Sebastian K. Šaumjan - P. A. Soboleva

FORMAL METALANGUAGE AND FORMAL THEORY AS TWO ASPECTS OF GENERATIVE GRAMMAR

As is known, metalanguage is a language by means of which another language is described. The latter is called object-language.

One and the same language may play the role of object-language and metalanguage. For example, the Russian language taken as an object of linguistic description is an object-language, the Russian language used for a linguistic description of Russian is a metalanguage.

There are non-formal and formal metalanguages. For example, the Russian language used in ordinary grammars of Russian as a means of describing the Russian language is a non-formal metalanguage. Formal metalanguage is an artificial language, defined by deductive rules of construction, and used to describe natural languages.

The problem of formal metalanguages for linguistic descriptions is a broad topic which we do not mean to exhaust. We shall restrict ourselves to clarifying the role of formal metalanguages in constructing generative grammars.

Every generative grammar is a formal theory. For example, a generative grammar of Russian is a formal theory of Russian, a generative grammar of English is a formal theory of English. We shall focus our attention on the question of correlation between formal metalanguages and formal linguistic theories understood as generative grammars. Let us take a concrete example: formal metalanguage and formal theory as two aspects of applicative generative grammar. This concrete example will make it possible, we hope, to draw certain general conclusions about the role of formal metalanguages in generative grammar.

We shall start with formal metalanguage in applicative generative grammar. We call this metalanguage a universal operator language.

The universal operator language is defined by the grammar which is an ordered quadruple

$$\langle V_l, V_s, F, R \rangle$$

where:

1. V_i is the lexicon of symbols whose denotata belong to a finite set of elementary categories called elementary episemions. The elementary episemions are denoted by the symbols α and β. The first is interpreted as the category of terms, and the second – as the category of sentences. The episemion β is called the distinguished episemion.

Complex categories called episemions are constructed from elementary episemions. The construction of episemions is effected with the help of the function Δ, which maps episemions into each other. The rules for constructing episemions are the following:
    (a) elementary episemions are episemions
    (b) if p and q are episemions, then $\Delta\,pq$ is also an episemion.
Rule (b) presented as a tree looks as follows:

(1)
$$\frac{p \quad q}{pq}$$

For example:

(2)
$$\frac{\alpha \quad \beta}{\Delta\alpha\beta}$$

The episemion $\Delta\alpha\beta$ may be interpreted as the category of one-place predicates (i.e. the function which maps terms into sentences), for example *burn, arrive, sleep, stay* etc.

(3)
$$\frac{\alpha \quad \alpha}{\Delta\alpha\alpha}$$

The episemion $\Delta\alpha\alpha$ may be interpreted as the category of term identifiers (i.e. the function which maps terms into terms), for example *big, good, John's, burning* (in *burning eyes, burning house*) etc.

(4)
$$\cfrac{\alpha \quad \cfrac{\alpha \quad \beta}{\Delta\alpha\beta}}{\Delta\alpha\Delta\alpha\beta}$$

The episemion $\Delta\alpha\Delta\alpha\beta$ may be interpreted as the category of two-place predicates (i.e. the function which maps terms into one-place predicates or, in other words, the function which maps two terms into sentences), for example *read, see, expect, like* etc.

$$(5) \quad \frac{\dfrac{\alpha \quad \beta}{\Delta\alpha\beta} \quad \alpha}{\Delta\Delta\alpha\beta\alpha}$$

The episemion $\Delta\Delta\alpha\beta\alpha$ may be interpreted as the category of depredicators (i.e. the function which maps one-place predicates into terms), for example -al (in arrival), -ation (in relaxation) etc.

$$(6) \quad \frac{\alpha \quad \dfrac{\alpha \quad \alpha}{\Delta\alpha\alpha}}{\Delta\alpha\Delta\alpha\alpha}$$

The episemion $\Delta\alpha\Delta\alpha\alpha$ may be interpreted as the category of denominative adjectivisers (i.e. the function which maps terms into term identifiers), for example -y (in stony path), -less (in heartless man), 's (in John's house), of (in the house of John) etc.

$$(7) \quad \frac{\dfrac{\alpha \quad \beta}{\Delta\alpha\beta} \quad \dfrac{\alpha \quad \alpha}{\Delta\alpha\alpha}}{\Delta\Delta\alpha\beta\Delta\alpha\alpha}$$

The episemion $\Delta\Delta\alpha\beta\Delta\alpha\alpha$ may be interpreted as the category of adjectivising depredicators (i.e. the function which maps one-place predicates into term identifiers), for example -ing (in burning house), -y (sleepy child) etc.

2. V_s is the lexicon of symbols whose denotata are elementary meanings. These symbols are called elementary semions. For example

T – an elementary semion which is interpreted as an object belonging to the category of terms.

P_1 – an elementary semion interpreted as an object belonging to the category of one-place predicates.

P_2 – an elementary semion interpreted as an object belonging to the category of two-place predicates.

A – an elementary semion interpreted as an object belonging to the category of term identifiers.

D_T – an elementary semion interpreted as an object belonging to the category of depredicators.

D_A – an elementary semion interpreted as an object belonging to the category of adjectivizing depredicators.

5

Sf_A – an elementary semion interpreted as an object belonging to the category of denominative adjectivisers etc.

