PHRASE STRUCTURE ANALYSIS AND SAUMYAN'S MODEL

EUGENE VERSTRAELEN
San Carlos University
(Received October 11, 1971)

1. INTRODUCTORY REMARKS

A Chomskyan analysis of language presupposes a Phrase Structure Analysis, developed by means of the so-called rewrite rules, to which the necessary transformation rules are applied.[1] If we wish to make Saumyan's model[2] comparable with traditional Phrase Structure Analysis, then some modifications in the latter are advisable. But once these modifications are made, then even an isomorphic one-to-one mapping of Saumyan's model on the Phrase Structure Analysis appears to be possible.

Two kinds of modification are treated in the following two paragraphs, viz. (a) the marker-modification and (b) the order-modification. A third modification will be dealt with in paragraph 5.

2. MARKER-MODIFICATION

In the traditional Phrase Structure Analysis, we may meet more or less the following rules:[3]

Sentence > *Noun Phrase* + *Verb Phrase*

Verb Phrase > *Verb* + *Noun Phrase*

Here '*Noun Phrase*' occurs twice but nevertheless there is a functional difference between the first and the second *Noun Phrase.* The difference is acknowledged by traditional grammarians by introducing the term '*Phrase Marker.*' Though these grammarians describe the difference between the two *Noun Phrases* satisfactorily, an exact definition of that item which constitutes the phrase marker is not given. I propose to define phrase marker as *the special functional combination with other linguistic forms in which the respective form is used.* Thus the first *Noun Phrase* has the phrase marker; it is used in combination with a *Verb Phrase.* The second *Noun Phrase* has the phrase marker; it is used as a part of a *Verb Phrase* (or, according to a rewrite rule not mentioned here, it is used in combination with a *Verb*).

However, if a certain form with a special phrase marker constitutes a functional feature in a language, then this should be labeled as such with its own name. E.g. the first

[1]The reader is here assumed to be acquainted with transformational grammar. Only those features of transformational grammar are treated which bear a direct relation to our topic.

[2]S. K. Saumyan: "The Applicational Generative Model," *Foundations of Language,* 1965, pp. 189–222; S. K. Saumyan and P. A. Soboleva: "Transformation Calculus as a Tool of Semantic Study of Natural Languages," *Foundations of Language,* 1965, pp. 290–336.

[3]These rules are not meant to be completely exact. They only serve to exemplify special features essential to our topic.

Noun Phrase could be labeled: *'Noun Phrase / Subject,'* or simply *'Subject Phrase'* (while another rewrite rule will indicate that *Subject Phrase* is realized as Noun Phrases or special forms of Pronouns in this special combination). Or if it is more convenient that *Subject Phrase* stand for forms which, abstracted from this concrete combination, can be used in different other combinations, then the label should be something like *Subject Phrase /'Subject /.'* to indicate this individual combination or phrase marker.

A *Noun Phrase* as such is never used in English; a *Noun Phrase* is always used with a phrase marker, i.e. in a concrete combination. *Noun Phrase* as such abstracts from a concrete combination. This abstract form is useful for language analysis, and I will return to it in paragraph 7. But the concrete form consisting of *'Noun Phrase* + this combination' is also of great importance and its treatment is not explicit but implicit and indirect only in the traditional *Phrase Structure Analysis.* These actual forms–as they are combined with other actual forms–constitute the interpretation of Saumyan's mathematical model.

3. ORDER-MODIFICATION

Let us return to the same rewrite rules mentioned in the preceding paragraph.

In the first rule it is indicated that the successive order is: first *Noun Phrase* and then *Verb Phrase.* We can give this successive order a kind of phonemic value. Just as the segmental phonemes are used to constitute morphs, in the same manner the successive order is one of the phonemic values for constituting morphemic forms. In the preceding paragraph we have agreed that *Subject Phrase / Subject /* is a morphemic form, realized by phonemes; in English one of the phonemic features constituting *Subject Phrase / Subject /* is the successive order. In traditional *Phrase Structure Analysis* the form *Noun Phrase* abstracts from the individual segmental phonemes; *Subject Phrase / Subject /* abstracts from any phonemic feature, including successive order. *Subject Phrase / Subject /* is a purely morphemic symbol which abstracts from all sub-morphemic features.

