meanings among words in a sentence.

6. MORPHOPHONEMIC RULES

We want to propose that the morphophonemic rules can be represented by a kind of operators operating on the words in the neighborhood. We include negation action, tense, case etc. to this level. We take the operators such as follows.

n (not), pr (present tense), ps (past tense)
inf (infinitive), sg (third person singular), pl (plural)
sub (subjective case), obj (objective case)

ing (ing-form of a verb), ed (past participle)
nn (nominalization of a verb), etc.

The functions of these operators are,

n + verb ⟶ do + not + verb
n + aux. verb ⟶ aux. verb + not
pr + verb ⟶ verb (present form)
ps + verb ⟶ verb (past form)
ing + verb ⟶ verb (gerund or present participle)
ed + verb ⟶ verb (past participle) etc.

Besides these, for the verb BE(b) and HAVE(h), the following three steps are to be applied in this order.

(1) b + b + verb ⟶ be + being + ed + verb
(2) b + verb ⟶ be + ing + verb
(3) h + be ⟶ have been

For example in the generation of a sentence shown in Fig. 5, the sentence obtained initially is

the father sg sub h b enjoy fresh breeze sg obj

Here the following operations are applied.

sg obj ⟶ obj sg, sg # ⟶ #
noun obj ⟶ noun, b enjoy ⟶ be + ing + enjoy
ing + enjoy ⟶ enjoying, h + be ⟶ have been
sg sub ⟶ sub sg, sg have ⟶ has
noun sub ⟶ noun

And we can get the final form,

the father has been enjoying fresh breeze

These operators can appear both in the phrase structure grammar and in the transformational rules, but the operations of these are supposed to be done after the application of transformational rules. But there occur many complicated situations for the sentences after the application of transformational rules and to what extent we can go on this line remains to be seen in the future.

7. EXPERIMENTS

We are doing experiments on sentence generation by a medium size computer KDC-1 installed in our university. The word size is about 450 (verb about 140, noun about 170, adjective about 80, adverb about 40, etc.) The phrase structure rules are about 80 of the form $z \rightarrow \rho_1 \cdot \rho_2 \cdots \rho_n$ (max. of n = 5 at present). The rules include the optional terms so that effectively the number of rules increases.

The transformational rules are now only fundamental ones such as, negation, passive transform, several kind of question form, nominalization of phrases, several kind of binary transformations, etc.

We treat the sentences such as

$S + V + O_1 + O_2$
$S + V + O + C$

by the binary transformation from two sentences of the form,

$$S + V + X \ : \ X = O_1 \text{ have } O_2 \longrightarrow S + V + O_1 + O_2$$
$$S + V + X \ : \ X = O \text{ be } C \longrightarrow S + V + O + C$$

Examples of generated sentences:

Working to set for new present-day of area be abroad.
To correspond to only research, different monograph work, be above such its branch of difficulty.
To be being between selection invitation to assemble, a completely immense technique be to know desirable boundary.
Of widely digital elimination of number, view of summary question which arise to be automatic here satisfy his directly attractive subject at only impossible reference.

We think that this study of sentence generation of English suggests something to the translation technique to English from Japanese which has no sentence structure like English.

In parallel to this study we are now experimenting just the reverse procedure of sentence generation. That is, we are trying to decompose a given sentence into several semantic units or set of kernels. This process will clarify in a subjective sense the amount of information contained in a given original sentence. Moreover this will contribute very much to the field of information retrieval and also to the clarification of the logic of a discource, such as question-answer problem.

REFERENCES

(i) V.H. Yngve: Random generation of English sentences, 1961 International Conference on Machine Translation of Languages and Applied Language Analysis, London, 1961

(ii) C.J. Fillmore: The position of embedding transformations in a grammar, The Ohio State University Rf Project, POLA No. 3, 1963

(iii) J.J. Katz & J.A. Fodor: The structure of a semantic theory, Language, Vol.39, No.2, 1963

(iv) S. Ceccato: Linguistic Analysis and Programming for Mechanical Translation, Gordon and Breach Science Pub. 1961