3. *F* is a so-called assignment function which places each elementary semion in correspondence with one or several episemions. The work of the assignment function may be represented by the formula

$$eX$$

where *e* is any episemion and *X* is a semion with which it has been placed in correspondence, for example

$$\alpha \quad T$$
$$\Delta\alpha\beta \quad P_1$$
$$\Delta\alpha\Delta\alpha\beta \quad P_2$$
$$\Delta\alpha\alpha \quad A$$
$$\Delta\alpha\beta\alpha \quad D_T$$
$$\Delta\Delta\alpha\beta\Delta\alpha\alpha \quad D_A$$
$$\Delta\alpha\Delta\alpha\alpha \quad Sf_A \qquad \text{etc.}$$

4. *R* – are rules for constructing combinations of elementary semions which are called semions. These rules are as follows:
(a) an elementary semion is a semion;
(b) if *X* is a semion which belongs to the episemion $\Delta\ pq$, and *Y* is a semion, belonging to the episemion *p*, then (XY) is a semion, belonging to the episemion *q*.
Rule (b) presented as a tree looks as follows:

$$(8) \qquad \frac{\Delta pqX \quad pY}{q(XY)}$$

We shall call *X* an operator, *Y* – an operand, and (XY) is the result of application of *X* to *Y*. Rule (b) is called the rule of application of semions.
Rule (b) can be illustrated in the following way:

$$(9) \qquad \frac{\Delta\alpha\beta P_1 \quad \alpha T}{\beta(P_1 T)}$$

If P_1 is interpreted as *stay* and *T* as *John* then the semion $P_1 T$ is interpreted as the sentence *John stays*. (Note that in the abstract operator language the operator always precedes the operand. So a natural lan-

guage analogue of the semion $P_1 T$ is *stays John* rather than *John stays*).
Below are given several more illustrations of rule (b):

$$(10) \qquad \dfrac{\dfrac{\Delta\Delta\alpha\beta\Delta\alpha\alpha D_A \quad \Delta\alpha\beta P_1}{\Delta\alpha\alpha(D_A P_1) \qquad \alpha T}}{\alpha((D_A P_1)T)}$$

If the elementary semion D_A is interpreted as *-ing*, P_1 as *burn* (intr.),
then the semion $D_A P_1$ is interpreted as *burning*. If T is interpreted as
house, then the semion $((D_A P_1)T)$ is interpreted as *burning house*. (The
exact order of elements being ((*-ing burn*) *house*)).

$$(11) \qquad \dfrac{\dfrac{\Delta\alpha\Delta\alpha\alpha Sf_A \quad \alpha T}{\Delta\alpha\alpha(Sf_A T)} \quad \dfrac{\Delta\alpha\beta\alpha D_T \quad \Delta\alpha\beta P_1}{\alpha(D_T P_1)}}{\alpha((Sf_A T)(D_T P_1))}$$

If Sf_A is interpreted as *'s*, T – as *John*, then $Sf_A T$ corresponds to *John's*.
If D_T is interpreted as *-al*, P_1 – as *arrive*, then $D_T P_1$ corresponds to *ar-
rival*. The derived semion $((Sf_A T)(D_T P))$ serves as a genotype analogue
of the noun phrase *John's arrival* (the exact order of elements being
((*'s John*) (*-al arrive*))).

Such is the grammar which defines the construction of the univer-
sal operator language. If an empirical interpretation is assigned to epi-
semions and semions, the universal operator language may serve as
a formal metalanguage for the description of any natural language.
The universal operator language has sufficient potentialities to construct
semions which serve as abstract analogues of sentences of any degree
of complexity. By way of an example we shall construct semions
which serve as abstract analogues of the following complex phrases:

I see a burning house (12) and
I expect John's arrival (13).

The tree below shows the construction of the genotype analogue
of the first phrase:

$$(12) \qquad \dfrac{\dfrac{\dfrac{\dfrac{\Delta\Delta\alpha\beta\Delta\alpha\alpha D_A \quad \Delta\alpha\beta P_1}{\Delta\alpha\alpha(D_A P_1) \qquad \alpha T^1}}{\Delta\alpha\Delta\alpha\beta P_2 \quad \alpha((D_A P_1)T^1)}}{\Delta\alpha\beta(P_2((D_A P_1)T^1)) \qquad \alpha T^2}}{\beta((P_2((D_A P_1)T^1))T^2)}$$

The final semion $((P_2 \, ((D_A P_1) \, T^1)) \, T^2)$ which is an object belonging to the category of β, i.e. sentence, is an abstract analogue of a two-place predicate sentence with a participle defining one of its terms.

The construction of the second semion is shown below:

$$
(13) \quad
\cfrac{
\cfrac{
\cfrac{\Delta\alpha\Delta\alpha\alpha Sf_A \quad \alpha T^1}{\Delta\alpha\alpha(Sf_A T^1)} \quad \cfrac{\Delta\Delta\alpha\beta\alpha D_T \quad \Delta\alpha\beta P_1}{\alpha(D_T P_1)}
}{
\cfrac{\Delta\alpha\Delta\alpha\beta P_2 \quad \alpha((Sf_A T^1)(D_T P_1))}{\Delta\alpha\beta(P_2((Sf_A T^1)(D_T P_1)))} \quad \alpha T^2
}
}{
\beta(P_2((Sf_A T^1)(D_T P_1)) T^2)
}
$$

The final semion $P_2((Sf_A T^1) \, (D_T P_1)) T^2$ which also belongs to the category of β, i.e. sentence, is an abstract analogue of a two-place predicate sentence with a nominalized phrase as one of its arguments.