Let us agree that the plus-symbol: + indicates the successive order, and the dash-symbol: – abstracts from the successive order. The marker – and order-modifications modify the traditional rewrite rule:

Sentence > Noun Phrase + Verb Phrase

into:

Predication > Subject Phrase / Subject / –Predicate Phrase / Predicate /

4. SAUMYAN'S MATHEMATICAL SYSTEM

Saumyan's model consists of a specific interpretation of an abstract mathematical system.

I will first give a short, basic description of the mathematical system:

a) Let α, β and any combination of α and β divided by a dot be sets (called episemia). $\alpha\beta$ is by definition: $\alpha . \beta$.

Thus: $\alpha\beta$

$\alpha . \alpha\beta$

$\alpha\alpha . \beta$

$\alpha . \beta\alpha$ etc.

are episemia.

b) The symbol preceding the dot (and α in case of $\alpha\beta$) is called argument.

Thus: $\alpha\beta$

$\alpha . \alpha \beta$
$\alpha . \beta \alpha$
have all α as argument.

And: $\alpha \beta . \alpha$
$\alpha \beta . \alpha \beta$
have $\alpha \beta$ as argument.

c) The symbol following the dot (and β in case of $\alpha \beta$) is called value.

Thus: $\alpha \beta$
$\alpha \alpha . \beta$
$\alpha \beta . \beta$
have β as value.

d) Let 'α'
'β'
'$\alpha \beta$' or
'$\alpha . \alpha \beta$' etc.
be a member (called semion) of the respective
α
β
$\alpha \beta$
$\alpha . \alpha \beta$ etc.

e) The argument of an episemion indicates that its semion should be combined with a semion of the episemion indicated by the argument.

Thus the argument α of the episemion $\alpha \beta$ indicates that the semion '$\alpha \beta$' should be combined with semion of the episemion α, thus with, e.g. 'α'. The resulting combination is 'α' '$\alpha \beta$' (or: '$\alpha \beta$' 'α'; the successive order of the combination is not considered essential). The '$\alpha \beta$' in '$\alpha \beta$' 'α' is called operator, the 'α' in '$\alpha \beta$' 'α' is called operand.

Here follows another example of combination (called: application):

$\alpha \beta . \alpha$ has the argument $\alpha \beta$. Thus '$\alpha \beta . \alpha$' can be applied to a member of $\alpha \beta$ e.g. '$\alpha \beta$' resulting in '$\alpha \beta . \alpha$' '$\alpha \beta$'.

f) The value of an episemion indicates that a result of the respective application is a semion of the episemion indicated by the value.

Thus the value β of the episemion $\alpha \beta$ indicates that a resulting application, e.g. 'α' . '$\alpha \beta$', is a semion of β. The value α of the episemion $\alpha \beta . \alpha$ indicates that a resulting application, e.g. '$\alpha \beta . \alpha$' '$\alpha \beta$', is a semion of α. Therefore '$\alpha . \alpha \beta$', being a semion of $\alpha . \alpha \beta$ with argument α, can be applied to '$\alpha \beta . \alpha$' '$\alpha \beta$', being a semion of α, resulting in: '$\alpha . \alpha \beta$' '$\alpha \beta . \alpha$' '$\alpha \beta$'.

5. SAUMYAN'S MODEL

Saumyan gives an interpretation of the mathematical system; viz. β = *Predication*, α = *Subject Phrase / Subject /*. If we operate with the applications developed in the preceding paragraph, we obtain an interpreted system on which the modified *Phrase Structure Analysis* can be mapped by an isomorphic perfect one-to-one relation. E.g.

Predication > Subject Phrase / Subject /–Predicate Phrase / Predicate /

$$\beta > \alpha - \alpha \beta$$

For *Predication* indicates a set and can be mapped on β interpreted as the same set.

Subject Phrase / Subject / can be mapped on α.

$\alpha\beta$ is a set of which the argument shows that its member should be combined with α (=*Subject Phrase / Subject /*) and the resulting application is a member of β (=*Predication*). Thus $\alpha\beta$ corresponds to *Predicate Phrase / Predicate /*.