(v) M. Nagao: An Approach to The General Theory of Natural Language, Preprints for Seminar on Mechanical Translation, U.S.-Japan Committee on Scientific Cooperation, Pannel II, April 1964

TABLE 1-1. SEMANTIC CATEGORIES (VERB)

Hand(110)		
	(111)	close, open, cover, fill, hold, drop, mark, plant, put, fire, take, draw, make
	(112)	keep, carry, use have, get, give, help, bear, raise, hold, take
	(113)	connect
Eyes etc.(120)		
	(121)	hear, find, see, look, watch
	(122)	say, speak, talk, tell, call, laugh, order, read, sing, state, cite, pronounce
Intelligent(130) behaviour		
	(131)	ask, answer, hear, find, order, cite, address
	(132)	add, get, receive, send, need, select, treat, eliminate, accept, arrange
	(133)	learn, read, find, write, see, process, apply, program, compute, speciaroze, compare, judge, indicate, understand, translate, know
	(134)	deal, treat
	(135)	interest, experience
Mental(140)		
	(141)	feel, remember, think, reflect
	(142)	enjoy, thak, love, fear, like,
	(143)	wish, hope, care, want
	(144)	stimulate, attract
Spiritual(160)		
	(161)	believe, know, mind, think, mean
	(162)	attempt, aim, intend
Meal(170)		drink, eat
Social(180)		
	(181)	build, found, publish, generate, develope, assemble, bridge
	(182)	kill, care
	(183)	pay, receive, buy, get, give, save, need, present, provide, precede, concede
	(184)	drive, fly, sail, ride, guide
	(185)	follow, lead, elect, participate
	(186)	live, work, make, cooperate, exist
	(187)	contribute, serve, help, save, pronounce
	(188)	consist, exist
	(189)	correspond, concern,
Body action(190)		
	(191)	play, show, try
	(192)	sit, stand, start, stop, put, set
	(193)	visit, call meet, show, see, appear, address
	(194)	sleep, rest, remain, wait
	(195)	walk, move, pass, leave
	(196)	come, go, reach, run, stay, arrive
Change of state(230)		
	(231)	change, turn, arise, remain
	(232)	start, go, come, drop, leave, begin
	(233)	extend, follow, increase, form

<u>Natural phenomena</u>(250)		
	(251)	rain, blow, cover, drop
	(252)	burn, fire

TABLE 1-2. SEMANTIC CATEGORIES (NOUN)

<u>Human beings</u>(300)		
man	(301)	boy, child, girl, person, man, woman, Jack, Betty, Nelson
family	(302)	brother, company, family, father, friend, people, sister, son, mother
social	(303)	king, soldier, god, president
	(304)	specialist, reader, scientist, professional, (group), blind, editor, debutant
	(305)	generation, group
<u>Nature</u>(340)		
celestial	(341)	sun, moon, earth
atmospheric	(342)	rain, wind, air
geographic	(343)	river, hill, mountain, road, land, field, sea Mt. Fuji, Kaatskill, Appalachia, Lake Biwa
minerals	(344)	rock, stone, gold, silver
water	(345)	water, sea, rain
<u>Large things</u>(360)		
movable	(361)	bus, train, car, ship
building	(362)	house, church, school, Kyoto Univ.
parts of 362	(363)	door, window, room
place	(364)	road, street, garden
<u>Articles</u>(370)		
books	(371)	book, picture, paper, story, Newsweek, Bible, handbook, library, monograph, article, summary literature, supplement, journal, proceeding, report, volume, text-book
foods	(372)	food, egg, bread, milk, corn,salt, pepper, water, bear
furniture	(373)	table, box, bed, dress
playthings	(374)	ball, teniss
	(375)	processor, machine, computer
<u>Mental action</u>(380)		
thinking	(381)	feason, idea, hope, mind, thought
feeling	(382)	love, life, fear
	(383)	aim, end
	(384)	readiness
	(385)	gratitude, thank, patience, acknowledgement
	(386)	knowledge, thought, view, opinion, reference, aspect, conception, comment, consideration, understanding
	(387)	sense, art, view
<u>Action</u>(390)		
	(391)	life, love
	(392)	war
	(393)	question, answer, speech, call, order, judgement, citation, problem example
	(394)	death, existence
	(395)	visit, watch, indication, advance, access, response, change