However complex the constructed semion may be, the process of construction fails to show how two primitive structures of the type *I see a house* and *The house is burning* are transformed into (12), the second becoming a participial phrase in (12), or how two primitive structures of the type *I expect smth.* and *John is arriving* are transformed into (13), the second becoming a nominalized phrase in (13).

Though the construction of semions accounts for the generation of the abstract analogues of sentences of any degree of complexity it does not aim to show such essential grammatical processes as nominalization or adjectivation of phrases, formation of complex predicates from simple ones or of simple sentences from complex sentences etc. So in spite of the abstract character of the universal operator language it does not allow us to rise above the level of taxonomic description of natural languages.

The necessity for the formal theory of natural languages follows from the two-level principle in linguistic studies. According to the two-level principle every natural language is stratified into two languages: the phenotype language and the genotype language. The phenotype language is a natural language as is given in immediate observation. The genotype language is hidden, not given in immediate observation; it is a construct language which consists of two sub-languages: the basic genotype language and the derived genotype language. The basic genotype language is the language of thought representation; the expressions of the basic genotype language which consist of elementary predicates may be identified with the thoughts which are to be embodied into a linguistic form. The embodiment

of thoughts into a lingustic form is effected in two stages. At the first stage the expressions of the basic genotype language, which are identified with thoughts, are transformed into the expressions of the derived genotype language, regarded as abstract linguistic forms which embody thoughts. At the second stage the expressions of the derived genotype language are transformed into the expressions of the phenotype language regarded as the concrete linguistic forms of thought embodiment. Thus, the derived genotype language may be considered as an intermediary language between the basic genotype language and the phenotype language.

The transformations used to convert phrases of the basis genotype language into phrases of the derived genotype language may be described with the help of logical operators called combinators.[1]

Since natural languages serve the purpose of communication between people, the grammar of every natural languages must possess the means for the transformation of thoughts into linguistic forms and, vice-versa, the means for the transformation of linguistic forms into thoughts.

Grammar viewed in such a way is a hypothetic transducer, which is not given in immediate observation, but which makes it possible to transform the phrases of the basic genotype languages into the phrases of the derived genotype language, and the phrases of the derived genotype language into the phrases of the phenotype language, and viceversa, the phrases of the phenotype language into the phrases of the derived genotype language, as well as the phrases of the derived genotype language into the phrases of the basic genotype language.

The formal theory of natural languages must use the universal operator language as a formal metalanguage for a description of the basic genotype language, the derived genotype language and the phenotype language.

Below we shall specify the type of rules for postulating the basic genotype language and the type of rules for obtaining the derived genotype language. The basic genotype language consists of one-, two- and three-place predicates obtained through the rule:

$$(14) \qquad S \rightarrow \begin{cases} P_1 \check{S} \\ P_2 \check{S} T \\ P_3 T^1 \check{S} T^2 \\ C j \check{S} \check{S} \end{cases}$$

---

[1] See. S. K. ŠAUMJAN (1971), pp. 127-133.

In (14) S is a global symbol of sentence, i.e. a semion belonging to the category β. P_1 in (14) is a one-place predicate with a sentence as its argument. Its episemion is ΔββΔ. P_2 is a two-place predicate with a sentence and a term as its arguments. Its episemion is ΔβΔαβ. P_3 is a three-place predicate with a sentence and two terms as its arguments. The episemion ascribed to it is ΔαΔβΔαβ. Cj is a two-place operator of conjunction with sentences as its arguments. Its episemion is ΔβΔββ. $\check{S}$ is a sentence which is an argument of P_1, P_2 and P_3 (an embedded sentence). Its episemion is β. The status of T has been discussed on page 65. It is an object belonging to the category of terms. In (14) it is an argument of P_1, P_2 and P_3.[2]

The one-place predicate P_1 corresponds to modal words in natural languages, such as *possibly, probably, perhaps* etc. It may also correspond to aspectual words, such as *begin, continue, stop.* The two-place predicate P_2 corresponds to volitive predicates such as *wish, want, like, dislike, hate* etc., emotive predicates such as *amuse, upset, distress, intimidate,* etc., mental perception predicates, such as *know, expect, think, believe* etc., P_2 may also correspond to some other classes of predicates which we shall omit for lack of space. The common feature which brings together these predicates is their status of two-place operators: applied to an embedded sentence and then to a noun-phrase they yield a sentence. For example: *want (come John) I, believe (come Mary) Nick.*

The three-place predicate P_3 corresponds to causative and communicative verbs such as *order, persuade, say, tell, beg, implore.* For example: *order man (open fire man) officer, tell John (come Mary) Nick.*

The above illustration of the two- and three-place predicates is given in the so-called hybrid language. The hybrid language is also a formal operator language. It answers the description given on pp. 63-66 except for the composition of V_S. If in the operator metalanguage proper V_S is the lexicon of symbols whose denotata are elementary meanings called elementary semions (predicates, terms), in the hybrid language the role of semions is played by the lexical stems of a concrete natural language. As a result a mixed language is obtained. Its grammatical component is the applicative grammar, and its lexical component is constituted by the lexical items of a concrete language.[3]

The embedded sentence $\check{S}$ in (14) is rewritten as either a one-place

---

[2] For greater detail see S. K. ŠAUMJAN (1971), pp. 137-139.

