

314.417  
7  
968

# COMPUTATIONAL LINGUISTICS

## VII

COMPUTING CENTRE OF THE HUNGARIAN ACADEMY OF SCIENCES  
BUDAPEST, 1968





# COMPUTATIONAL LINGUISTICS

## VII.

COMPUTING CENTRE OF THE HUNGARIAN ACADEMY OF SCIENCES  
BUDAPEST, 1968

Editorial Board

Bálint Dömölki, Ferenc Kiefer (editor), János S. Petőfi, György Szépe,  
János Szelezsán, Éva B. Szöllősy (technical editor), Dénes Varga

Felelős kiadó: Balázs János

682601 MTA KESZ Sokszorosító. F.v.: Szabó Gyula



## Contents

E. Agricola	
Prospects of solution of syntactic and semantic ambiguities	5-19
A. Gladky	
On the syntactic description of sentences (in Russian with an English summary)	21-44
F. Kiefer	
On the problem of word order	45-65
J. Kunze	
The treatment of non-projective structures in the syntactic analysis and synthesis of English and German	67-77
S. J. Petőfi	
Notes on the semantic interpretation of verbal works of art	79-105
D. Varga	
Syntactic analysis on the basis of continuous junction of syntactic units	107-113
REVIEW	
Yu. D. Apresyan	
Idei i metody sovremennoj strukturnoj lingvistiki, Moskva, 1966. p. 300 (P. Szántó) (in Russian)	117-123



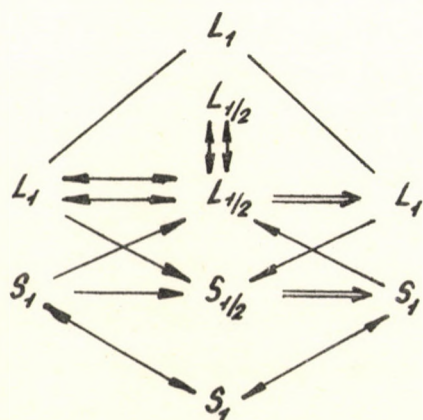


# PROSPECTS OF SOLUTION OF SYNTACTIC AND SEMANTIC AMBIGUITIES

Erhard Agricola

## Introduction

The basic data of any kind of translation analysis of the source language is the surface structure of a sentence in a text, i.e. the string of information-carrying units in a particular sequence, whose inner deeper relation is to be found in the form of the correct actual tree of the explicit syntactic and semantic structures. One of the main obstacles to automatic analysis of random, unprepared texts is the fact that most of these basic data - by their very nature as language - will already present ambiguous elements. The decision about the actual syntactic structure among many potential ones and about the correct semantic variant - and this means at the same time the choice of the appropriate translation equivalents in the target language - is made in ordinary communication by the application of a complex network of contextual information taken from every component of the total meaning of a sentence or an utterance (insofar as no extralinguistic knowledge has to be used). The basic relations between the components are as follows:



$L_1$  = unambiguous lexemes

$L_{1/2}$  = ambiguous lexemes

$S_1$  = unambiguous syntactic elements

$S_{1/2}$  = ambiguous syntactic elements

→ = one-sided relations

↔ = mutual relations

⇔ = multiple relations

⇔⇔ = processes for resolving ambiguities



For the purposes of machine translation a model of these relations must be constructed which is able to account exhaustively for the linguistic information content, right to its objective limits.

### Syntactic ambiguity

The dependency grammars used for analysis so far are in principle capable of assigning to each sentence in a text the complete dependency tree of its word-class elements (dependency relation/DR) and the syntactic values of these relations (dependency type/DT). However, since in German and in English the underlying information, taken in isolation, is ambiguous in 60 to 70 per cent of all cases, a process of syntactic analysis is only fully automatic if the grammar is put in a position to decide on the structure of the lexical content from among many potential trees. To do this the syntactic part of the meaning of the analysed sentence itself must be used first of all as far as possible. The solution of this problem requires the treatment of all possible forms of syntactic ambiguity (polysyntacticity) in common; the consideration of all of these under the same analytical and contextual conditions; the reduction of diverse occurrences to a few general, constant types of causes, of the effects, and of the solution of the ambiguity, of the standardisation and systematisation of the basic types established and finally the investigation of the contextual relations obtaining as a matter of principle between the elements which cause and those which resolve ambiguity.

The ambiguity of the DR and the DT with respect to any two word-forms, and thus to the polysyntacticity of the whole sentence, has, according to the above consideration, the following basic forms:

1. Ambiguous criteria of the linear sequence of words (wordsequential ambiguity WSA) and/or for ambiguous formal syntactic components of the word-forms, sub-categorised according to:



word-form ambiguity with regard to inflexion (Schrauben; seemed) =  
= WA2/3,

with regard to word-class (laut; means) = WA4,

with regard to inflexion and word-class (reifen; saw) = WA5.

These causes of ambiguity result in the following:

(a) more than one DR, but with the same DT = syntactic ambiguity of  
the basic type SAI: basic form 1/2 (= DR)  
A/A (= DT)

Examples: Kurse für Verbandsmitglieder im Ausland, a red number window.

(b) more than one DR, these having different DT = syntactic ambiguity  
of the basic type SAIII: basic form 1/2 (= DR)  
A/B (= DT)

Examples: darin werden während des Transports beschädigte Apparate repariert;  
you don't know how good Calvados tastes.

2. (a) Unambiguous criteria of sequence and of word forms but several  
(only transformationally separable) syntactic values of one and the same DT  
(multiple dependency/MD).

Examples: das Schleifen der Kolbenringe;  
she is ready to kiss.

or (b) ambiguous criteria of sequence (WSA) and/or of word-forms (WA)

Examples: in Kraków kann man sich verlieben,  
twenty Polish students.

or (c) ambiguous criteria of sequence (WSA) combined with multiple dependency  
(MD)

Examples: André verlor alle folgenden Abende;  
How do you feel these cold days?

These causes give rise to only one DR, though this has more than one  
DT = syntactic ambiguity, basic type SAII: basis form 1/1 (= DR)  
A/B (= DT)

Every occurrence of polysyntacticity can be derived from one of these  
constant types but its actual status regarding ambiguity or non-ambiguity is  
relative: how, when and whether the instances of ambiguity are identified as  
such depends on the type, direction and state of the analytic process. Further-



more the degree of reduction or cancellation (and thus also of the effect on the total meaning) of instances identified depends in each case on the presence of a particular set and combination of actualising contextual information (in the wider sense). From the point of view of the degree of interference in the information caused by syntactic ambiguity and to what extent it can be solved, we can recognise about eleven "degrees" of syntactic ambiguities of sentences.

This double relativity, the diversity and possibilities of combination of the sequence of word-classes which potentially give rise to ambiguity have defied all attempts at solution by previous methods, namely the use of tables with "contexts" of the surface structure which were meant to give possible actualisations for some frequent instances of ambiguity. The new system for the classification of the instances of potentially syntactically ambiguous sequences of word-classes uses the following criteria: the three basic types mentioned SA I, II and III; within the basic types a classification according to the type of cause and their combinations; within these types a sequence according to the "level" affected (Phrase, Clause, Sentence and Transitions). According to these criteria more than eighty per cent of constantly recurring instances in German and English can be detected and identified as roughly 100 minimal constant types of characteristic sequences of word classes. Of these 30 to 40 are present in both languages and are in part mutually "translatable."

Example of a type:

(70) Type: SA III-23    Cause: WA4    Translatability: TØ    Level: P(P)

E: (F)	V <sub>7</sub>	N
1	2	3
y	1/3	2/1
	FE/NA	D0/FE

variants: NA = NY;  
D0 = SN

(Avoid infection) by killing germs.

Visiting relatives (can be boring).

Besides these easily classifiable types there also occur in the texts:

(a) Instances of types superimposed on each other within the same word-class sequence, which according to their distance in the system operate as



"type variants" (as in the above example), as "subtypes" (deviations in individual word-class information), and as membership of different types because the cause and the level affected are different. For example for each of the frequent occurrences of the English word-class sequence  $V_{ing} N$  it is possible that it will belong to one or more of ten types.

(b) Instances of connected types, i.e. 2 to 4 types which always occur together, as an underlying reason for polysyntacticity and are mutually determining in its cancellation (example above). The possibilities of connection can likewise be classified into types and are to be understood as a modification of the three basic types SA I, II and III (three basic types and ten modified basic types). They represent the 13 minimal models of syntactically ambiguous structures abstracted from the reason, the word sequences, and the affected level.

Example:

Modified, abstracted basic type SA III.2.1. (= SA III + SA I)

P	Q	R
1	2	3
2/3		1/2
A/B		A/A

Realised in the types:

(76) Zehntausend Görlitzer und polnische Bürger versammelten sich.

(77) Britain has raised tariffs.

(87) ....., dass er diesen Auftrag beschleunigt und grosszügig ausführt.

The analysis must furthermore be able among other things to recognise the following non-typified instances and to allot them places in the system in order to cancel them if need be:

(a) individual instances (occurring relatively infrequently and applying only to a small number of lexical collocations), e.g.: ungebeiztes Holz wird modern; he flew across the sea to land.

(b) free (not connected) combinations of more than one type of ambiguity, e.g.: It is too hot to eat; a more extreme example: People who apply for



marriage licences wearing shorts or pedal pushers will be denied licences  
(Kuno): 12 word forms with ambiguous DT's with a total of 33 variants.

(c) incomplete structures (e.g. ellipses, headlines, omissions in coordinating and comparative constructions) er liebt seinen Vater mehr als seine Mutter; Fliegen nicht nur Plage sondern auch Gefahr; Girl skier safe; Sketch Spy Story Flops.

Between the ambiguous syntactic meanings and those which cancel ambiguity three general relations obtain (solution type/SSA). These can be actualised:

(a) by the sole concrete possibility of a particular word-class variant or of a dependency type situated on the dependency tree above or below unambiguous or actualised structural units (= SSA 1)

(b) by exclusion as far as only one occurrence of a particular element in a structural unit is admissible or obligatory (= SSA 2)

(c) by congruity of case, number and gender forms (= SSA 3).

In addition "semi-lexical" (i.e. syntactically utilisable and fixed lexical) criteria are possible, such as lists of compounds, idioms, particular word sub-classes (= SSA 4).

Examples for particular cancellation types in the context of the ambiguous form widerwillig (adjective as predicate/adverb modifying verb):

Sie nannten den Täter widerwillig (AD/AC; not solved)

Sie nannten der Täter widerwillig und nach langem Zögern (AD+AB; SSA 1)

Sie nannten den Täter widerwillig und asozial (AC + AC; SSA 4)

We recognise a system of indirect relations between the types of cause and the types of solution, which may serve as a basis for analytic operations: the types of ambiguous word-class sequences fall into 13 groups (and some variants), the criterion for the mapping of which is the particular combination of cause elements in the sequence of their occurrence in the analysis. To each of these groups is ascribed one (or more) of the general solution types, which is potentially capable of being used for solution and which has to be sought within the structure of the analysed sentence.



For example:

WA4	WSA	WA 2/3	DR	MD	DT	Ambiguity type	SSA			
							1	2	3	4
1	1	1	1/1	1	A/A	SA 0 (Unambig)				
1	1	1	1/1	2	A/B	SA II 27, 29-34	+			+
						28	+	+		+
1	1	2	1/1	1	A/B	SA II 41		+	+	+
1	2	1	1/1	2	A/B	SA II 35-40	+			+
...	...	...	...	...	....	.....	.....	.....	.....	.....

The search operations are thus reified and at the same time limited because they are performed purposely in a particular part of the dependency tree and according to definite elements, and because all the word-class elements and connections which cause ambiguity, the affected level and the number of variants to be sought, are known.

The different structural variants of an ambiguous word-class sequence are however not usually identified in one analytic step. Therefore a further system of relations and operations between them had to be set up so that at the occurrence of the first potential variant the possibility of an instance of polysyntacticity is considered and all the (partial) dependency trees of the explicit structure corresponding to the ambiguous word-class sequence can be set up, from amongst which the actual one is to be chosen by means of the appropriate solution operation. With the theoretical description of these systems and operations one deficiency in the adequate representation of the syntactic analytic process and thus an important obstacle to its transfer into machine form has been removed. But the fact has thus also become known that a complete syntactic analysis in isolation is impossible because about 60 or 70 per cent of ambiguity affecting information originally can on average be reduced to only 10 to 30 percent by purely syntactic meaning. The other occurrences can only be decided on the basis of semantic values which are unambiguous or whose ambiguity has already been solved. This problem however cannot be solved by the sporadic or ad hoc introduction of supplementary semantic information. Rather the solution presupposes the insight and systematic formulation of a complete system of semantic units and of their underlying interrelations in



this system and in the texts of language in use, and also the relations, which must be regarded as a special sub-group, between the semantic and the syntactic elements of meaning.