(396) work, help, treatment
(397) voyage, play, trip
(398) research, work, study, aid, contribution
(399) selection, processing, use, application, programming, elimination, addition, presentation, publication, specialization

Abstract(400)

(401) name, word
(402) thing, matter, something, state
(403) way, form, mean, point
(404) case, matter, measure, course, use
(405) cause, change, end
(406) color, sight
(407) sound, voice

(410)

(411) group,
(412) gap, point, boundary, link
(413) interest, value, validity
(414) limitation, difficulty, success, problem
(415) kind, respect, branch, feature

Social terms(420)

(421) money
(422) bank, company, shop
(423) law, right
(424) city, country, street, town
(425) world, country, Japan, America, nation

(430)

(431) data, language, word, German, English, French, reference
(432) technique, science, manner, way, methodology, originality
(433) particular, detail, summary, assembly
(434) translation, generation, development, appearance comparison, arangement, cooperation
(435) information, fact, source, consequence
(436) subject, material, list, listing title,

(470)

(471) field, area, circle, center
(472) course, state, live, manner, way, stage
(473) context, implication, content, indication, source

TABLE 1-3. SEMANTIC CATEGORIES (ADJECTIVE)

Qualifying

state of the object

outward(530)

color	(531)	black, blue, green, red, white
shape	(532)	round, plain
length	(533)	long, short
height	(534)	deep, high, low
extent	(535)	wide, narrow
size	(536)	large, little, small, big, least
	(537)	single, individual, numerous, multiple

subjective(550)			
	beauty	(551)	beautiful, pretty
	fair	(552)	fair, clear, fine
	free	(553)	free, fresh
	full	(554)	full, complete
		(555)	undisputed
internal(570)			
	material	(571)	gold,
	weight	(572)	light, heavy
	hardness	(573)	hard, soft
	temperature	(574)	cold, warm, hot
	new, old	(575)	new, old, fresh
	soft,	(576)	soft, swee, fresh
social state(590)			
	wealth	(591)	rich, poor, cheep, modest
	position	(592)	great, deep
	fact	(593)	true, correct, natural
	restruction	(594)	free, strict, major
		(595)	skilled, detailed
		(596)	commercial, available
		(597)	neighboring
mental state(600)			
	sentiment	(601)	glad, happy, sad
	intimacy	(602)	dear, ready, desirable, familiar
		(603)	aware, wary, afraid
		(604)	active, attractive
		(605)	proud
time(620)		(	
		(621)	fast, quick
		(622)	ready
		(623)	late, recent, current, present, final
valuation(640)			
	good, bad	(641)	good, bad, right, modest, valuable
		(642)	very, immense
		(643)	natural
abstruct(660)			
		(661)	general, collective, common, particular, special, comprehensive, standard, specific
		(662)	next, certain
		(663)	present, absent, conventional
		(664)	related, informed, concerned
		(665)	complex, detailed, bound, complate
relation(670)			
		(671)	right, left
		(672)	close, near, intermediate
		(673)	direct, conclusive, extensive
		(674)	individual
		(675)	possible, impossible, necessary, enough
	(680)		
		(681)	digital , linguistic, automatic spectral, t technical, scientific
		(682)	leterary, professional
		(683)	world wide