[3] See S. K. ŠAUMJAN (1971), pp. 134-135.

or a two-place predicate of the type $\check{P}_1 T$, $\check{P}_2 T^1 T^2$, where $\check{P}_1$ and $\check{P}_2$ are predicates with terms as their arguments (rather than embedded sentences as in case of P_1, P_2 and P_3 in (14)) or as predicates, such as in (14):

$$(15) \qquad S \rightarrow \begin{cases} \check{P}_1 T \\ \check{P}_2 T^1 T^2 \\ P_1 \check{S} \\ P_2 \check{S} T \\ P_3 T^1 \check{S} T^2 \\ Cj \check{S} \check{S} \end{cases}$$

If the names of the predicates: modal Md, aspective As, volitive Vl, mental perception Mt, causative Ca and communicative Cm are introduced in the genotype language as constants which are substituted for the variables P_1, P_2 and P_3 as in (16)-(18):

$$(16) \qquad P_1 \rightarrow \begin{cases} Md \\ As \end{cases}$$

$$(17) \qquad P_2 \rightarrow \begin{cases} Vl \\ Em \\ Mt \end{cases}$$

$$(18) \qquad P_3 \rightarrow \begin{cases} Ca \\ Cm \end{cases}$$

a great variety of sentences can be obtained which constitute an essential part of the basic genotype language.

For example, a basic genotype language sentence

$$(19) \qquad Md(Mt(\check{P}_1 T^1) T^2)$$

is obtained by rule (14), a double application of rule (15), rules (16) and (17). It may be interpreted as *Perhaps, you think she is wrong.* A basic genotype language sentence

$$(20) \qquad CaT^1(CaT^2(\check{P}_2 T^3 T^2) T^1) T^4$$

is obtained by rule (14), a double application of rule (15) and rule (18). It may be interpreted as *Jane[4] persuaded Mary[1] (that) Mary[1] order the pupils[2] (that) pupils[2] work in the garden[3].*

Note 1. The identical indices of the terms point to the identity of the object of causation and the subject of the situation caused.

Note 2. Rules (14) – (18) are but a possible example of the generative rules of the basic genotype language.

By applying the rules of derivation to the basic genotype language we obtain the derived genotype language which includes basic and derived genotype sentences.

In the process of derivation an essential role is played by the combinators or operators on functions.[4]

The combinators may be also called transformational operators.

In the derivations considered below we shall use the following combinators:

B – the compositor of functions. It composes the predicates of the matrix and the embedded sentences (Cp. " predicate – raising " in the transformational grammar).

The combinator *B* as well as the other combinators may be applied both to basic and derived genotype sentences according to the rule

$$(21) \qquad\qquad X(YZ) \to BXYZ$$

where X is the predicate of the matrix sentence, Y is the predicate of the embedded sentence, and BXY is a complex predicate with Z as its argument.

For example

$$(22) \qquad\qquad Md(\overset{\smile}{P}T) \to BMd\overset{\smile}{P}T$$

illustrated in the hybrid language as

>   *Perhaps (come John)* → *B may come John*

and interpreted as

>   *Perhaps, John will come* → *John may come*

The compositor *B* is also used for simulating[5] such transformations as

>   *I expect John will come* → *I expect John to come*

---

[4] On the combinators see H. B. CURRY (1958), p. 153; H. B. CURRY (1963), p. 118. See also S. K. ŠAUMJAN (1971).

[5] The role of combinators in simulating transformations and obtaining semantic fields is discussed in S. K. ŠAUMJAN and P. A. SOBOLEVA (1973).

L – the confluentor of functions with identical arguments:

(23) $XZ \ (YZ) \ \to \ LXYZ$

where X is the predicate of the matrix, Y – of the embedded sentence, and Z is the identical argument of both sentences, for example:

(24) $CaT^1(\breve{P}_1 T^1)T^2 \ \to \ LCa\breve{P}_1 T^1 T^2$

which simulates the transformation of causation:

> ★ *Mary persuaded John, John came* →
> *Mary persuaded John to come*

The combinator L fulfils the function, analogical to " predicate-raising ", " subject-raising ", and " equi-*NP*-deletion ", applied simultaneously.

W – the duplicator, which, applied to a function with two identical arguments, deletes one of them, turning a two-place function into a one-place function:

(25) $XYY \ \to \ WXY$

Rule (25) is used in the process of deriving genotype analogues of subjective infinitives, for example

(26) $Vl(\breve{P}T^1)T^1 \ \to \ BVl\breve{P}T^1 T^1 \ \to \ W(B \ VL \ \breve{P})T^1$

illustrated in the hybrid language as

> *like (sing John¹) John¹* → *B like sing John¹ John¹* →
> *W(B like sing) John¹*

and interpreted as

> ★ *John likes, John sings* → ★ *John likes John to sing* →
> *John likes to sing*

Cv – the converter of arguments:

(27) $XZY \ \to \ CvXYZ$

where X is the predicate, Y and Z are its arguments. The derivation with the help of Cv is analogous to the transformations " flip " (" psych-movement ") and passivization (see foot-note 5).