### Semantic Ambiguity

A system which is capable of deciding in an analysis on the correct semantic structure of a sentence we shall call an "operative language thesaurus." This is not meant to mean a grouping of synonyms or an ordering according to conceptual units as provided by the usual non-alphabetic dictionaries. The attribute "language" is rather meant to emphasise that linguistically fixed and objectively identifiable units and relations serve as a basis and that the totality of linguistic (syntactic and semantic) components are included. "Operative" means here the purpose of the thesaurus, namely that on the basis of the information represented by the system of its organisation it carries out in the analysis of the information represented by the system of its organisation it carries out in the analysis of texts certain operations of decision. The set of elements i.e. the sememes (semantic variants) of a section of the vocabulary, is already known, and its effect on the relations in the analytic process is observable. The following considerations are meant to provide and discuss an overview of the forms of the relations in the context between semantic units on the one hand and between syntactic and semantic units on the other, and possible feed-back from such evidence of the effectiveness of a thesaurus to its structure. The examination of these questions was provisionally confined to relations within German.

A first step towards systematisation distinguishes those basic forms of semantic relations and operations between units which stand in a specific and direct syntactic relation to each other, from those which have indirect syntactic relations or none at all. To the former belongs the establishment of the compatibility of the sememes of two word-forms (or several, e.g. forms bound obligatorily by prepositions or verb valencies) and proceed either:



without realisation or reduction procedures, i.e. the lexemes are both unambiguous or one or both of them are ambiguous and all sememes of one lexeme are compatible with all those of the other:

$$L_1^a \quad L_1^b \longrightarrow L_1^a + L_1^b : \quad \underline{\text{Schaumwein kaufen, exportieren}}$$

$$L_{1-3}^a \quad L_1^b \longrightarrow L_{1-3}^a + L_1^b : \quad \underline{\text{Wein kaufen, exportieren}}$$

("Weinstöcke/ -trauben/Getränk")

$$L_{1-3}^a \quad L_{1-2}^b \longrightarrow L_{1-3}^a + L_{1-2}^b : \quad \underline{(\text{den}) \text{ Wein behandeln}} \text{ ("einwirken/abhandeln")}$$

or with reduction in the number of sememes to be combined of two word forms, one or both of which are ambiguous:

$$L_{1-3}^a \quad L_1^b \longrightarrow L_{2-3}^a + L_1^b : \quad \underline{\text{Wein darreichen}}$$

$$L_{1-3}^a \quad L_{1-3}^b \longrightarrow L_{1-2}^a + L_2^b \quad L_3^a + L_3^b : \quad \underline{\text{Wein spritzen}} \text{ ("sprudeln/übersprühen/$$

mit Mineralwasser  
versetzen")}

or with the realisation of one of several sememes of the same word-form or one sememe of both word-forms by establishing the unique compatibility of these sememes:

$$L_{1-3}^a \quad L_{1-3}^b \longrightarrow L_1^a + L_2^b : \quad \underline{\text{Wein bauen}} \text{ ("errichten/anpflanzen/vertrauen")}$$

$$L_{1-3}^a \quad L_{1-4}^b \longrightarrow L_2^a + L_1^b : \quad \underline{\text{Wein lesen}} \text{ ("sammeln/Schrift lesen/...")}$$

A further form is the actualisation of each several sememes of two or more word-forms as members of an idiom.

The second kind of relation is the actualisation of one of several sememes of a word-form by establishing the actual one of several potential syntactic relations of this word form to one or more other word-forms. This relation is particularly conspicuous in the form of the connection between sememes and



between the valencies of the verb and actual valency occurrences. The number of verbs which can be rendered unambiguous solely by the actual type of their syntactic relations is not very large, but they are usually words (from the general vocabulary) with a relatively high frequency of occurrence:

$$L_{1-5}^a S_1 \longrightarrow L_1^a + S_1^a; \quad L_{1-5}^a S_1^b \longrightarrow L_2^a + S_1^b \text{ etc:}$$

eine Verordnung ist ergangen/eine Einladung ist (an jemanden) ergangen/er hat sich (im Schatten) ergangen/es ist ihm schlecht ergangen/eine Flut von Schmähungen ist über ihn ergangen. Mostly however the ambiguity is not solved, but merely reduced

$$(L_{1-4}^b S_1^d \longrightarrow L_{1,3}^b + S_1^d). \quad \text{Outside the verbal domain there are not many}$$

types or occurrences of this principle because the main condition for it is not satisfied, namely the occurrence of valency conditioned, obligatory dependent word-forms, and the change of valency requirements. Comparable to this would be the realisation of ambiguous nouns in many cases by the use of the article, the possible differences in meaning in the attributive as opposed to the predicative use of adjectives, and also the solution by already solving the word-class ambiguity (lichten verb/adjective), and lastly ambiguity caused by the coincidence of inflected forms and cancelled by case or number concord:

Boxen in der Strassenbahn sind/ist die neueste Einrichtung.

The third basic form of relation is the realisation of one of several possible correct syntactic structures by establishing the compatibility of a sememe of word-form and a sememe of only one of two other word-forms: this is a question of the possibility of solving semantically a syntactic ambiguity which has not yet been solved syntactically. A sequence of word-classes such as "noun-preposition-noun-preposition-noun" is always syntactically ambiguous as to the subordination of the second prepositional phrase (basic type SA I):

Die Tankstelle mit dem Reklameschild an der Ecke.

When this structure has a certain (other) semantic filling (intra-structural solution) one of the possibilities is chosen on the basis of semantic compatibility:



$$S_{1-2} \quad L_1^a \quad L_1^b \quad L_1^c \longrightarrow S_1^a + L_1^a + L_1^b \quad S_2^a + L_1^a + L_1^c :$$

Die Tankstelle mit dem Reklameschild am Giebel/... am Gründungstag.

On the other hand the original version, ambiguous also semantically, can only be made unambiguous through its relations to sememes outside the ambiguous structural component (extra-structural solution):

Die Tankstelle mit dem Reklameschild an der Ecke des Daches; ...an ihre vorderen Ecke/... an einer einträglichen Ecke.

These considerations would apply also to the semantic reconstruction of coordinated, elliptical and incomplete structures, a largely unexplored field, which we can do no more than mention here.

Establishing the compatibility and the realisation of the sememes of two word-forms which are syntactically indirectly related (remote compatibility) takes place after the compatibility of the sememes of all pairs of syntactically directly related lexemes has been examined, and serves the purpose of bringing out the whole correct and unambiguous meaning of the macrosyntagm. It is in this way that information above all is gained which is needed for the realisation of ambiguous forms for which the immediate syntactic environment is not sufficient:

$L_{1-3}^a + L_{1-2}^b :$       (Der) Wein (wird) behandelt

$\longrightarrow L_{1-3} + L_{1-2} :$     Der Wein wird in diesem Kapitel/nach Anbaugebieten behandelt.

$\longrightarrow L_{1-3} + L_1 :$       Der Wein wird mehrere Jahre/mit aller Sorgfalt) in Eichenfassern behandelt.

But the general and universal uncontradictory nature of the semantic marking of larger syntactic units must also be established.

The basic forms of semantic relations and operations between units which are not related syntactically have other functions besides those already mentioned: they produce "Isotopie" (continuation of the topic), meaning transfer



(correlation or substitution relations) and they give realisation far beyond the sentence boundaries. Between the word-forms of the partners in a sememe there is in this case no syntactic relation of subordination. There arises an additional structure on semantic criteria which is independent of the syntax. Qualitatively also these relations (with the exception of pronominal relations) represent something else. The problem is not to decide about the compatibility of sememes from differing areas of meaning in a given syntactic connection, but to establish the semantic equivalence or similarity of sememes in separate structures.

The most general form is resumption of the topic by pronouns. Many sub-classes of the pronoun apart from their normal syntactic dependency on a word-form, set up a further relation, namely the substitution relation to a second form. This refers by means of grammatical congruence more or less unambiguously to the second form which is to be repeated. Between the sememes of both wordforms, which need not occur in the same sentence, the same decisions as to compatibility are reached which would be required for them in an equivalent direct syntactic relation:

$$L_1^a (L_1^b) L_1^c \longrightarrow L_1^c + L_1^a :$$

die Sagen der Saar von ihren Quellen bis zur Mündung/von ihren Quellen bis zu den literarischen Bearbeitungen.

In the one case ihren is a correlate of Saar, and in the other of Sagen: the decision as to which and as to the actual sememes of Quellen is made on the basis of the remote compatibility of Saar - Quelle - Mündung and Sage - Quelle - Bearbeitung.

The resumption of the topic by synonyms (including repetition of word-forms) is the very form of "Isotopie". The main element of the continuity of a topic is the relation between at least one sememe of a sentence and a sememe of the preceding sentence (or one of several preceding sentences), which expresses itself as proximity or equality of meaning and, if the case arises, serves as the realisation of an ambiguous lexeme. In contrast to the correlation relations of the pronouns, there are no formal morphological indications:



$L_{1-2}^a \quad L_1^b \longrightarrow L_2^a \approx L_1^b :$     Der Düker gleitet in den Fluss. Das Wasser verringert das Gewicht dieser Druckleitung.

On the other hand there are some, albeit very generalised, syntactic conditions which determine and restrict the kind and position of the sememe partners to be related, such as the condition that information can be gained only from elements of identical or syntactically equivalent word-classes and sentence elements, or from certain positions in the structure of the sentence. The most frequent carriers of the relation of "Isotopie" are nouns, both between themselves and in interchange with pronouns, while the other word-classes and other nouns in the same sentence mainly have the function of varying the topic and continuing its context.

"Isotopie" with paraphrasing is a more complex form of the same relation, namely the statement of certain semantic equivalences between the sememes of lexicalised elements and their paraphrases, and between paraphrases themselves. Both these entities and their relations can be regarded as expansion and condensation of meaning.

Expansion process:       $L_1^a \approx E_1^a (L_1^b + L_1^c + \dots);$

Condensation process:     $E_1^a (L_1^b + L_1^c + \dots) \approx L_1^a$

Dichter/Goethe/Dichterfürst = der Genius von Weimar/ der 24-jährige Verfasser des Weimar/der Grösste unserer Literatur/der Weimarer Minister-Dichter.

Syntactically an expansion can, subject to context, type, situation and the necessary balance between economy and redundancy, extend to a whole macrosyntagm of complex structure: for this reason it is often necessary to reduce or transform complex sentences to their basic structures before the actual operation to ascertain equivalence, with which the derivation of basic semantic structures comparable to each other runs parallel.

A conception of the structure of the thesaurus system:

It seems quite possible that a reproduction of the linguistic operations and decisions of the analytic process between elements standing in the



relationships mentioned can be gradually approached. On the basic features of the structure of the system, i.e. on the thesaurus in its narrower sense, its units and information, we are able at present to develop the following conceptions:

The distinct semantic units are the sememes of the lexemes, paralexemes and idiomatic constructions. The sememes are however not primarily understood here to be structured combinations of semes, noemes or other abstract elements, but rather a sememe is characterised by the totality of potential obligatory or optional relations to which (on the basis of its lexeme form) it can enter in a syntactically correct structure. The thesaurus is consequently above all a system of paired, mutual relations between two points, where each point represents different specific combinations of conditions and properties of compatibility.

Thus there is produced an arrangement of relations between combinations of properties which at the same time represent sets of sememes corresponding to these combinations. However, proximity and remoteness in the system do not denote degrees of semantic relatedness but those of similarity regarding the kind of permissible syntactic-semantic combinations and the kind of choice of possible sememe partners. It is precisely in the nature of these relations however where on the one hand formally identical sememes of polysemic lexemes are distinguished and on the other the incompatible groupings are isolated. The operation of decision as to choice and admissibility of contextual sememe relations for all the relations mentioned (except for "Isotopie") consists in the fact that whether possible or necessary relations are stated between the two thesaurus ranges corresponding to feature combination of contextual partners is tested for each case.

The features and conditions which characterise sememes and thesaurus points are taken from diverse positions of three inventories. The first inventory consists of values of the underlying syntactic dependencies capable of obtaining between any two elements of particular word-classes, and the further division of these into the relations between word sub-classes, including those which can only be distinguished transformationally. Some hundreds of syntactic values of this degree of delicacy as characteristics of relations are



available. The second inventory of features contains conditions representing generalised semantic values which put into effect generally valid conditions and restrictions in the making of certain syntactic structures. These include for example the constraints on sememe pairs with characteristics such as "animal", "human", "abstract" etc. For cases in which the decision on compatibility or realisation has not yet been reached further limitations on combination have to be made by features from the third inventory. The elements of this one represent reduced or abstracted semantic features belonging to each single sememe of a given group of sememes as part of its meaning. The degree of abstraction corresponds approximately to such labels as "evaluation," "process", "simultaneity", "enumeration", "circumstance", "dimensionality", "condition" ect. With features such as these a considerable portion of the relations can be differentiated, which obtain between relatively distinct sememes of polysemic lexemes and their possible partners, e.g. the five sememes of the adjective licht or the two sememes of Leber (as an organ and as food). A borderline case is presented by zweite where the criteria "enumeration", "time-sequence" and "value judgement" coincide in many cases.