$>$ in both cases are symbols indicating that what follows is a subset of *Predication* or β. In the traditional *Phrase Structure Analysis* the subset is probably understood to be improper (perhaps on account of the actual English structure). In Saumyan's model the subset may be proper, and this agrees better with the general abstract analysis applicable to any language. E.g. in the Cebuano Bisaya language, *Predication* is not necessarily realized as consisting of a combination of *Subject Phrase / Subject /* and *Predicate Phrase / Predicate /*. *'ambut* ('I do not know') is a member of the set *Predication* (i.e. is a 'β' belonging to β) which is not morphemically structured consisting of *Subject Phrase / Subject /* and *Predicate Phrase / Predicate /* (i.e. it is not 'α' '$\alpha\beta$'). Therefore, in the modified *Phrase Structure Analysis* the arrow is followed by a subset which may be proper or improper.

Here follows another example of the modified *Phrase Structure Analysis* and Saumyan's Model:

Predicate Phrase / Predicate / > modified finite Verb–finite Verb modifier

$$\alpha\beta > \alpha\beta—\alpha\beta \cdot \alpha\beta$$

Again it should be stressed that the application '$\alpha\beta$' '$\alpha\beta \cdot \alpha\beta$' is a member of the proper or improper subset of $\alpha\beta$.

In '$\alpha\beta$' '$\alpha\beta \cdot \alpha\beta$' the form '$\alpha\beta \cdot \alpha\beta$' is the operator and '$\alpha\beta$' the operand belonging, respectively, to $\alpha\beta \cdot \alpha\beta$ and $\alpha\beta$. The argument in $\alpha\beta \cdot \alpha\beta$ shows that its member '$\alpha\beta \cdot \alpha\beta$' should be applied to '$\alpha\beta$' (to a *Predicate Phrase / Predicate /*). The value in $\alpha\beta \cdot \alpha\beta$ shows that the respective application, viz. '$\alpha\beta$' '$\alpha\beta \cdot \alpha\beta$' is a member of $\alpha\beta$. Here follow some concrete interpretations:

'$\alpha\beta \cdot \alpha\beta$' = 'fast' or 'apples'
'$\alpha\beta$' = 'eats'

The application '$\alpha\beta$' '$\alpha\beta \cdot \alpha\beta$' = 'eats fast' or 'eats apples'. This application is a member of $\alpha\beta$, and is applicable to 'α'. 'α' '$\alpha\beta$' '$\alpha\beta \cdot \alpha\beta$' = 'he eats fast' or 'he eats apples'. In this example we see that in this concrete combination 'fast' and 'apples' are morphemically the same (cf. paragraph 7).

6. THE ONE-TO-ONE MAPPING

The isomorphic mapping with the modified *Phrase Structure Analysis* must be explained in greater detail.

The modified *Phrase Structure Analysis* consists of the following:

a) At the left side of the arrow a set is indicated.
b) At the right side, components are indicated which are essentially related to each other (according to the marker-modification) and which together constitute a subset of the set indicated at the left side.

In Saumyan's model the components are also essentially related according to the argument of one of the components. The components constitute a subclass as indicated by the value of one of the components. E.g. in:

$\alpha - \alpha\beta$

the argument in $\alpha\beta$ relates it essentially to α; the value in $\alpha\beta$ indicates the class β to which the application of the respective semia belongs.

No other operations are defined in the system of the modified *Phrase Structure Analysis;* neither in Saumyan's model.

This shows that there is a perfect one-to-one mapping between the two systems.

In this paragraph, however, Saumyan's symbol-system is used in a manner which does not appear in Saumyan's papers. This is done purposely in order to show more clearly the isomorphism between the modified Phrase Structure Analysis and Saumyan's model.

E.g. Saumyan uses neither the arrow: $>$ nor the whole line:

$\beta > \alpha - \alpha\beta$

Saumyan would explain this formula more or less in the following way:

'α' '$\alpha\beta$' means: application of two semia which are members, respectively, of the episemia α and $\alpha\beta$, which application is included in the episemion β. Saumyan's symbolism is more succinct. My way of using his symbols is unnecessarily circumstantial, but it more clearly shows the isomorphic relation.

Because the isomorphic one-to-one mapping constitutes a crucial point in this paper, some additional clarifications may be welcome.

How should a formula of the modified Phrase Structure Analysis be translated into a formula of Saumyan's model?

'Predication' (or *'Sentence'* in the terminology of Chomsky) is translated into β.

'Subject Phrase /S/' into α.