K_i – the operator of introducing a dummy argument:

$$(28) \qquad VXY \to \begin{cases} K_1 VZXY \\ K_2 VXZY \\ K_3 VXYZ \end{cases}$$

where V is the predicate, X and Y are its arguments and Z is a dummy argument introduced either immediately after V (if K_1 is applied), or after X or Y (if K_2 and K_3 are applied correspondingly), for example

$$(29) \qquad Vl(\check{P}T^1)T^2 \to K_1 \ VL \ T_\emptyset(\check{P}T^1)T^2$$

which may be interpreted as

$$* \ I \ like, \ John \ sings \to I \ like \ it \ (that) \ John \ sings$$

The operator K may be used to model the derivation of various it-constructions in English. The combination of K_i and Cv is similar in effect to extraposition.

U – the operator which exchanges the roles of the function and its argument:

$$(30) \qquad \qquad XY \to UYX$$

The application of this operator will be illustrated later. The combinator U is used in the process of the derivation of the genotype analogues of verbal nouns, gerundial phrases and cleft sentences.

H – the operator which, applied to a predicate of P_i ($i = 1, 2, 3$) type changes it into a predicate of $\check{P}$ type by nominalizing an embedded sentence, for example

$$(31) \qquad \qquad P\check{S}T \to H \ \check{P}(R_T \ \check{S})T$$

which may be interpreted e.g. as

$$I \ believe \ John \ will \ come \to I \ believe \ that \ John \ will \ come$$

The operator R_T is a nominalizer belonging to the category $\Delta\beta\alpha$ and interpreted as $that$ or a complex of morphemes $prep...'s...-ing$, for example

(32) $Em \ (\check{P}T^1 T^2)T^3 \rightarrow HEm(R_T \ (\check{P}T^1 T^2))T^3$

which may be given an interpretation of

> * *John is amused, Harry plays the piano* →
> *John is amused at Harry's playing the piano*[6]

Several more combinators are used in applicative generative grammar but we shall not dwell upon them here. For lack of space we shall also omit their categorial status.

Besides the combinatory derivation rules there are rules of a different kind, such as relativization, pronominalization and some others. Here we shall give the rule of relativization:

(33) $Cj \ (XY) \ (ZY) \rightarrow Z(R_A(XY)Y)$

Where X is the predicate of the first conjoint sentence, Z – the predicate of the second conjoint sentence and Y is the identical argument of both X and Z. R_A is the operator of relativization, a semion belonging to the category $\Delta\beta\Delta\alpha\alpha$. The full notation of the relativizer is $\Delta\beta\Delta\alpha\alpha$ R_A. The notation shows that R_A is the operator which applied to a sentence, maps it into a term identifier. The interpretation of (33) is ungrammatical for English and further derivation rules are necessary to obtain the analogues of complex sentences with relative clauses or with participial phrases.

Now we have sufficient means at our disposal to obtain the genotype analogues of (12) and (13) as well as of other complex constructions with all kind of complements.

The analogue of (12) is obtained from the basic genotype sentence $Cj \ (\check{P}^1 T^1) \ (\check{P}^2 T^1 T^2)$ by applying the rule of relativization and two combinatory rules with the compositor B and duplicator W. Below the steps of the derivation are given in a column:

(34)
> 1. $Cj \ (\check{P}^1 T^1) \ (\check{P}^2 T^1 T^2)$
> 2. $\check{P}^2(R_A(\check{P}^1 T^1)T^1)T^2$
> 3. $\check{P}^2(BR_A\check{P}^1 T^1 T^1)T^2$
> 4. $\check{P}^2(W(BR_A\check{P}^1)T^1)T^2$

---

[6] A different way of deriving gerundial structures with the help of the combinator U is discussed in S. K. ŠAUMJAN and P. A. SOBOLEVA (1973).

For the sake of clarity, derivation (34) may be repeated in the terms of the so-called hybrid language, i.e. a language the grammatical component of which is formed by applicative grammar and the lexical component – by the lexical items of a concrete natural language, such as English, for example:

(35)
1. Cj *(burn house) (see house I)*
2. *see(*R_A*(burn house) house) I*
3. *see(*BR_A *burn house house) I*
4. *see(W(*BR_A *burn) house) I*

The basic structure in (34) and (35) is interpreted as a semigrammatical sentence

  ★ *I see a house and the house is burning.*

The second line cannot be given a grammatical interpretation, it roughly corresponds to a hypothetical sentence

  ★ *I see the house is burning, a house*

The third line is obtained as a result of the application of the compositor B to the operator R_A and after that of BR_A to P^1, yielding a complex function BR_AP^1. This complex function has two identical arguments T^1 and T^1. If the first argument is pronominalized (we omit the rule here) then the third line can be given a grammatical interpretation:

  I see a house which is burning

The fourth step of the derivation consists in the application of the duplicator W which eliminates one of the identical terms, thus yielding a genotype analogue of (12), i.e.

  I see a burning house.