At this point, before the actual substantial analysis of sememes of seme-combination begins, the tasks and the possibilities of a thesaurus, which can only perform inner-linguistic contextual operations, end. Whether and to what extent the criteria of the third inventory are so adequate as to be used for the systematic establishment of equivalence relations between sememes remains to be seen by further research. Presumably however the basic concept of a thesaurus as outlined here must be elaborated into a wider, quite different structure, with semes as elements, if suprasyntagmatic relations are to be analysed.







## ОБ ОПИСАНИИ СИНТАКСИЧЕСКОЙ СТРУКТУРЫ ПРЕДЛОЖЕНИЯ.

А. В. Гладкий

В последнее время в лингвистике получили распространение два способа описания синтаксической структуры предложения: с помощью систем составляющих и с помощью деревьев подчинения <sup>\*</sup>/ . Каждый из этих способов имеет свои достоинства и недостатки. Преимущество использования деревьев подчинения состоит, в частности, в том, что они позволяют учитывать не только факт наличия синтаксических связей между теми или иными словами, но и направление этих связей. Кроме того, аппарат деревьев подчинения позволил обнаружить явление проективности, что существенно обогатило наши знания о синтаксическом строении предложения. В то же время деревья подчинения дают не вполне естественное описание синтаксических связей в тех случаях, когда направление последних нерелевантно, и в особенности тогда, когда реальные связи имеются не между отдельными словами, а между сочетаниями слов. Для описания подобных случаев больше приспособлены системы составляющих.

---

<sup>\*</sup> / Неформальное изложение обоих способов и описание взаимоотношения между ними имеется, например, в [1], формализованное — в [2].



Однако различные типы синтаксических связей сплошь и рядом соседствуют не только в одном языке, но и в одном предложении. Поэтому представляется целесообразным разработать "комбинированную" систему описания структуры предложения — так, чтобы в рамках этой системы были бы достаточно удобно представимы как те виды синтаксических связей, для которых направление релевантно, так и те, для которых оно нерелевантно, и связи могли бы устанавливаться не только между словами, но и между группами слов, причём эти группы сами могли бы быть снабжены некоторой внутренней синтаксической структурой.

В настоящем докладе предпринимается попытка построить систему указанного выше типа. Предлагаемый способ описания структуры предложения является одновременным обобщением понятий системы составляющих и дерева подчинения; оба эти понятия содержатся в нём как частные случаи. Этот способ обладает, в частности, той особенностью, что фигурирующие в нём в качестве "синтаксических единиц" группы слов не обязательно заполняют отрезки, а могут быть и разрывными; однако на их взаимное расположение накладываются некоторые ограничения типа проективности.

Новый способ формального описания синтаксической структуры предложения позволяет предложить также некоторое новое содержательное толкование этой структуры для ряда случаев, обычно вызывающих затруднения при синтаксическом анализе, и в особенности тех, которые при использовании деревьев подчинения трактуются как непроективные. При этом толковании синтаксической структуры проективность / в некотором обобщенном смысле / рассматривается как универсальное правило, не допускающее исключений. Такая интерпретация представляется нам более естественной, чем традиционная.



Впрочем, рассмотрения, относящиеся к интерпретации, носят пока что предварительный характер.

# 1. СИСТЕМА ОПИСАНИЯ СИНТАКСИЧЕСКОЙ СТРУКТУРЫ ПРЕДЛОЖЕНИЯ

/ Формальное изложение /.

Пусть  $\Pi$  — непустое конечное множество, на котором определено отношение линейного порядка / обозначаемое в дальнейшем символом  $<$  /. Элементы множества  $\Pi$  будут называться т о ч к а м и. В лингвистической интерпретации  $\Pi$  будет представлять собой предложение, а точки — вхождения в это предложение "слов".

Наименьший и наибольший элементы произвольного множества  $E \subseteq \Pi$  будут обозначаться соответственно  $\text{Inf } E$  и  $\text{Sup } E$ .

Для произвольных точек  $a, b$ ,  $a \leq b$ , положим

$$(a, b)_{\text{Def}} = \{ x \mid a < x < b \}$$

$$[a, b]_{\text{Def}} = \{ x \mid a \leq x \leq b \}$$

Множества вида  $/ a, b /$  будем называть и н т е р - в а л а м и, множества вида  $[a, b]$  — о т р е з к а м и.

Очевидно, интервал тогда и только тогда является отрезком, когда он не пуст; отрезок тогда и только тогда является интервалом, когда он не содержит  $\text{Inf } \Pi$  и  $\text{Sup } \Pi$ .

Как обычно, выражение " $/$  точка  $/$   $b$  лежит между  $/$  точками  $/$   $a$  и  $c$  " означает  $a < b < c \vee c < b < a$



Каждое непустое множество  $E \leq \Pi$  единственным образом представляется в виде

$$E = \bigcup_{i=1}^m [a_i, b_i]$$

где  $m \geq 1$  и при  $m > 1$  для любого  $i = 1, \dots, m-1$  интервал  $(b_i, a_{i+1})$  не пуст.

Отрезки  $[a_1, b_1], \dots, [a_m, b_m]$  мы будем называть компонентами множества  $E$ .

Нам понадобится ещё понятие циклической компоненты множества  $E \leq P$ , определяемое следующим образом. Пусть  $[a_1, b_1], \dots, [a_m, b_m]$  — все компоненты  $E$ , записанные в порядке возрастания. Если при этом  $a_1 = \inf \Pi$  и  $b_m = \sup \Pi$ , то циклическими компонентами  $E$  будут множества  $[a_2, b_2], \dots, [a_{m-1}, b_{m-1}], [a_m, b_m] \cup [a_1, b_1]$ . В противном случае циклические компоненты  $E$  совпадают с компонентами  $E$ .

Пусть  $E_1, E_2 \leq \Pi$ . Легко видеть, что  $E_1$  тогда и только тогда содержится в одной из циклических компонент множества  $\Pi \setminus E_2$ , когда  $E_2$  содержится в одной из циклических компонент множества  $\Pi \setminus E_1$ . Мы будем говорить, что  $E_1$  и  $E_2$  зацепляются, если  $E_1 \cap E_2 = \emptyset$  и  $E_1$  не содержится ни в одной циклической компоненте множества  $\Pi \setminus E_2$ , и что  $E_2$  расщепляет  $E_1$ , если  $E_1 \cap E_2 = \emptyset$  и  $E_1$  не содержится ни в одной компоненте множества  $\Pi \setminus E_2$ . Нетрудно заметить, что  $E_1$  и  $E_2$  зацепляются тогда и только тогда, когда  $E_2$  расщепляет  $E_1$  и  $E_1$  расщепляет  $E_2$ .

Упорядоченную пару  $(S, \longrightarrow)$ , где  $S$  — некоторое множество непустых подмножеств  $\Pi$  и  $\longrightarrow$  — бинарное отношение на  $S$ , мы будем называть системой синтаксических групп /ССТ/ на  $\Pi$ , если она удовлетворяет приводимым ниже аксиомам  $\mathcal{L}1 - \mathcal{L}3, R1 - R5$



$\mathcal{L}1$ .  $S$  содержит  $\Pi$  и все одноэлементные подмножества

$\mathcal{L}2$ . Если  $E_1, E_2 \in S$ , то либо  $E_1 \cap E_2 = \emptyset$ , либо  $E_1 \subseteq E_2$ , либо  $E_2 \subseteq E_1$ .

$\mathcal{L}3$ . Если  $E_1, E_2 \in S$  и  $E_1 \cap E_2 = \emptyset$ , то  $E_1$  и  $E_2$  не зацепляются.

Для формулировки следующих нам придётся предварительно ввести некоторые новые определения и обозначения.

Элементы множества  $S$  мы будем называть **синтаксическими группами** /  $SG$  /.

Если  $E_1, E_2$  — две  $SG$ ,  $E_1 \subset E_2$  <sup>\*</sup> и не существует  $SG$   $E'$ , такой, что  $E_1 \subset E' \subset E_2$ , мы будем говорить, что  $E_1$  непосредственно вложена в  $E_2$ . Очевидно, каждая  $SG$ , отличная от  $\Pi$ , непосредственно вложена в одну и только одну  $SG$ .

Далее, назовём **дугой** пару синтаксических групп  $(E, F)$ , для которой место  $E \rightarrow F$ , и **путём** — последовательность  $SG$   $E_1, E_2, \dots, E_k$  ( $k > 1$ ), в которой для каждого  $i = 1, \dots, k-1$  пара  $(E_{i-1}, E_i)$  есть дуга. Пусть  $E_1, \dots, E_k$  будем называть **путём из**  $E_1$  в  $E_2$ .

Будем писать  $E \Rightarrow F$ , если существует путь из  $E$  в  $F$ . Очевидно, из  $E \rightarrow F$  следует  $E \Rightarrow F$ .

Если  $E \rightarrow F$ , будем говорить, что  $F$  **подчинена**  $E$ , если  $E \Rightarrow F$  — что  $F$  **зависит от**  $E$ .

Сформулируем теперь аксиомы  $R1 - R5$

$R1$  Если  $E_1 \rightarrow E_2$ , то  $E_1$  и  $E_2$  непосредственно вложены в одну и ту же  $SG$ .

<sup>\*</sup>/ Запись  $E_1 \subset E_2$  означает  $E_1 \subseteq E_2 \text{ f } E_1 \neq E_2$



R2. Невозможно  $E \Rightarrow E$

R3. Если  $E_1 \rightarrow E$  и  $E_2 \rightarrow E$ , то  $E_1 = E_2$

R4. Если  $E, E_1, E_2$  — попарно непересекающиеся СГ и  $E_1 \rightarrow E_2$ , то множества  $E$  и  $E_1 \cup E_2$  не зацепляются.

Если  $E_1, E_2, E_3, E_4$  — попарно непересекающиеся СГ, такие, что  $E_1 \rightarrow E_2, E_3 \rightarrow E_4$ , то множества  $E_1 \cup E_2$  и  $E_3 \cup E_4$  не зацепляются.

Аксиомы  $\mathcal{L}1, \mathcal{L}2$  представляют собой обычные аксиомы для системы составляющих / см., например, [1], § 2 / \* /.

Аксиомы R2, R3 означают, что каждая связная компонента графа  $/C, \rightarrow/$  является деревом / точнее, прадеревом в смысле [4]\*\* /.

В силу аксиомы R4 все СГ, входящие в такое дерево / если оно отлично от дерева, состоящего из одной СГ  $\Pi$  / непосредственно вложены в одну и ту же СГ.

Чтобы представить в несколько более наглядном виде аксиомы R3, R4 и R5, сделаем следующее замечание. Будем говорить, что пары точек  $/a, b/$  и  $/c, d/$  разделяют друг друга, если  $a, b, c, d$  попарно

\* / Тем не менее  $\mathcal{C}$  не является, вообще говоря, системой составляющих в смысле [1], поскольку СГ не обязаны быть отрезками.

\*\* / Граф  $/M, \rightarrow/$  называется прадеревом, если :  
 а/ он не содержит контуров / замкнутых путей / ; б/ в нём имеется единственная вершина / называемая корнем / , в которую не входит ни одна дуга; в/ в каждую вершину, отличную от корня, входит точно одна дуга.



различны и одна из точек  $c, d$  лежит, а другая не лежит между  $a$  и  $b$  / или, что то же самое, одна из точек  $a, b$  лежит, а другая не лежит между  $c$  и  $d$ /. Очевидно, множества  $E_1, E_2 \subseteq \Pi$  такие, что  $E_1 \neq \emptyset, E_2 \neq \emptyset, E_1 \cap E_2 = \emptyset$  тогда и только тогда зацепляются, когда существуют точки  $a, b \in E_1$  и  $c, d \in E_2$ , такие, что пары  $/ a, b /$  и  $/ c, d /$  разделяют друг друга.

Теперь аксиомы  $\mathcal{L}3, R4$  и  $R5$  могут быть переформулированы так:

$\mathcal{L}3'$  Если  $E_1, E_2 \in \mathcal{C}$  и  $E_1 \cap E_2 = \emptyset$ , никакие две пары точек  $(a_1, b_1)$  и  $(a_2, b_2)$ , где  $a_1, b_1 \in E_1$  и  $a_2, b_2 \in E_2$  не разделяют друг друга.

$R4'$  Если  $E, E_1, E_2$  - попарно непересекающиеся СГ и  $E_1 \longrightarrow E_2$ , то никакие две пары точек  $(a, b)$  и  $(c, d)$ , где  $a, b \in E$  и  $c, d \in E_1 \cup E_2$  не разделяют друг друга.