TABLE 2. Semantic categories attached to words

VERB

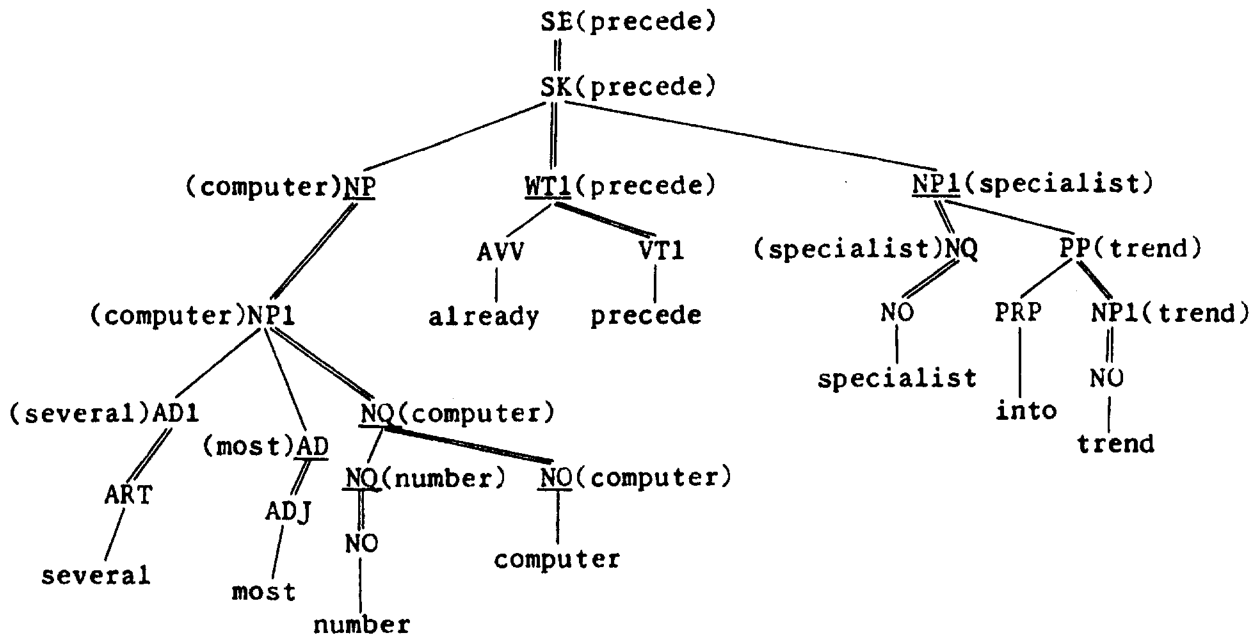
add	P1 :	132
	P2 :	730 760
	P3 :	300
	P4 :	369 420
	P5 :	990
answer	P1 :	131
	P2 :	730 740
	P3 :	300
	P4 :	300 393 395 396
bear	P1 :	112
	P2 :	722 741
	P3 :	300 320 340 360
	P4 :	381 400
	P5 :	960
begin	P1 :	232 260
	P3 :	300
	P4 :	390
believe	P1 :	160
	P2 :	743 744
	P3 :	300
	P4 :	300 380 400
bring	P1 :	196
	P3 :	300 320 341 361 380
	P4 :	300
	P5 :	980
build	P1 :	181
	P3 :	300
	P4 :	346 361 364 368
hear	P1 :	121 131
	P3 :	300 320
	P4 :	390 380 400 412 413 414 418
help	P1 :	112 186
	P3 :	300 341 342 360 380
	P4 :	300 320 380 390 400
sea	P1 :	193 121 150
	P3 :	300 320
	P4 :	300 310 320 330 340 360 380
think	P1 :	160 130
	P3 :	300 425
	P4 :	380 390 402 403 405 425
enjoy	P1 :	142
	P3 :	300 320 425
	P4 :	340 372 391 395 397 463

NOUN

answer	P1 :	393
	P2 :	510 520 550 590 640 660
apple	P1 :	330
	P2 :	510 520 531 532 536 551 575
bank	P1 :	368
	P2 :	536 575 591 592 640 660
bird	P1 :	322
	P2 :	510 520 530 551 578
book	P1 :	362
	P2 :	510 520 530 551 572 641
change	P1 :	413
	P2 :	520 536 550 590 620 640
country	P1 :	384 385
	P2 :	510 520 536 551 590
flower	P1 :	330
	P2 :	531 536 551 571 576 591
hour	P1 :	410
	P2 :	597 594 620 640 660
order	P1 :	393
	P2 :	536 553 573 621 640 660
fear	P1 :	391 395
	P2 :	534 536 578 592 593
love	P1 :	382 383
	P2 :	550 574 576578 593 595
right	P1 :	388 411
	P2 :	554 592 593 594 640
sight	P1 :	405
	P2 :	534 535 551 574 597 660