The derived structure (13) is obtained also in four steps, the basic sentence being $Mt(P^1T^1)T^2$. See the derivation below:

(36)
1. $Mt(\check{P}^1T^1)T^2$
2. $HMt(R_T(\check{P}^1T^1))T^2$
3. $HMt(BR\check{P}^1T^1)T^2$
4. $HMt(UT^1BR_T\check{P})T^2$

In terms of the hybrid language (36) is presented as follows:

(37)
1. *expect (arrive John) I*
2. *H expect (R_T(arrive John)) I*
3. *H expect (BR_T arrive John) I*
4. *H expect (U John BR_T arrive) I*

R_T is the nominalizing operator with the episemion $\Delta\beta\alpha$ i.e. R_T is the operator which, being applied to a sentence, maps it into a term. The categorial status of H here is $\Delta\Delta\beta\Delta\alpha\beta\Delta\alpha\Delta\alpha\beta$ i.e. H is the operator which changes a two-place predicate with an embedded sentence and a term into a two-place predicate with two terms as its arguments.

The first line of derivation (37) is interpreted as

I expect John is arriving

The second line may correspond to a sentence of the type

I expect that John is arriving.

The third line cannot be given a grammatical interpretation. Here the combinator B composes the functions R_T and P producing a composite function BR_TP roughly corresponding to a verbal noun of the *arrival* type. To make the interpretation grammatical the combinator U is applied at the fourth step, turning T into a function and BR_TP into its argument. The fourth line may be interpreted then as (13) i.e. the sentence *I expect John's arrival.*

If we now compare the two possible ways of obtaining the genotype analogues of (12) and (13) we shall see that applicative grammar provides us with the means of simulating natural languages both at the surface level – taxonomic description – and at the deep level. The application of derivation rules to the sentences of the basic genotype language simulates the process of transforming the deep semantic representation of sentences into their surface images.

We shall now attempt to show how the application of different sets of derivation rules to one and the same basic genotype sentence generates a bundle of the derived genotype sentences which may be interpreted as a semantic field of phrases. Let us take several sentences with the emotive verb *please* and obtain their genotype analogues. The sentences are:

(38)   *Mary is pleased that she found the dictionary*
(39)   ? *That she found the dictionary pleased Mary*

(40)   *It pleased Mary that she found the dictionary*
(41)   *Mary is pleased to find the dictionary*
(42)   *It pleased Mary to find the dictionary*

The genotype analogues of (38) – (42) are obtained from the basic genotype sentence

(43) $Em \; (\check{P}T^1 T^2)T^2$

as a result of applying the derivation rules with the combinators discussed above, their choice and order differing in each particular case.

The genotype analogue of (38)[7] is obtained as a result of the application of derivation rule (31). See (44) (below the derivations are given in columns, the first line being a basic sentence i.e. the sentence of the zero degree of derivation).

(44)   1. $Em(\check{P}T^1 T^2)T^2$
            2. $HEm(R_T(\check{P}T^1 T^2)T^2)$

The genotype analogue of (39) is obtained as a result of the applications of rules (31) and (27). See (45)

(45)   1. $Em(\check{P}T^1 T^2)T^2$
            2. $HEm(R_T(\check{P}T^1 T^2))T^2$
            3. $Cv_{12}(HEm)T^2(R_T(\check{P}T^1 T^2))$

In derivation (45) the combinator of conversion Cv is applied to the result of derivation (44) which changes the places of the term and the embedded sentence, yielding the derived structure with a nominalized embedded sentence as its final argument (interpreted as the subject) (Cp. " Flip " or " psych-movement " in transformational grammar).

The genotype analogue of (40) is obtained as a result of four rules – (31), (28), (27). See (46)

(46)   1. $Em(\check{P}T^1 T^2)T^2$
            2. $HEm(R_T(\check{P}T^1 T^2))T^2$
            3. $K(HEm)T_\emptyset(R_T(\check{P}T^1 T^2))T^2$
            4. $Cv_{13}(K(HEm))T^2(R_T(\check{P}T^1 T^2)T^2)T_\emptyset$

---

[7] We do not touch upon the problem of pronominalization. This problem is discussed in the papers S. K. ŠAUMJAN (1971) and S. K. ŠAUMJAN, P. A. SOBOLEVA (1973).

We shall illustrate derivation (46) in the hybrid language. See (47):

(47)
1. *please (find dictionary Mary) Mary*
2. *H please (that (find dictionary Mary)) Mary*
3. *K(H please) it (that (find dictionary Mary)) Mary*
4. *Cv_{13}(K(H please))Mary (that (find dictionary Mary)) it*

The first line i.e. the basic sentence, is interpreted as

★ *Mary is pleased Mary (she) found the dictionary*

The second line is interpreted as (38).

The third line, obtained after the introduction of the dummy argument *it* is interpreted as an ungrammatical sentence

★ *Mary is pleased (about) it that she found the dictionary·*

The fourth line, in which the combinator Cv_{13} changes the places of the first and the third arguments gives the analogues of (40) (Cp " extraposition " and " flip ").

The genotype analogue of (41) is obtained as a result of two rules – (21) and (25). See (48)

(48)
1. $Em(\check{P}T^1 T^2)T^2$
2. $BEm(\check{P}T^1)T^2T^2$
3. $W ((BEm) (\check{P}T^1))T^2$

Derivation (48) is illustrated in the hybrid language. See (49):

(49)
1. *please (find dictionary Mary)Mary*
2. *B please (find dictionary) Mary Mary*
3. *W((B please) (find dictionary) Mary*

In the second line the combinator *B* composes the predicates *Em-please* and $\check{P}T^1$-*find dictionary*. The complex predicate *BEm ($\check{P}T^1$)* is interpreted as an infinitive complex *is pleased to find the dictionary* (Cp. " predicate raising "). In the third line the combinator *W* deletes the identical terms T^2 *Mary*, yielding (41).