$R5'$  Если  $E_1, E_2, E_3, E_4$  - попарно непересекающиеся СГ, такие, что  $E_1 \longrightarrow E_2, E_3 \longrightarrow E_4$ , то никакие две пары точек  $(a, b)$  и  $(c, d)$ , где  $a, b \in E_1 \cup E_2$  и  $c, d \in E_3 \cup E_4$  не разделяют друг друга.

Таким образом, аксиомы  $\mathcal{L}3, R4$  и  $R5$  могут быть истолкованы как некоторые условия проективности.

Системы составляющих и проективные отношения синтаксического подчинения  $/ [I], \S 3 /$  представляют собой "вырождённые" частные случаи ССГ. Именно, всякая система составляющих может рассматриваться как ССГ  $/ \mathcal{C}, \longrightarrow /$ , в которой все СГ являются отрезками и отношение  $\longrightarrow$  пусто / т.е.  $E_1 \longrightarrow E_2$  не имеет места ни для каких  $E_1, E_2 \in \mathcal{C} /$ , а всякое проективное отношение синтаксического подчинения / "проективное дерево подчинения" / - как ССГ  $/ \mathcal{C}, \longrightarrow /$ , распадающаяся, как граф, точно на две связных компоненты,



одна из которых состоит из всех одноэлементных подмножеств  $\Pi$ , а другая — из одного множества  $\Pi$  / так что никаких СГ, отличных от  $\Pi$  и неоднородных, нет /.

Сформулируем несколько утверждений о свойствах ССГ; их доказательств / впрочем, довольно простых / приводить не будем.

УТВЕРЖДЕНИЕ 1. Если две различные СГ  $E_1$  и  $E_2$  непосредственно вложены в одну и ту же СГ, то  $E_1 \cap E_2 = \emptyset$

СЛЕДСТВИЕ. Если  $E_1, E_2$  — СГ, такие, что  $E_1 \Rightarrow E_2$  то  $E_1$  и  $E_2$  не зацепляются.

Пусть  $D = \{E_1, \dots, E_k\}$  — подмножество  $C$ , такое, что соответствующий подграф графа  $/C, \longrightarrow/$  т.е. граф  $/D, \longrightarrow_D/$ , где  $\longrightarrow_D$  — отношение на  $D$ , индуцированное отношением связей / тогда он является прадеревом /. Множество  $E_1 \cup \dots \cup E_k$  мы будем называть в этом случае связным куском. В частности, всякая СГ есть связной кусок.

УТВЕРЖДЕНИЕ 2. Если  $H_1, H_2$  — связные куски и  $H_1 \cap H_2 = \emptyset$ , то  $H_1$  и  $H_2$  не зацепляются.

СЛЕДСТВИЕ 1. Если  $E, E_1, E_2$  — СГ, такие, что  $E_1 \Rightarrow E_2$ , и  $E$  не лежит на пути из  $E_1$  в  $E_2$ , то множества  $E$  и  $E_1 \cup E_1 \cup E_2$  не зацепляются.

СЛЕДСТВИЕ 2. Если  $E_1, E_2, E_3, E_4$  — СГ, такие, что  $E_1 \Rightarrow E_2$ ,  $E_3 \Rightarrow E_4$  и пути из  $E_1$  в  $E_2$  и из  $E_3$  в  $E_4$  не пересекаются, то множества  $E_1 \cup E_2$  и  $E_3 \cup E_4$  не зацепляются.

УТВЕРЖДЕНИЕ 3. Пусть  $/C, \longrightarrow/$  есть ССГ на  $\Pi$  и



пусть  $D$  — некоторое множество связанных кусков этой ССГ, содержащее  $\Pi$  и все одноточечные СГ и удовлетворяющее следующему условию: если  $H_1, H_2 \in D$  то либо  $H_1 \cap H_2 = \emptyset$  либо  $H_1 \leq H_2$  либо  $H_2 \leq H_1$ . Определим на  $D$  бинарное отношение  $\longrightarrow_D$  следующим образом:  $H_1 \longrightarrow_D H_2$  тогда и только тогда, когда  $E_1 \Longrightarrow E_2$ , где  $E_1$  и  $E_2$  — корни прадеревьев, соответствующих  $H_1$  и  $H_2$ , и при этом не существует  $H \in D$ , такого, что  $E_1 \Longrightarrow E \Longrightarrow E_2$ , где  $E$  — корень прадеревья, соответствующего  $H$ . Тогда  $\langle D, \longrightarrow_D \rangle$  также есть ССГ.

Введём теперь ещё две аксиомы.

R6 Если  $E_1, E_2 \in C$  и  $E_1 \rightarrow E_2$ , то  $E_1$  не расщепляет  $E_2$ .

R7 Если  $E_1, E_2, E_3 \in C$  и  $E_1 \rightarrow E_2 \rightarrow E_3$ , то  $E_1$  не расщепляет  $E_2 \cup E_3$ .

Чтобы пояснить смысл этих аксиом, заметим, что если  $E_1, E_2 \leq \Pi$ ,  $E_1 \neq \emptyset$ ,  $E_2 \neq \emptyset$ ,  $E_1 \cap E_2 = \emptyset$ , то  $E_1$  тогда и только тогда расщепляет  $E_2$ , когда существуют точки  $a \in E_1$ ,  $b, c \in E_2$ , такие, что  $a$  летит между  $b$  и  $c$ . Пользуясь этим, мы можем переформулировать R6 и R7 так:

R6 Если  $E_1, E_2 \in C$  и  $E_1 \rightarrow E_2$ , то ни для наших трёх точек  $a, b, c$ , где  $a \in E_1$ ,  $b, c \in E_2$ ,  $a$  не лежит между  $b$  и  $c$ .

Таким образом, аксиомы R6 и R7 могут быть истолкованы как некоторые условия сильной проективности / [I], § 3 /.

ССГ, удовлетворяющую аксиомам R6 и R7, мы будем называть с и л ь н о й ССГ, сокращённо СССГ.

УТВЕРЖДЕНИЕ 4. Пусть  $H_1$  и  $H_2$  — связанные куски в СССГ и  $E_1, E_2$  — корни соответствующих прадеревьев. Если  $H_1 \cap H_2 = \emptyset$  и  $E_1 \Longrightarrow E_2$ , то  $H_1$  не расщепляет  $H_2$ .



СЛЕДСТВИЕ. Если  $E_1, E_2$  - СГ в СССТ, такие, что  $E_1 \Rightarrow E_2$  то  $E_1$  не расщепляет  $E_2$ .

УТВЕРЖДЕНИЕ 5. Пусть  $H_1, E_2, E_3$  - попарно непересекающиеся связные куски в СССТ и  $E_1, E_2, E_3$  - корни соответствующих прадеревьев. Если  $E_1 \Rightarrow E_2 \Rightarrow E_3$ , то  $H_1$  не расщепляет  $H_2 \cup H_3$

СЛЕДСТВИЕ. Если  $E_1, E_2, E_3$  - СГ в СССТ, такие, что  $E_1 \Rightarrow E_2 \Rightarrow E_3$ , то  $E_1$  не расщепляет  $E_2 \cup E_3$

УТВЕРЖДЕНИЕ 6. Пусть  $/C, \rightarrow /$  - СССТ на  $\Pi$ . Если  $D_n \rightarrow_D$  определяются так же, как в утверждении 3, то  $(D, \rightarrow_D)$  также есть СССТ.

## II. ЗАМЕЧАНИЯ ОТНОСИТЕЛЬНО ИНТЕРПРЕТАЦИИ. ПРИМЕРЫ.

Описание синтаксической структуры предложения с помощью систем синтаксических групп обладает, как нам представляется, большей "гибкостью", чем обычные описания с помощью систем составляющих или деревьев подчинения. Преимущества нового способа описания становятся наиболее наглядными тогда, когда естественно трактовать те или иные группы слов как "синтаксически целостные", не отказываясь в то же время от введения направленных синтаксических связей.

В русском языке такое положение возникает, в частности, в следующих случаях:

I./ Предложения, содержащие сложные формы глагола или



сочетания модальных глаголов с инфинитивом, например:

/ 1 / Подобные<sub>1</sub> законы<sub>2</sub> мы<sub>3</sub> будем<sub>4</sub> называть<sub>5</sub> логическими<sub>6</sub>

---

$$\{ 4, 5 \} = A$$

$$A \rightarrow 3 ; A \rightarrow 2 ; A \rightarrow 6 ; 2 \rightarrow 1 ;$$

/ 2 / Зимой<sub>1</sub> здесь<sub>2</sub> будет<sub>3</sub> работать<sub>4</sub> каток<sub>5</sub>

---

$$\{ 3, 4 \} = A$$

$$A \rightarrow 5 ; A \rightarrow 1 ; A \rightarrow 2$$

/ 3 / Я буду<sub>1</sub> там<sub>2</sub> работать<sub>3</sub> завтра<sub>4</sub>

---

$$\{ 2, 4 \} = A$$

$$A \rightarrow 1 ; A \rightarrow 3 ; A \rightarrow 5$$

/ 4 / Мы<sub>1</sub> должны<sub>2</sub> это<sub>3</sub> расстояние<sub>4</sub> пробежать<sub>5</sub> быстро<sub>6</sub>

---

$$\{ 2, 5 \}$$

$$A \rightarrow 1 ; A \rightarrow 4 ; A \rightarrow 6 ; 4 \rightarrow 3 ;$$


---

ж /

Все примеры будут записываться по следующей схеме. Сначала выписывается предложение с нумерованными вхождениями слов; номера вхождения считаются элементами множества  $\Pi$ . Далее даётся перечень всех нетривиальных СГ / т.е. не одноэлементных и отличных от  $\Pi$  /; эти СГ обозначаются простыми латинскими буквами. Наконец, выписывается перечень зависимостей между СГ.



2./ Предложения, содержащие предложные группы, например:

/ 5 / Он<sub>1</sub> учится<sub>2</sub> в<sub>3</sub> университете<sub>4</sub>

---

$$\{ 2, 5 \} = A$$

$$2 \rightarrow 1 ; 2 \rightarrow A$$

/ 6 / Он<sub>1</sub> учится<sub>2</sub> в<sub>3</sub> Новосибирском<sub>4</sub> университете<sub>5</sub>

---

$$\{ 3, 4 \} = A$$

$$2 \rightarrow 1 ; 2 \rightarrow A ; A \rightarrow 4$$

3./ Предложения, содержащие частицу не и некоторые другие частицы, например:

/ 7 / Иванов<sub>1</sub> вчера<sub>2</sub> не<sub>3</sub> приехал<sub>4</sub>

---

$$\{ 3, 4 \} = A$$

$$A \rightarrow 1 ; A \rightarrow 2 ; 4 \rightarrow 3$$

/ 8 / Иванов<sub>1</sub> не<sub>2</sub> вчера<sub>3</sub> приехал<sub>4</sub>

---

$$\{ 2, 3 \} = A$$

$$4 \rightarrow 1 ; 4 \rightarrow A ; 3 \rightarrow 2$$

/ 9 / Иванов<sub>1</sub> ведь<sub>2</sub> не<sub>3</sub> приехал<sub>4</sub>

---

$$\{ 2, 3, 4 \} = A ; \{ 3, 4 \} = B$$

$$A \rightarrow 1 ; B \rightarrow 2 ; 4 \rightarrow 3$$



4./ Предложения с однородными членами, например:

/ IO / Он<sub>1</sub> прочёл<sub>2</sub> газету<sub>3</sub> и<sub>4</sub> журнал<sub>5</sub>

---


$$\{ 3, 4, 5 \} = A$$

$$2 \longrightarrow 1 ; 2 \longrightarrow A$$

/ II / Он<sub>1</sub> работал<sub>2</sub> и<sub>3</sub> отдыхал<sub>4</sub>

---


$$\{ 2, 3, 4 \} = A$$

$$A \longrightarrow 1$$

Такие предложения могут содержать парные союзы, например:

/ I2 / Он<sub>1</sub> не<sub>2</sub> прочёл<sub>3</sub> ни<sub>4</sub> газету<sub>5</sub>, ни<sub>6</sub> журнал<sub>7</sub>

---


$$\{ 2, 3 \} = A \quad \{ 4, 5, 6, 7 \} = B ; \{ 4, 6 \} = C$$

$$A \longrightarrow 1 ; A \longrightarrow B ; 3 \longrightarrow 2$$

/ I3 / Он<sub>1</sub> не<sub>2</sub> только<sub>3</sub> работал<sub>4</sub> , но<sub>5</sub> и<sub>6</sub> отдыхал<sub>7</sub>

---


$$\{ 2, \dots, 7 \} = A ; \{ 2, 3, 5, 6 \} = B$$

$$A \longrightarrow 1$$

5./ Сложные предложения, например:

/ I4 / Человек<sub>1</sub> , которого<sub>2</sub> ты<sub>3</sub> вчера<sub>4</sub> видел<sub>5</sub> , уже<sub>6</sub> уехал<sub>7</sub>

---


$$\{ 2, 3, 4, 5 \} = A$$

$$7 \longrightarrow 1 ; 7 \longrightarrow 6 ; 1 \longrightarrow A ; 5 \longrightarrow 3 ; 5 \longrightarrow 2 ; 5 \longrightarrow 4$$



/ I5 / Я<sub>1</sub> узнал<sub>2</sub>, что<sub>3</sub> он<sub>4</sub> приехал<sub>5</sub> от<sub>6</sub> Иванова<sub>7</sub>

---


$$\{3, 4, 5\} = A; \{4, 5\} = B; \{6, 7\} = C,$$

$$2 \rightarrow 1; 2 \rightarrow A; 2 \rightarrow C; 5 \rightarrow 4$$

/ I6 / Деревня<sub>1</sub>, где<sub>2</sub> скучал<sub>3</sub> Евгений<sub>4</sub>, была<sub>5</sub>

---

прекрасный<sub>6</sub> уголок<sub>7</sub>

---


$$\{2, 3, 4\} = A$$

$$5 \rightarrow 1; 5 \rightarrow 7; 1 \rightarrow A; 7 \rightarrow 6; 3 \rightarrow 4; 3 \rightarrow 2$$

/ I7 / Иванов<sub>1</sub> говорил<sub>2</sub>, а<sub>3</sub> Петров<sub>4</sub> молчал<sub>5</sub>

---


$$\{1, 2\} = A \quad \{4, 5\} = B$$

$$2 \rightarrow 1; 5 \rightarrow 4$$

В частности – сложные предложения с парными союзами, например:

/ I8 / Если<sub>1</sub> я<sub>2</sub> приду<sub>3</sub>, то<sub>4</sub> он<sub>5</sub> уйдёт<sub>6</sub>

---


$$\{2, 3\} = A; \{5, 6\} = B; \{1, 4\} = C$$

$$3 \rightarrow 2; 6 \rightarrow 5$$

Иногда для описания синтаксической структуры предложения используются "размеченные системы составляющих", в которых для каждой не однозначной составляющей среди всех непосредственно вложенных в неё составляющих выделяется одна, называемая главной. Такое описание может быть пред-



ставлено, как ССТ, обладающая тем свойством, что для каждой СГ множество всех непосредственно вложенных в неё СГ, если оно не пусто, является "деревом высоты I". Например:

/ I9 /      Смелый      охотник      убил      медведя

---

$$\{1, 2\} = A ; \quad \{3, 4\} = B$$

$$B \longrightarrow A ; \quad 3 \longrightarrow 4 ; \quad 2 \longrightarrow I.$$

Разумеется, одно и то же предложение может допускать несколько разных описаний с помощью ССТ, причём это не обязательно связано с различиями в понимании смысла предложения. Так, предложение / I9 / может быть описано, например, с помощью ССТ, состоящей из тех же СГ, что и приведённая выше, но с пустым отношением / обычная система составляющих /, или с помощью ССТ, имеющей только тривиальные СГ и отношение  $\longrightarrow$ , задаваемое таблицей:  $3 \longrightarrow 2 ; 3 \longrightarrow 4 ; 2 \longrightarrow I$  / обычно дерево подчинения /. Правда, обычная система составляющих содержит в данном случае менее полную информацию о структуре предложения, чем размеченная. Другой пример: в предложении / I6 / вполне естественно было бы ввести ещё три СГ  $B = \{3, 4\}$ ,  $C = \{5, 6, 7\}$  и  $D = \{6, 7\}$  и заменить дуги  $3 \longrightarrow 2$ ,  $5 \longrightarrow I$  и  $5 \longrightarrow 7$  дугами  $B \longrightarrow 2$ ,  $C \longrightarrow I$  и  $5 \longrightarrow D$ .

Неоднозначности доказанного типа возможны, впрочем, и при описании структуры предложения с помощью обычных систем составляющих.

Наиболее существенным преимуществом систем синтаксических групп нам представляется то обстоятельство, что они позволяют в значительной степени / а при некотором расширительном толковании синтаксических связей, возможно и



полностью / "снять" явление непроективности.

Встречающиеся в русском языке непроективные структуры могут быть условно разделены на два типа – "нейтральные" – допустимые в любых предложениях, в том числе тех, которые не преследуют никаких целей, кроме сообщения информации / как в научно-технических текстах /, и "стилистически окрашенные", допустимые лишь в предложениях, преследующих какие-либо дополнительные цели, например, выражение отношения говорящего к сообщаемому в предложении факту.

"Нейтральные" непроективности в большинстве случаев / если не всегда / появляются, как кажется, в силу того обстоятельства, что в обычных деревьях подчинения не учитывается существование между словами связей разной степени близости, так что те группы слов / не обязательно идущих подряд /, между которыми имеется "более близкая" связь, должны в дереве "более далёкой" связи выступать как элементарные единицы / вершины /. Поэтому использование ССГ позволяет устранять такие непроективности. Проиллюстрируем это на следующих примерах, заимствованных из книги. Л. Н. Иорданской [3]. / Деревья предложений / 20 / – / 25 /, приведённые в [3] на стр. 16–17, непроективны /.

/ 20 / Он<sub>1</sub> написал<sub>2</sub> такое<sub>3</sub> письмо<sub>4</sub> , что<sub>5</sub> ...

---


$$\{ 3, 4 \} = A ; \{ 5, \dots \} = B$$

$$2 \rightarrow 1 ; A \rightarrow B ; 3 \rightarrow 2$$

/ 21 / Лингвисты<sub>1</sub> язык<sub>2</sub> рассматривают<sub>3</sub> как<sub>4</sub> систему<sub>5</sub>

---


$$\{ 2, 3 \} = A ; \{ 4, 5 \} = B$$

$$A \rightarrow 1 ; A \rightarrow B ; 3 \rightarrow 2$$



/ 22 / Это<sub>1</sub> более<sub>2</sub> понятная<sub>3</sub> книга<sub>4</sub> , чем<sub>5</sub> та<sub>6</sub>

$$\{2, \dots, 6\} = A ; \{5, 6\} = B ; \{2, 3\} = C$$

$$A \rightarrow 1 ; 4 \rightarrow C ; C \rightarrow B ; 3 \rightarrow 2$$

/ 23 / Почему<sub>1</sub> из<sub>2</sub> многих<sub>3</sub> этим<sub>4</sub> занимаются<sub>5</sub> только<sub>6</sub>

некоторые<sub>7</sub> ?

$$\{2, \dots, 7\} = A , \{2, 3, 6, 7\} = B ; \{2, 3\} = C$$

$$A \rightarrow 1 ; 5 \rightarrow 4 ; 5 \rightarrow B ; 7 \rightarrow 6$$

/ 24 / В<sub>6</sub> этот<sub>2</sub> список<sub>3</sub> не<sub>4</sub> вошли<sub>5</sub> ни<sub>6</sub> элемент<sub>7</sub> А<sub>8</sub> ,

ни<sub>9</sub> элемент<sub>10</sub> В<sub>11</sub>

$$\{1, 3\} = A ; \{4, 5\} = B ; \{6, \dots, 11\} = C ; \{6, 9\} = D ; \{7, 8\} = E ; \{10, 11\} = F$$

$$5 \rightarrow C ; 5 \rightarrow A ; 7 \rightarrow 8 ; 10 \rightarrow 11 ; A \rightarrow 2 ;$$

/ 25 / О<sub>1</sub> докладе<sub>2</sub> дала<sub>3</sub> отзыв<sub>4</sub> комиссия<sub>5</sub>

$$\{1, 2\} = A ; \{3, 4\} = B$$

$$B \rightarrow 5 ; B \rightarrow A ; 3 \rightarrow 4$$

Непроективные деревья подчинения дали бы также приведённые выше предложения / 2 / , / I2 / , / I3 / .

Что касается "стилистически окрашенных" непроективных конструкций, то относительно них можно заметить следующее. Представляется бесспорным, что в предложениях, содержащих такие конструкции имеют место некоторые отклонения от "идеальной правильности" синтаксической структуры, причём эти отклонения как раз и обеспечивают выполнение предложением



его добавочных функций. Внешний эффект подобного отклонения обычно состоит в изменении нормального порядка слов, так что некоторое слово ощущается как "особо выделенное". Например, в предложении

/ 26 / Хорошую<sub>1</sub> ты<sub>2</sub> компанию<sub>3</sub> нашёл<sub>4</sub>!

интуитивно ощущается как "выделенное" слово хорошую. При представлении структуры предложения / 26 / с помощью дерева подчинения это явление трактуется как нарушение проективности, в результате которого слово хорошую оказывается на необычном месте. Но более естественно, быть может, трактовать это явление иначе — как разрыв обычных синтаксических связей: слово хорошую присоединяется не к слову компанию, а ко всему предложению; нарушения проективности при этом не происходит. Применение ССГ позволяет легко формализовать такую трактовку. Соответствующее описание структуры предложения / 26 / выглядит так:

$$A = \{ 2, 3, 4 \}$$

$$A \rightarrow 1; 4 \rightarrow 2; 4 \rightarrow 3$$

Разумеется, морфологически слово хорошую остаётся при этом связанным со словом компанию.

Таким способом целесообразно, быть может, описывать и те случаи, когда эффект "выделения" некоторого слова достигается без изменения порядка слов, например, с помощью интонации в устной речи или подчёркивания на письме. Так, если в предложении:



/ 27 / Ты<sub>1</sub> нашёл<sub>2</sub> хорошую<sub>3</sub> компанию<sub>4</sub> !

---

делается ударение на слове хорошую, то этот факт может быть отражён в следующем описании структуры предложения:

$$A = \{ 1, 2, 4 \}$$

$$A \rightarrow 3; 2 \rightarrow 1; 2 \rightarrow 4$$

/ в то время как "нормальная" ССГ для данного предложения имеет вид:

$$2 \rightarrow 1; 2 \rightarrow 4; 4 \rightarrow 3 /.$$

Предложенная трактовка "выделения" не может быть непосредственно приложена к тем случаям, когда выделяется прилагательное с предлогом, например:

/ 28 / В<sub>1</sub> хорошую<sub>2</sub> ты<sub>3</sub> компанию<sub>4</sub> попал<sub>5</sub> !

---

Автору кажется возможным предложить описывать такие конструкции как возникающие в результате эллипсиса из "идеальных предложений" типа

/ 28 \*/ \*В<sub>1</sub> хорошую<sub>2</sub> ты<sub>3</sub> в<sub>4</sub> компанию<sub>5</sub> попал<sub>6</sub> ! \*

---

в которых предлог при прилагательном может рассматриваться как своего рода "морфологический показатель", обеспечивающий, так сказать, "согласование" этого прилагательного с предложной конструкцией. Таким образом, ССГ для "предложения" / 28 \*/ будет иметь вид:

---

\* Как известно, подобные конструкции реально встречаются в русской народной поэзии.



$$\{3, 4, 5, 6\} = A; \{1, 2\} = B; \{4, 5\} = C.$$

$$A \longrightarrow B; 6 \longrightarrow 3; 6 \longrightarrow C,$$

а для предложения / 28 / - вид:

$$\{3, 4, 5\} = A; \{1, 2\} = B.$$

$$A \longrightarrow B; 5 \longrightarrow 3; 5 \longrightarrow 4.$$

Приведем теперь два более сложных примера описания структуры предложения с помощью ССТ.