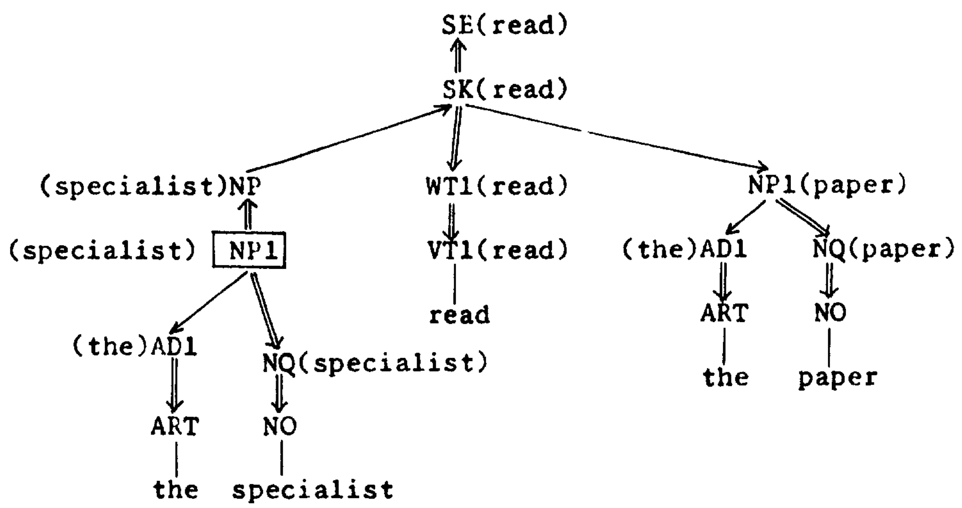
ADJECTIVE

old	P1 :	575 581 623
fresh	P1 :	575 576 553
strong	P1 :	573 582
great	P1 :	592
plain	P1 :	532 596
free	P1 :	553 594
right	P1 :	538 593 641



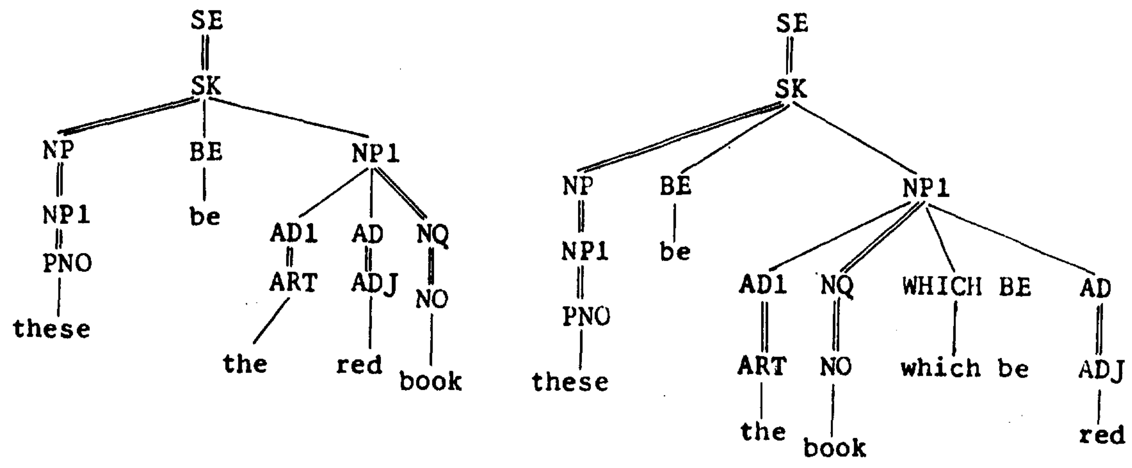
Several most number computer already precede specialist into trend.

Fig.1 Sentence generation from the axiom and a verb "precede"



The specialist read the paper.