The first rewrite rule has *'Predication'* or β at the left side of the arrow.

At the right side of the arrow are the eventual component classes (i.e.: immediate constituents). In case of different variants, these variants are submorphemic (cf. paragraph 7).

One component class of β, indicated at the right side of the arrow, is necessarily α, which is interpreted as *'Subject Phrase /S/'.*

The other component class–and also any other component class in the succeeding rewrite rules–is indicated by a composite symbol with its fitting argument and value. In this case by $\alpha\beta$ with argument α and value β. This $\alpha\beta$ corresponds to *'Predicate Phrase /P/'.*

In each following rewrite rule a component class taken from the right side of a preceding rewrite rule is at the left side, and its component classes are at the right side. These component classes are indicated again by symbols with fitting arguments and values, which correspond to the more or less arbitrary labels used in the modified *Phrase Structure Analysis.*

From this it is clear that Saumyan's symbol system is in a certain respect more perfect than the *Phrase Structure Analysis.* Any composite symbol indicates clearly, by means of its argument and value, which level of immediate constituent is meant. The arbitrary labels of the *Phrase Structure Analysis* do not convey this information. Therefore we can only find the corresponding symbol for a label by tracing all the preceding rewrite rules. Thus the translation of one system to the other can be done systematically and each label corresponds to a symbol. But if the total translation of all preceding rewrite rules is not yet fully carried out, then the exact identity of an arbitrary label is not clear, while the identity of Saumyan's symbols is always clear by a simple inspection of the argument and value contained in the symbols.

Since the composite symbols function in a special manner on account of their arguments and values, a formula containing composite symbols may not be well-formed, and thus meaningless. E.g. $\beta > \alpha - \alpha\beta$ (or: $\beta > \alpha\beta - \alpha$, because successive order is not essential) is meaningful: the argument α in $\alpha\beta$ indicates that $\alpha - \alpha\beta$ refers to a meaningful application, and the value β in $\alpha\beta$ indicates that at the left side there should be β. Therefore $\beta\alpha > \alpha - \alpha\beta$ is meaningless, because at the left side there is no β, but $\beta\alpha$.

7. SOME REMARKS AND CONCLUSIONS

In paragraph 6 I have shown—perhaps not rigorously, but sufficiently convincingly—that there is a one-to-one mapping between the modified *Phrase Structure Analysis* and Saumyan's Model. This means that all the features deductible from Saumyan's model belong also to the modified *Phrase Structure Analysis.*

Let us call those features which are essential in both systems *morphemic features.* The features from which the two systems abstract are called *submorphemic features.*

An example of a submorphemic feature is: the functional successive order of the different forms (cf. paragraph 3). It is submorphemic as far as it is a kind of phonemic feature which with other phonemic features constitutes the morphemic relationship between the respective forms.

A second example of a submorphemic features is the following:

In the traditional *Phrase Structure Analysis,* both the *Subject Phrase / Subject /* and the *Predicate Phrase / Predicate /* contain a *Noun Phrase*; i.e. different morphemic forms contain an identical sub-form, if abstraction is made from their actual combinations. In English, e.g., this identity of *Noun Phrase* is obtained if abstraction is made from the special successive order in the total sentence. This identity of forms is in Saumyan's Model something submorphemic. Nevertheless it is something essential in Language analysis. (Saumyan, however, has expanded his system by which this defect is probably removed. He has applied this expanded system to Russian, but I do not know enough Russian to be able to criticize his work.)

Saumyan has indicated that morphemic transformations are not necessary for the analysis of language.[4] One can easily prove this thesis once the one-to-one mapping of the two systems is established. In Saumyan's model the ultimate result of an application depends totally on the value of the respective episemion, while the application itself does *not* depend on the value (but on the argument). Thus, also in the modified *Phrase Structure Analysis* the rewrite rules can be formed in such a way that any morphemic result is possible by choosing the proper constituents. Only for the submorphemic features (e.g. concerning the successive order of forms) transformations may remain necessary.

This last conclusion, viz. that any morphemic transformation can be reduced to a rewrite rule, is important for language analysis. The great, unwieldy, mass of transformations, used in transformational grammar, can be done away with, with the result that the grammar becomes neat and clear.

---

[4]Cf. Saumyan: *op. cit.*, p. 205.