The analogue of (42) is obtained in five steps: to derivation (48) two more lines are added (rules (28) and (27)). See (50):

(50)
4. $K(W((BEm)\ (\overset{\smile}{P}T^1)))T_{\emptyset}T^2$

5. $Cv_{23}(K(W((BEm)\ (\overset{\smile}{P}T^1))))T^2T_{\emptyset}$

The set of derivations from the basic sentence (43) may be shown in the tree below. See (51):

(51)
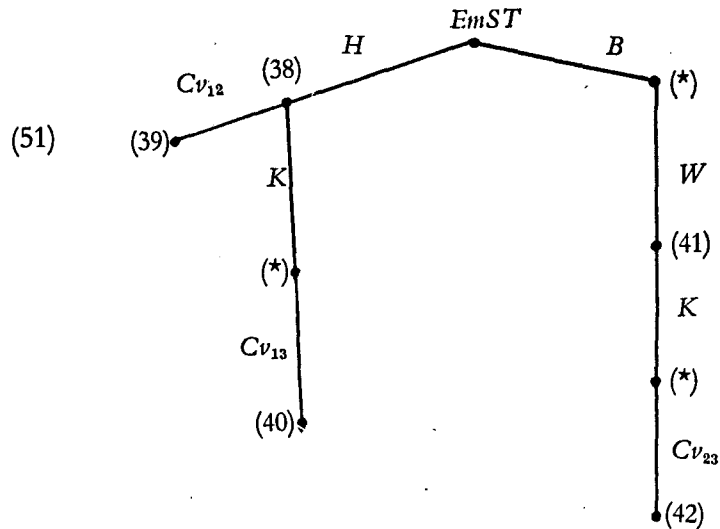
At the root of the tree we have a basic genotype sentence. The lines indicate the combinators which are applied to the basic or derived sentences in our derivations. The numbers (38) – (42) refer to the sentences which interpret the derived genotype sentences, (*) denotes the absence of a grammatical interpretation in English. Similar trees of derivation may be drawn for other basic sentences of the genotype language.

Examples of such trees for the basic sentences $Md\check{S}$, $As\check{S}$, $Mt\check{S}T$, $Vl\check{S}T$, $CaT^1\check{S}T^2$ and $CmT^1\check{S}T^2$ may be found in the paper of S. K. ŠAUMJAN and P. A. SOBOLEVA (1973).

The basic genotype sentence at the root of the tree may be understood as a generalized semantic representation, which we call a standard macrosituation. The derived genotype sentences at the other nodes are semantic representations called derived macrosituations. Interpretations (38) – (42) are concrete semantic representations called microsituations.

The tree in (51) may be regarded as a possible scheme of derivation

of the emotive semantic field in English mapped into the genotype language.

The purpose of (34) – (49) was to show that the combinators may be used to form a unified system of derivations necessary for the construction of the derived genotype language. In (38) – (49) the genotype sentences were derived from one and the same basic genotype sentence, differing from each other only in the choice, number and order of application of the combinators. Hence, the possibility of simulating the hierarchical organisation of grammatical synonymy as well as the process of embodying thought in linguistic forms.

Let us now turn to the definition of formal theories of natural languages.

The formal theory of natural languages is a formal system which distinguishes between the following components:

1. The basic genotype language is defined by the schemes of sentences serving as semantic representations. Each semantic representation corresponds to a definite situation. Semantic representations are formed from elementary semions. Elementary semions are interpreted as elementary meaning units. Only two types of elementary semions are allowed – elementary predicates and elementary terms – and two types of sentences: 1) sentences with one-, two- and three-place elementary predicates and an embedded sentence as one of their arguments, 2) sentences which consist of sentences connected by means of a conjunctor.

2. The rules of derivation of the derived genotype language sentences from the sentences of the basic genotype language. The sentences of the basic genotype language are regarded as the sentences of the zero degree of derivation, hence the basic genotype language is regarded as part of the derived genotype language which is a sum of the zero degree of derivation sentences. Every rule of derivation of the derived genotype language sentences is a substitution of some expression containing combinators applied to the elementary predicates instead of some expression of the basic genotype language. The rules of derivation of the derived genotype language may be called combinatory rules because they consist in the substitution of the expressions with the various types of combinators applied to elementary predicates in place of the basic genotype language expressions.

3. The rules of derivation of phenotype language sentences from sentences of the derived genotype language. These rules may be called phenotype rules. Every phenotype rule is a substitution of: a) a certain

6

expression comprising phenotype semions into a certain expression of
the derived genotype language. Or: b) into a certain phrase obtained
according to (a). The phenotype rules meet the requirement according
to which any application of the phenotype rule must not break the
applicative structure of the sentence. Sentences obtained as a result of
the application of the phenotype rules are called phenotype sentences.

4. The rules of linear transformations.

In the genotype language expressions have a standard linear structure:
an operator precedes its operand. This standard structure is retained by
the phenotype sentences. The aim of linear transformations is to con-
vert the standard structure into a linear structure characteristic of the
sentences of the corresponding natural language.