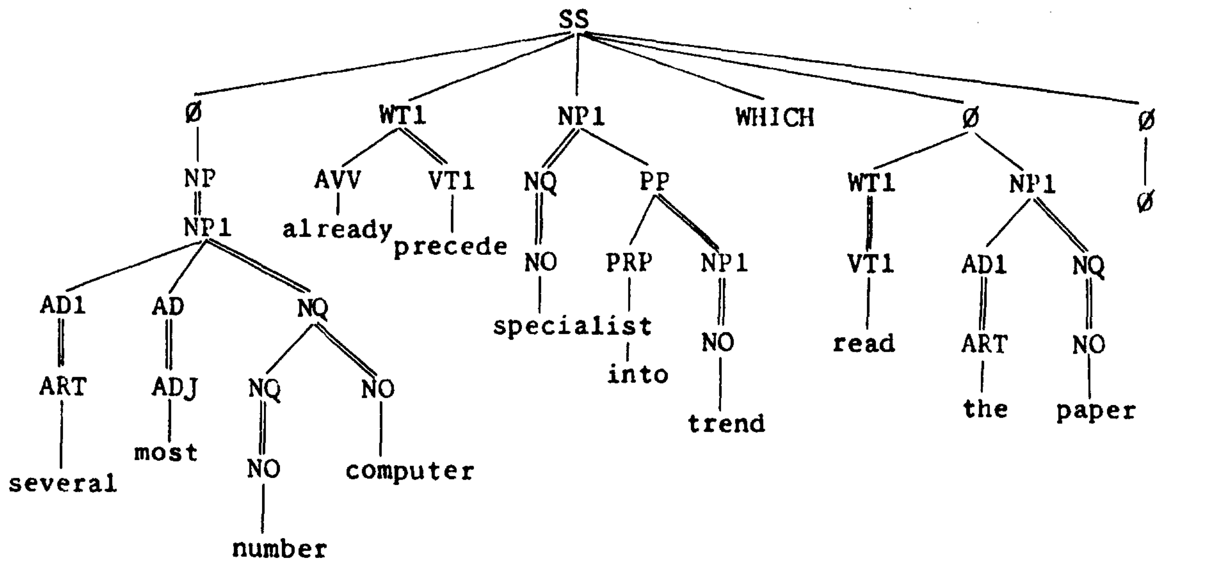
Fig.2 Sentence generation from NP1 and a noun "specialist"



These be the red book. These be the book which be red.

NP1 : Ø·AD·NQ ⟶ Ø·NQ·WHICH·BE·AD

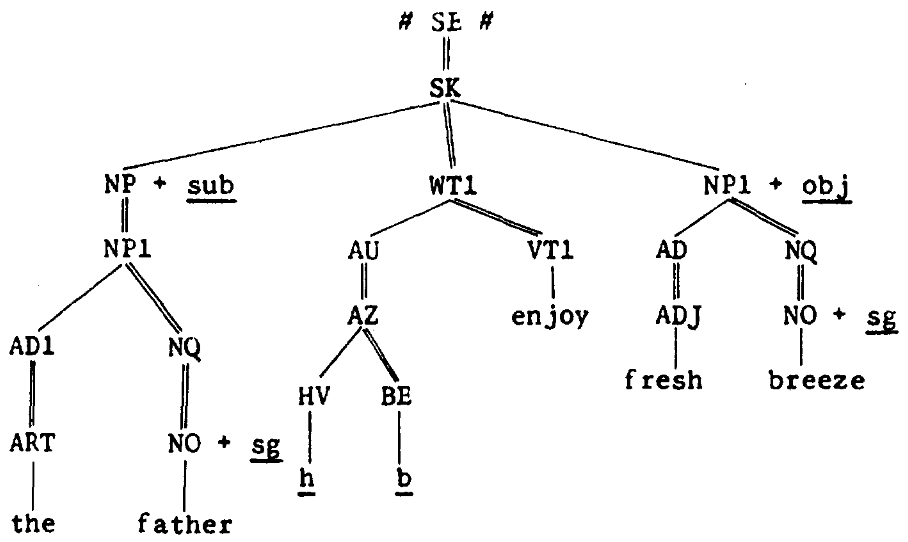
Fig.3 Transformation in a phrase.



Several most number computer already precede specialist into trend which read the paper.

SS : Ø·WT1·NP1·Ø·CM·NP1·Ø ⟶ 1·2·3·WHICH·7·4

Fig.4 Transformation between two sentences.



the father sg sub h b enjoy fresh breeze sg obj

Fig.5 Application of morphophonemic rules as operators.