/ 29 / Например<sub>1</sub>, если<sub>2</sub> требуется<sub>3</sub> при<sub>4</sub> любых<sub>5</sub> а<sub>6</sub>

---

вычислять<sub>7</sub> а<sub>8</sub>, то<sub>9</sub> машина<sub>10</sub> должна<sub>11</sub> каждый<sub>12</sub>

---

раз<sub>13</sub> перед<sub>14</sub> умножением<sub>15</sub> уже<sub>16</sub> полученного<sub>17</sub>

---

результата<sub>18</sub> на<sub>19</sub> а<sub>20</sub> сравнивать<sub>21</sub> число<sub>22</sub> уже<sub>23</sub>

---

произведённых<sub>24</sub> умножений<sub>25</sub> с<sub>26</sub> n<sub>27</sub>

---

$$\{2, \dots, 27\} = A; \{3, \dots, 8\} = B; \{10, \dots, 27\} = C; \{2, 9\} = D$$

$$\{3, 7\} = E; \{11, 21\} = F; \{4, 6\} = G; \{12, 13\} = H;$$

$$\{14, 15\} = I; \{19, 20\} = J; \{26, 27\} = K$$

$$A \rightarrow 1; E \rightarrow 8; E \rightarrow G; G \rightarrow 5; F \rightarrow 10; F \rightarrow 22; F \rightarrow K$$

$$F \rightarrow H; F \rightarrow 1; 13 \rightarrow 12; 1 \rightarrow 18; I \rightarrow J; 18 \rightarrow 17; 17 \rightarrow 16$$

$$22 \rightarrow 25; 25 \rightarrow 24; 24 \rightarrow 23$$



/ 30 / В<sub>1</sub> зависимости<sub>2</sub> от<sub>3</sub> сюжета<sub>4</sub> работы<sub>5</sub> бывают<sub>6</sub>  


---

добрыми<sub>7</sub> или<sub>8</sub> злыми<sub>9</sub>, коварными<sub>10</sub> или<sub>11</sub> готовы-  


---

ми<sub>12</sub> на<sub>13</sub> самопожертвование<sub>14</sub>, комически<sub>15</sub> ог-  


---

раниченными<sub>16</sub> или<sub>17</sub> всезнающими<sub>18</sub> мудрецами<sub>19</sub>,  


---

но<sub>20</sub> во<sub>21</sub> всех<sub>22</sub> случаях<sub>23</sub> авторское<sub>24</sub> отношение<sub>25</sub>  


---

писателя<sub>26</sub> к<sub>27</sub> такому<sub>28</sub> роботу<sub>29</sub> напоминает<sub>30</sub>  


---

скорее<sub>31</sub> отношение<sub>32</sub> к<sub>33</sub> человеку<sub>34</sub>, чем<sub>35</sub> к<sub>36</sub>  


---

машине<sub>37</sub>.

$\{1, \dots, 19\} = A$ ;  $\{21, \dots, 37\} = B$ ;  $\{1, 2\} = C$ ;  $\{3, 4\} = D$   
 $\{6, \dots, 19\} = E$ ;  $\{7, \dots, 19\} = F$ ;  $\{7, 8, 9\} = G$ ;  $\{10, \dots, 14\} = H$   
 $\{12, 13, 14\} = I$ ;  $\{13, 14\} = J$ ;  $\{15, \dots, 19\} = K$ ;  $\{15, 16\} = L$   
 $\{18, 19\} = M$ ;  $\{21, 23\} = N$ ;  $\{27, 29\} = O$ ;  $\{31, \dots, 37\} = P$   
 $\{31, 35\} = Q$ ;  $\{32, 33, 34\} = R$ ;  $\{34, 35\} = S$ ;  $\{36, 37\} = T$   
 $\{E \rightarrow 5; E \rightarrow C; C \rightarrow D; 6 \rightarrow F; 12 \rightarrow J; 16 \rightarrow 15$   
 $19 \rightarrow 18; 30 \rightarrow 25; 30 \rightarrow P; 30 \rightarrow N; N \rightarrow 22; 25 \rightarrow 24$   
 $25 \rightarrow 26; 25 \rightarrow O; O \rightarrow 28; 32 \rightarrow S.$



В заключение заметим, что во всех приведенных примерах, кроме / 22 / и / 23 /, построенные нами ССТ являются сильными. Можно было бы, конечно, и предложения / 22 / и / 23 /, описать с помощью СССТ, но такое описание было бы, пожалуй, менее естественным. Вопрос об адекватности СССТ для описания структуры предложения требует специального рассмотрения.

## Л И Т Е Р А Т У Р А

1. А. В. Гладкий, Лекции по математической лингвистике для студентов. Новосибирск, 1966.
2. Е. В. Падучева, О способах представления синтаксической структуры предложения. Вопросы языкознания, 1964, № 2.
3. Л. Н. Иорданская, Автоматический синтаксический анализ, ч.П., Новосибирск, 1967.
4. К. Берж, Теория графов и её применение. М., 1962.



## ON THE SYNTACTIC DESCRIPTION OF SENTENCES

### (Summary)

There are two basic tools for the formal syntactic description of sentences used in recent linguistic works: phrase structures and dependency relations (dependency trees). The first of these methods enables us to account for the existence of different degrees of "nearness" in syntactic connections and to lay down connections between the word groups, not only between words; the second one works just with connections between words and does not differentiate among the levels of syntactic nearness. In the description of this type the connections are directed.

This paper is an attempt to construct some generalized formal system for the syntactic descriptions of sentences which would enable us to work with directed connections and at the same time to lay down the connections between word groups (which may be disconnected in the sense of word order) as well as to differentiate among the levels of "nearness". The author hopes that by means of this system in the syntactic description one can, in particular, avoid non-projectivity.

The basic notions of the system are laid down by means of an axiom system as follows.

The sentence is formally represented as a finite set  $\Pi$  with a total ordering noted by  $<$ . The elements of  $\Pi$  are called points.

An ordered pair  $(C, \rightarrow)$ , where  $C$  is a set of non-empty subsets of  $\Pi$ , and  $\rightarrow$  is a binary relation on  $C$ , we call a system of syntactic groups (SSG) on  $\Pi$  (and elements of  $C$  are called in this case syntactic groups - SG's/, if the following axioms are fulfilled:



- $\mathcal{L}1$ .  $C$  contains  $\overline{T}$  and all one-element subsets of  $\overline{T}$ .
- $\mathcal{L}2$ . If  $E_1, E_2 \in C$ , then either  $E_1 \cap E_2 = \emptyset$  or  $E_1 \leq E_2$  or  $E_2 \leq E_1$ .
- $\mathcal{L}3$ . If  $E_1, E_2 \in C$  and  $E_1 \cap E_2 = \emptyset$ , then there are no points  $a_1, b_1 \in E_1, a_2, b_2 \in E_2$  such that either  $a_2$  lies between  $a_1$  and  $b_1$  (i.e.  $a_1 < a_2 < b_1$  or  $b_1 < a_2 < a_1$ ) and  $b_2$  does not lie between  $a_1$  and  $b_1$ , or  $b_2$  lies and  $a_2$  does not lie between  $a_1$  and  $b_1$ . (We say in this case that  $E_1$  and  $E_2$  do not interlace).
- $R1$ . If  $E_1 \rightarrow E_2$ , then there exists a SG,  $E$  such that  $E_1 \leq E_2 \leq E$  and there exists no SG,  $E', E''$  such that  $E_1 \leq E' \leq E, E_2 \leq E'' \leq E, E_1 \neq E' \neq E, E_2 \neq E'' \neq E$ .
- $R2$ . There is no sequence  $E_1, E_2, \dots, E_n$  ( $n \geq 1$ ) such that for every  $i, 1 \leq i \leq n-1, E_{i-1} \rightarrow E_i$  and  $E_1 = E_n$ .
- $R3$ . If  $E_1 \rightarrow E$  and  $E_2 \rightarrow E$ , then  $E_1 = E_2$ .
- $R4$ . If  $E, E_1, E_2$  are pairwise disjoint SG's and  $E_1 \rightarrow E_2$  then  $E$  and  $E_1 \cup E_2$  do not interlace.
- $R5$ . If  $E_1, E_2, E_3, E_4$  are pairwise disjoint SG's and  $E_1 \rightarrow E_2, E_3 \rightarrow E_4$  then  $E_1 \cup E_2$  and  $E_3 \cup E_4$  do not interlace.
- $\mathcal{L}3, R4$  and  $R5$  are some projectivity properties.

It is evident that usual phrase structures (constituent structures) and dependency trees are degenerate special cases of SSG's.

Some propositions about properties of SSG's are formulated. Then various examples of SSG's are given for Russian sentences.



## ON THE PROBLEM OF WORD ORDER<sup>x/</sup>

Ferenc Kiefer

1. Early works on transformational grammar (e.g. Chomsky 1957) considered word order changes as straightforward effects of optional transformational rules. In later works (e. g. Chomsky 1965), too, word order was looked upon as a mere stylistic matter and, as a consequence, the question was raised as to whether word order changes should be accounted for in grammar at all. In other words, Chomsky and others were inclined to relegate the problem of word order in toto to performance rather than to consider it as a matter of competence. To show that this contention is not quite justified is one of the aims of the present paper. We shall put forth, though tentatively, the general principles of handling the problem of word order. The present considerations are based on my earlier study of word order in Hungarian (Kiefer 1967). Here I will, however, treat several points differently. Recent development in the theory of grammar (Bierwisch, Fillmore 1968 a and b, McCawley a.o.) make several improvements on my earlier treatment possible. Furthermore, so far I have focussed my attention on the role of emphasis in determining word order. It is clear, however, that apart from emphasis the topic-comment relation constitutes another important factor in word order changes (Daneš, Firbas, Heidolph, Novak and most recently Sgall; with respect to Hungarian cf. also Elekfi, Dezső, Dezső-Szépe). Finally, here I will not restrict myself to Hungarian though Hungarian seems to be particularly appropriate for a study of word order.

---

<sup>x/</sup> To appear in Recent Developments in Linguistics (M. Bierwisch and K.E. Heidolph, eds.). Mouton and Co. The Hague.



2. Emphasis is not only a syntactic problem because it can be predicted by syntactic rules as claimed by Lu (Lu 1965) but also, and more importantly, because several syntactic rules are triggered by a syntactic feature which might be referred to as Emph. The question of how this Emph gets into syntax will be one of our main concern. Before going into this problem let me adduce some of the syntactic constructions that are determined by Emph.

a/ In English the two best-known examples are

/i/ the cleft-sentences like

It is Mary who I want to marry.

It is the big book that I have read.

/ii/ and the emphatic do

I do hope that she will come.

b/ In French the constructions with "c'est ... qui" and "c'est ... que" cannot be explained without Emph. The same holds with respect to the repetition of personal pronouns like in

Moi, je ne comprends rien.

c/ In German, apart from some word order changes, there are several (transformational) rules that make use of Emph. E.g.

nicht ein  $\implies$  kein

is obligatory except in case of emphasis, because then we may have either

nicht ein ... (sondern zwei)

nicht ein Buch ... (sondern ein Heft)

or

kein Buch ... (sondern ein Heft)

d/ The ambiguity of the Russian sentence

матѣ бугѣла гоучѣ

can only be explained by means of Emph. The two possibilities are

матѣ бугѣла гоучѣ

матѣ бугѣла гоучѣ

On the other hand,

матѣ бугѣл сын

is only grammatical if "матѣ" is emphatic.



e/ In Hungarian if the verb "van" (to be) is used existentially, it must be emphatic. In this case, some word order changes are obligatory. Consider, for example,

Péter van az osztályban.

Peter is the class-in

There is somebody called Peter in the class.

versus

Péter az osztályban van.

Peter is in the classroom.

On the other hand,

Levelet olvas János.

letter reads John

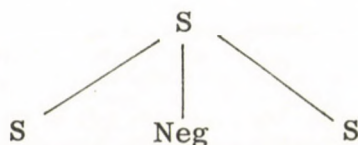
is only possible if "levelet" is emphasized:

Levelet olvas János.

It is a letter that John is reading.

These examples will suffice to show that emphasis is an important syntactic problem.

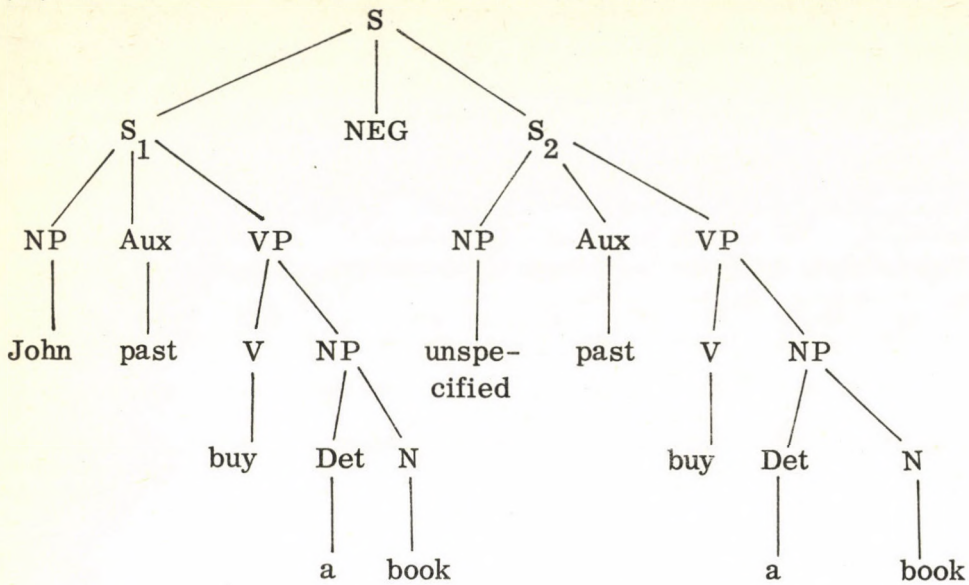
Undoubtedly, there are several types of emphasis. In the case of emotional emphasis everything can receive emphasis (any morpheme or even any phoneme). Therefore, no rules govern this kind of emphasis. Consequently, it falls outside the scope of a competence model. In contrast to emotional emphasis, logical emphasis reveals a well-defined syntactic structure that can roughly be rendered by the following diagram



where the two lower S's are almost identical, they differ in one lexical item only. This fact has already been observed by Lu who propounds the following deep structure for the sentence "John bought a book." (op. cit. p. 41.):



/1/



In order to get the surface structure "John bought a book" one has simply to apply some transformational rules to /1/:

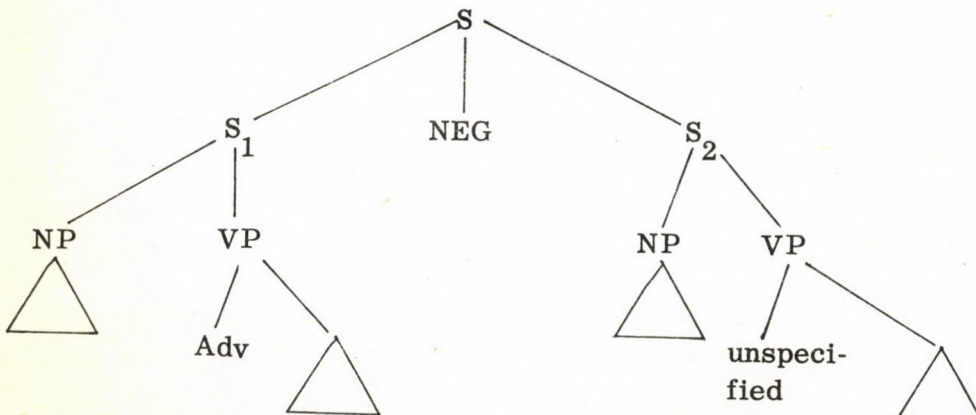
T1: Delete the whole  $S_2$ .