The rules of linear transformations are based on the use of two com-
binators of permutations. The use of the permutation combinators is
interesting from the typological point of view, since the standard linear
structure may be regarded as a typological standard, the derivation
from which may be accounted for by different sets of permutation
combinators.

5. Morpho – phonological rules.

Morpho – phonological rules consist in the substitution of either
phonological strings instead of phenotype expressions, or of some pho-
nological strings instead of other phonological strings.

In accordance with the components of the formal theory of natural
languages considered above the scheme of deduction in this theory
looks as follows:

$$G_1, \ ..., \ G_n F_1, \ ..., \ F_m L_1, \ ..., \ L_p \Phi_1, \ ..., \ \Phi_q$$

G_1 is a sentence of the basic genotype language. G_n is a sentence of the
derived genotype language. In case $n = 1$, G_n is a sentence of the zero
degree of derivation. F_1 is the initial sentence of the phenotype compo-
nent. F_m is the final sentence of the phenotype component. L_1 is a
sentence obtained at the first step of the application of the linear transfor-
mational rules. L_p is a sentence obtained at the final step of the linear
transformation rules.

Φ_1 a morpho-phonological representation of the sentence obtained
at the first step and Φ_q – at the last step of the application of the morpho-
phonological rules.

The formal theory of natural languages distinguishes between the
concept " macrosituation " and the concept " microsituation ". A mac-

rosituation is an abstract situation at the level of generalized semantic representation. The microsituations pertaining to the given macrosituations are its various interpretations at the level of concrete semantic representations. An example of an " emotive " macrosituation: " An animate passive subject (patient) experiences an emotion in connection with the action it performed ". This macrosituation may be realized in the following microsituations: *Mary is pleased that she found the dictionary, John is upset that he lost the pen, Jane is confused that she upset the cup* etc. An example of a causative macrosituation: " An animate active subject (agent) causes an animate object (patient) to perform an action ". This macrosituation may be realized in the following microsituations: *The officer ordered the men to open fire, The mother persuaded her son to take a walk, The teacher makes the pupils work* etc.

In connection with the distinction between macro and microsituations we distinguish between two levels of derivation in the formal theory of natural languages: derivations at the level of macrosituations and derivations at the level of microsituations. As to the formal theory of natural languages it must provide for the rules of derivation at the level of macrosituations. The level of microsituation is an interpretative level which is a function of a special component which may be called the dictionary of concrete lexemes. The dictionary of concrete lexemes consists of dictionary entries corresponding to separate concrete lexemes.

Every dictionary entry must contain the necessary grammatical information providing for its inclusion as a concrete semion interpretant of the macrosituation in a derivation set.

In conclusion we shall consider the concept " semantic fields ". In connection with the semantic field we shall split the concept " macrosituation " into the concept " standard macrosituation " and " derived macrosituation ". The standard macrosituation is a semantic representation corresponding to the sentences of the basic genotype language. For example $Em(\breve{P}T^1\ T^2)T^2$ which roughly corresponds to the " emotive situation " cited above or $CaT^1\ (\breve{P}\ T^2\ T^1)\ T^3$ which is very close to the above causative situation. However, the latter can be given a more exact interpretation: " an animate subject (agent) brings pressure to bear on an animate object (patient) and the latter performs an action ".

The derived macrosituation is a semantic representation corresponding to the sentences of the derived genotype language. For example $H\,Em\,(R_T(\breve{P}\ T^1\,T^2))T^2$ which exactly corresponds to the above " emotive " situation of $L\ Ca(\breve{P}T^2)T^1\,T^3$ which exactly corresponds to the

"causative" situation in its previous formulation, i.e. "an animate active subject causes an animate passive object to perform an action".

In the formal theory of natural languages considered above the semantic field is a bundle of derivations simulating a set of derived macrosituations from a basic genotype language sentence which simulates a standard macrosituation. See (51). The standard and the derived macrosituations are in relations of meaning inclusion.

In accordance with the destination between macro- and microsituations we distinguish between the semantic fields at the level of macrosituations and the semantic fields at the level of microsituations: macrosemantic fields and microsemantic fields. Each macrosemantic field corresponds to a finite set of microsemantic fields which are its interpretations.

On the basis of what has been considered above it is possible to draw the following general conclusions pertaining to any generative grammar.

The term "generative grammar" must be understood in two senses: as the grammar of a formal metalanguage used by the formal theory of natural languages and as the formal theory of natural languages proper.

A formal theory of natural languages must be based on the two-level principle which distinguishes between the two levels of abstraction: the phenotype and genotype levels. The genotype level is not given in immediate observation but is postulated as a hypothetical object which is a theoretical construct.

It is necessary to distinguish between the two levels of formal derivation: the level of macrosituations and the level of microsituations. Grammar as a formal theory of natural languages has to do with macrosituations. As to the level of microsituations it has to do with a special dictionary which is compiled for the needs of the generative grammar.

REFERENCES

H. B. Curry, *Combinatory logic*, vol. 1, Amsterdam, 1958.

H. B. Curry, *Foundations of mathematical logic*, New York, 1963.

S. K. Šaumjan, *Genotipičeskij jasyk i formalnaja semantika*, in « Problemy structurnoj linguistiki », 1971.

S. K. Šaumjan, P. A. Soboleva, *Komplementacija v applicativnoj grammatike*, in « Problemy structurnoj linguistiki », 1973.