T2: Delete NEG

T3: Add stress to the word in  $S_1$  whose counterpart in  $S_2$  is unspecified.

This proposal leaves, however, several important questions unanswered. First of all, in which way can structures like (1) be generated? Secondly, the aforementioned deletion transformations seem to contradict the well-known conditions of deletability (Cf. Chomsky 1965, p. 182).

As to the first problem, notice that in Hungarian there are many adverbials that can occur in structures like (1) but cannot receive emphasis. What I mean by that is that one may also have



where "unspecified" will be replaced by some adverb.



Consider, for example,

+ Folyton nézett rám.  
continually looked-he me-at  
+ Végül ment el.  
finally went-he away  
+ Minduntalan jött haza.  
incessantly came-he home

versus

Folyton rám nézett.  
Végül elment.  
Minduntalan hazajött.<sup>1/</sup>

One may have

(3) Péter folyton olvasott és nem olvasott ritkán.

Peter continually read and not read rarely

This fact indicates that structures like (1) can be interpreted as a deep structure of an emphatic sentence only if the lexical items do not prove the contrary. Before proceeding, however, another remark is necessary. Instead of (3) we may also have

(4) Péter folyton olvasott és nem ritkán olvasott

where the adverb "folyton" is emphatic. This is, however, not logical emphasis but rather a metalinguistic emphasis that does not affect meaning.<sup>2/</sup>

One can adduce quite easily a great number of further examples where emphasis is not possible. So, for instance, articles cannot take emphasis in many languages. Although one may have in Hungarian

<sup>1/</sup> Notice, incidentally, that in emphatic sentences the particle comes always after the verb.

<sup>2/</sup> This kind of emphasis can be conceived either as a correction of some sort of mispronunciation or of the inappropriate use of a word. In (4) one has the latter case. This means, that (4) involves somehow that one should not say "rarely" with respect to one's reading. From a semantic point of view, "folyton" cannot be contrasted with other adverbs. In other words, one is unable to grasp the meaning of "folyton" if it is emphatic and not used metalinguistically. It is, so to speak, void of meaning.



Az ember ment el otthonról és nem egy ember.

Here emphasis has a metalinguistic character and when we assert that no emphasis is possible than this is to be taken in the sense that logical emphasis is out of the question. The same holds for emphasis in the following examples

Der Mann ging weg und nicht die Mann.

Der Mann ging weg und nicht ein Mann.

It is interesting to note that whenever the indefinite article is emphasized it becomes the numeral one.<sup>3/</sup>

I do not want to dwell on this problem any further.<sup>4/</sup>

It seems to be beyond any doubt that emphasis is not a purely syntactic but also a lexical problem. If so, then either structures like (1) are not the most appropriate way to handle it or it must be supplemented by some conditions as to the lexical items.

In fact, one might think of other possibilities of generating emphatic sentences. This is apparently supported by some further syntactic properties of emphasis, the most important being the fact that no simple sentence can have more than one emphatic constituent. Here "simple" means that the deep structure of the sentence contains only one S. This suggests the following rule for the introduction of emphasis as a syntactic feature:

(5)  $S \longrightarrow (\text{Emph}) \dots NP \quad VP$

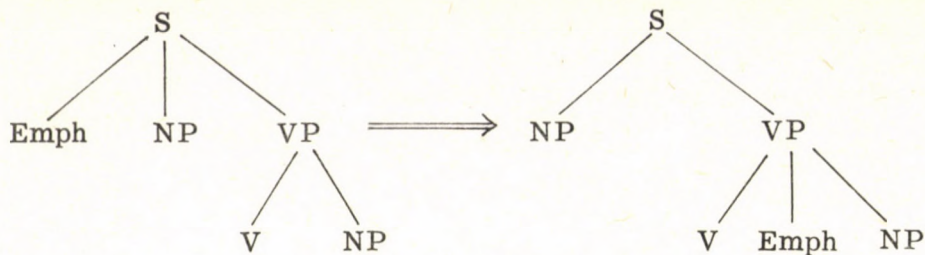
Then one wants Emph, optionally introduced in (5), to be attached to the right nodes in the base P-markers. Otherwise one would need extra filtering transformations, i.e. transformations that do not perform other tasks except for filtering out the wrong deep structures, a solution which seems to me quite counterintuitive. One simply does not like rules in a grammar that do not play a creative role and just filter. Of course, also simplicity is violated by this solution. In order to avoid this one may stipulate other solutions. As far as I can see two possibilities present themselves. One could introduce

<sup>3/</sup> Lu observes that the indefinite article in English, if emphasized, should be interpreted as "one" or "a single". Does not this indicate that indefinite articles cannot take emphasis? (Cf. Lu, op. cit. p. 40.)

<sup>4/</sup> See for further details Kiefer 1967. pp. 4-53.



transformational rules into the base that attach Emph just to the right node. The effect of these transformational rules would be something like this



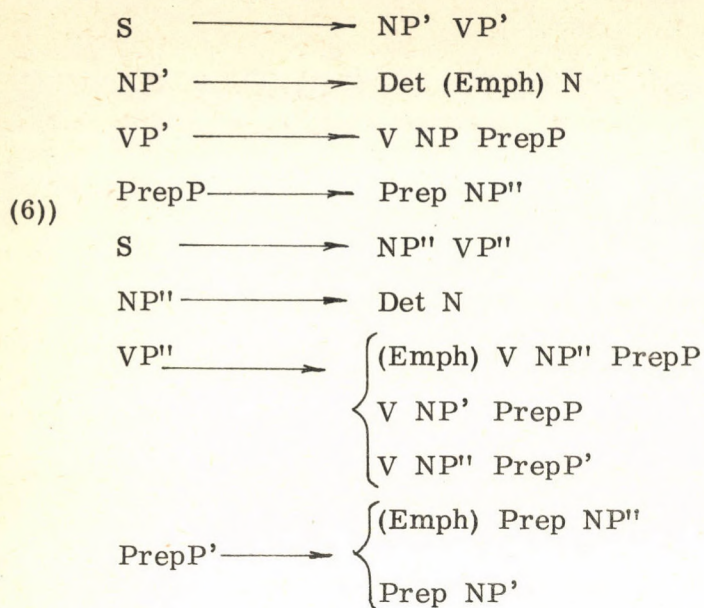
These transformations, called *attachement transformations* (cf. Kuroda) can perform the task desired. They are, however, against the principle that no transformational rules should work in the base. If we allow transformational rules to operate in the base, then I see no way of drawing a line between deep and surface structure, between the base of a grammar and its transformational component.<sup>5/</sup>

In view of the above one should make use of transformational rules in the base only if no other possibility is available. As far as I know such a strong argument could not be found up to now.

A further possibility would be to generate the correct structures for emphatic sentences by means of categorial (phrase structure) rules. This can be done if we increase the number of rules, for example, in the following way

<sup>5/</sup> One cannot argue here that transformational rules are needed anyway in the base if one accepts the more plausible alternative of the organisation of the base propounded in Chomsky 1965. The lexical insertion rules are, in fact, transformational rules but transformational rules of a special kind. They do not alter the syntactic structure of the sentence structures generated by the base.





It is clear that it is not a system of rules like (6) that we want. In the case of more constituents the number of rules grows still more rapidly. Therefore, at least for the time being, one must drop the claim that **Emph** should be introduced by categorial rules. It would seem that (1) is still the best way of handling emphasis.<sup>6/</sup>

The lexical items that cannot take emphasis will be marked so in their lexical characterization. Thus we must add to (1) the following condition:

Condition 1. Emphasis must not be derived from structures like (1) if this contains in  $S_1$  in the position corresponding to the node labelled "unspecified" a lexical item that is negatively specified with respect to emphasis.

Apart from Condition 1. further conditions must be imposed on (1). These will refer to the deletability of  $S_2$  in (1).

Compare the following types of emphatic sentences with each other:

<sup>6/</sup> It is conceivable that a generative semantic approach will offer us new possibilities. At present it would, however, be too early to put forward sane proposals to this effect. The scattered remarks and ideas about a generative semantics have not been developed to a full-fledged theory as yet. (Cf., for example, Bierwisch, McCawley.)



Your handbag is light and not heavy.

(7) This problem is difficult and not easy.

The boy is fat and not meagre.

John is at home and not Paul.

(8) The mail arrived and not the newspaper.

He read the book and not the newspaper.

The book is mine and not yours.

(9) He is going to do the job and not me.

I will read this book and not that one.

In (7) the adjectives form antonymous pairs, the assertion of one of them implies the negation of the other and conversely, the negation of one of them implies the assertion of the other. Here  $S_2$  can be deleted because the lexical entry referred to as "unspecified" can be inferred on the basis of  $S_1$ . In (9) the scope of the contrast is well-defined but not the particular instance. In the first sentence of (9) "mine" may also contrast with "his, her, ours, yours, theirs". In other words, the deletion of  $S_2$  is only justified under the presupposition that not only the scope but also the instance of the item "unspecified" is known. Similar conclusions can be drawn with respect to (8), though there the scope of emphasis is less clear. Now we are able to formulate the conditions of deletability for (1).

Condition 2. In (1)  $S_2$  can be deleted only if either (a) the item "unspecified" can uniquely be inferred from the corresponding term in  $S_1$  or (b) the presuppositions arising from the situational context of (1) allow this inference.

I cannot go into the question of presuppositions in more detail here (cf. Fillmore 1968 b) but it is plain that presuppositions play an essential role in the case of deletion.<sup>7/</sup>

<sup>7/</sup> It should be made clear that presuppositions are also necessary for the general interpretation of emphatic sentences. Notice, that no emphatic sentence can initiate a discourse. Each emphatic sentence presupposes a certain kind of discourse (cf. Kiefer 1967. pp. 122-155.). So far as I can see, the theory of sentence presuppositions is a very promising step towards a new theory of discourse.



The structure (1) together with Condition 1. must be considered what Perlmutter calls output conditions of the base (of. Perlmutter). Such conditions are necessary for various reasons.

In most cases, emphasis requires in free-word order languages special word order. Hence Condition 2. cannot be lumped together with Condition 1. (1) and Condition 2. form an output condition on the transformational component of grammar.

To sum up, a structure like (1) and Conditions 1. and 2. will account for the syntactic description of emphasis except for the word order changes that will be described in a subsequent section.<sup>8/</sup>

3. While the syntactic description of emphasis is relatively novel, the topic-comment relation has been investigated for some time, especially by the Prague linguists (Mathesius, Daneš, Dokulil, Firbas, Novak, Sgall - to mention only some of them). The topic-comment relation, called functional sentence perspective in the Prague school, was also used as an argument to set up a stratificational model of language (cf., for example, Sgall). The functional sentence perspective was separated from the syntactic description of sentences on the ground that it does not tell us anything about the syntactic structure but about the ways in which a given syntactic structure is used in the process of communication. Thus, word order is conceived as belonging to the pragmatic level of linguistic description rather than to the syntactic structure proper. In the process of communication each communicational unit (which is not necessarily a sentence) is supposed to convey new information. Moreover, if language is aptly used, each communicational unit adds new information to something that is already known. Hence communicational units can be split up into two types of information: the novel information is conveyed by the comment/s, the already known information is the topic of the communicational unit. This distinction is indicated in language by stress (hence intonation)

<sup>8/</sup> This is not all that I can say and that can be said about emphasis, of course. Here I had to content myself with a brief summary of my earlier and present thoughts about this topic. A more detailed discussion would go beyond the scope of the present paper.



word order, etc. We may now ask how the topic-comment relation can be accounted for in a generative description of language. There have already been some attempts to account for this relation in the framework of generative grammar (Staal, Dezső-Szépe). I shall comment upon these later on. First let me point out that the topic-comment relation is in several points akin to emphasis. At the same time it differs, essentially, from the latter.

(1) First of all, there is no way to describe the topic-comment relation by means of a structure like (1) because

- (a) to be a comment does not involve negation;
- (b) it does not involve a parallel structure.

Notice, for instance, that the topic of a sentence<sup>9/</sup> is by no means determined by its position in the sentence. Consider, for example,

- (i) I saw Sally to read a book.
- (10) (ii) She read an interesting book.

In (10)(i) under the presuppositions that we do not initiate a discourse by it and that "I saw Sally" is already known, we may say that "to read a book" is the comment.

(2) In languages with free word order the impact of the topic-comment relation is in general different from that of emphasis. In Hungarian, for example, in emphatic sentences the emphatic constituent must always precede the verb and in the case of verbs with particle the particle must come after the verb (except for the case when the verb itself receives emphasis). Thus, one has

A postára János vitte el a levelet.

the post-office to John carried-he away the letter

It was John who took the letter to the postoffice.

János a postára elvitte a levelet.

János elvitte a postára a levelet.

etc.

where the comment can be "a postára", "elvitte a levelet", "a levelet" etc. In other words, comment is much less determined by word order than emphasis.

<sup>9/</sup> For simplicity's sake we shall identify the communicational unit with "sentence", though this is not necessarily so.



The stress pattern is more significative from the point of view of comment than word order, though the latter can also be decisive (see below).

(3) It is clear that in emphatic sentences the emphatic constituent is supposed to convey the novel information. Therefore, the emphatic constituent must be considered the comment of the sentence. Thus, every emphatic constituent is the comment of a sentence but not every comment is emphatic.

(4) Emphasis affects meaning, not so the topic-comment relation. If we follow Fillmore's distinction between meaning and presupposition (cf. Fillmore 1968 b) then we may say that negation (and also question, command etc.) affects the meaning of sentences but not their presuppositions. But we have (at least) two kinds of presuppositions, one concerns the presuppositions originating in the lexical items of the sentence, the other comes from the sentence structure. Fillmore's distinction holds for the former type of presuppositions but not for the latter. Take, for instance, Fillmore's example

- (11) (i) Open the door!  
(ii) Don't open the door!

The presuppositions concerning "open" and "door" and "you" are the same for (11) (i) and (ii). At the same time, the meanings of (i) and (ii) are different. On the other hand, however, if I say

- (12) (i) I open the door.  
(ii) I do not open the door.

Here there will be a difference in presuppositions as well, the sentence (12)(ii) being an answer to a question or command, i.e. it cannot initiate a discourse while (12)(i) can. In this sense some meaning changing operations will also change the presupposition. Since the topic-comment relation will only affect the stress pattern or word order etc. of sentences, only the presuppositions with respect to the possible positions in the communication process of the given sentence will change but not their meaning. In the case of emphasis, however, besides some changes in the presuppositions also the meaning of the sentence will change.

I cannot but agree with Sgall who says that "the functional sentence perspective, as well as the means of its realization, has a systematic character and a full description of a language system as a system of "forms"



and "function" is not possible without respecting it." (Sgall, op. cit. p. 206.) In other words, the topic-comment relation must be accounted for in a competence model of language. Let us say that the topic of a sentence is the/unmarked category/and the comment is the/marked category. Furthermore, let us denote comment by Com. The question is now how Com is to be assigned to a sentence or its parts.

To begin with, we must seek an answer to the following questions: (i) What can be Com? (ii) How many Com's can occur in a sentence?

As to (i) it is easy to see that a sentence that initiates a discourse is a Com. Hence Com can be assigned to whole sentences. On the other hand, it has been observed that there is a close connection between questions and Com's (of., for example, Hatcher). Moreover, it has been stated that Com can be assigned only to constituents that can be questioned (Staal). It is a well-known fact that one questions mostly noun phrases.<sup>10/</sup>

It seems to me, however, that not only must Com be assigned to sentences in some cases but also to verb phrases. In fact, one can ask a question about the whole verb phrase or about the activity, state etc. expressed by the verb. Take, for instance, the sentences

- |      |                     |                                                    |
|------|---------------------|----------------------------------------------------|
|      | What are you doing? | I <u>am walking</u> .                              |
| (13) | What are you doing? | I <u>am writing a letter</u> .                     |
|      | What are you doing? | I <u>am feeding my dog with bread and butter</u> . |

versus

- |      |                                            |                                                    |
|------|--------------------------------------------|----------------------------------------------------|
|      | What are you writing?                      | I am writing <u>a letter</u> .                     |
| (14) | What are you feeding your dog with?        |                                                    |
|      |                                            | I am feeding my dog <u>with bread and butter</u> . |
|      | Who are you feeding with bread and butter? |                                                    |
|      |                                            | I am feeding <u>my dog</u> with bread and butter.  |

<sup>10/</sup> Katz and Postal even argue that one can question only noun phrases. Cf. Katz-Postal pp. 79-120.



In (13) we are questioning the verb phrase, hence in the answer the verb phrases are to be considered as Com's (underlined with staggered lines). In (14) we are asking about some parts of the verb phrase.

We leave deliberately the problem open of how questions are treated in the base. It might very well be that grammatical categories that appear at the surface as nouns, verbs, adjectives etc. must be considered a single category in the base component. It is clear, however, that (13) and (14) must be handled differently.

As to (ii) we must, once again, disagree with Staal who claims that every sentence has only one topic. Even if we restrict ourselves to simple sentences, this is not true. This is simply a consequence of the symmetrical behavior of topic and comment. That more than one Com can occur in a simple sentence is without any doubt because of questions like

Who is going to kill whom?

Who sends what to whom?

Let us now take a sentence  $S_1$  with three Com's and one topic. This should be followed in the discourse by  $S_2$  that repeats the three Com's of  $S_1$  and replaces the topic of  $S_1$  by something else. Thus,  $S_2$  will have three topics and one Com. If we agree, however, that topic is the unmarked category, we do not have to bother about it. It is enough if we can manage somehow to indicate the possible or necessary Com's for a given sentence.

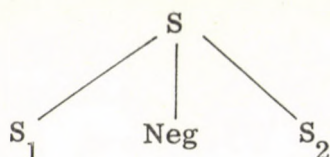
Several authors have pointed out that Com determines the place of break (or breaks) and the stress pattern in a sentence. Thus, Com plays an essential role in determining the phonological structure of sentences (Daneš, Pala, Dezső-Szépe. Elekfi a.o.). This means that the assignement of Com must come before the phonological component is put to work. A rule like

(15)  $S \longrightarrow (\text{Com}) \dots \text{NP VP}$

would obscure the issue because Com does not affect the meaning proper of sentences. Furthermore, Com depends in many languages on word order and the presuppositions the speaker-listener makes about some order. Therefore,



(15) does not tell us anything about what is really going on. Here it is quite impossible to set up a structure like



as in the case of emphasis because on the basis of Com the syntectic structure of the preceding sentence in the discourse cannot be inferred (as it can in the case of emphasis). On the basis of Com we can only state what information has already been previously mentioned.

As far as I can see, in view of the structure of grammar as presently conceived, Com should be assigned to sentences or their parts after the word order rules have already done their work. Com can then be attached to P-markers by means of transformational rules. These rules can be either obligatory or optional. They are obligatory in case of emphasis or in case of some word order. If you say in German

Den Abendstern sah er.

then one has either

Den Abendstern sah er und nicht den Mond.

i.e. with emphasis on "Abendstern" or

Er sah einen Stern am Himmel. Den Abendstern sah er.

where "Abendstern" is to be assigned Com but not Emph.

The assignment of Com entails different presuppositions. These must be accounted for by a theory of presuppositions that will tell us how sentences with various presuppositions should be interpreted. Is this to mean that also surface structures must be interpreted semantically? This seems to me mainly a terminological issue. Nevertheless, sentences must be interpreted with respect to their presuppositions but I am completely in the dark as to how this should be done.<sup>11/</sup>

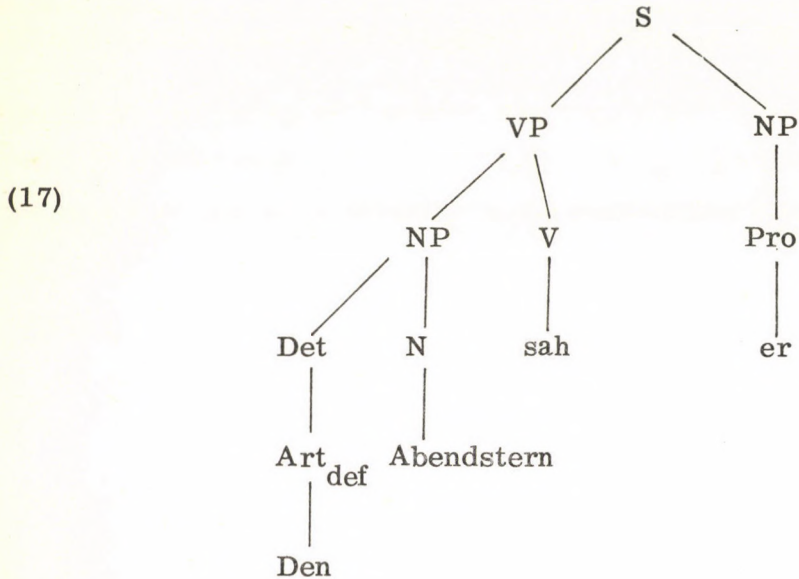
<sup>11/</sup> It may be that a new component will be needed, a sort of pragmatical component, that will take care of these phenomena. It is also possible that a "revised" stratificational approach will suit better our purpose. At present, however, all this cannot be more than mere speculation.



In order to illustrate what am after let me take the sentence

(16) Den Abendstern sah er.

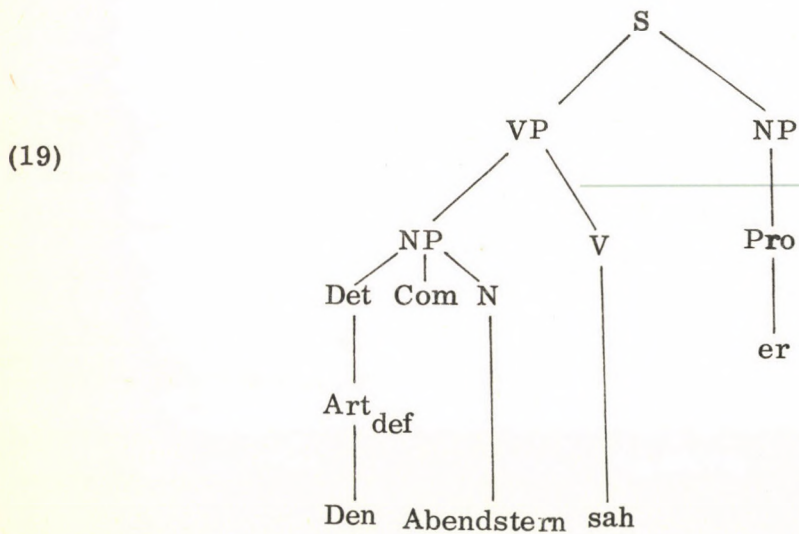
again. Let us now leave aside emphasis. The (surface) structure of (16) is approximately



Here the transformational rule

(18) 
$$\begin{array}{ccccccc} \text{Art}_{\text{def}} & \text{N} & \text{V} & \text{NP} & \Longrightarrow & 1 & \text{Com} & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 & & & & & & \end{array}$$

obligatorily applies. We obtain the structure





On interpreting (19) one must assign to (19) the presupposition that whenever it is uttered everything except for "Abendstern" is taken for already being known (from a previous sentence or by some other means). What we want to communicate is the fact that we are talking about something called "Abendstern". Everything else in the sentence is not important, it is important to the extent only that it is necessary in order to express ourselves in some linguistically acceptable way.

In languages with much freer word order than German (Hungarian, Russian etc.) the rule (18) would not apply obligatorily but the interpretation of (19) would be the same.

4. What has been said so far has an important bearing on word order. I have already pointed out that in Hungarian "emphatic" word order reveals two important features:

- (i) the emphatic constituent must always precede the verb (except, of course, if the verb itself is emphatic); as a consequence a sentence where the verb stands in initial position can only be emphatic if the verb receives emphasis.
- (ii) in case of verbs with particle the particle must come after the verb (once again, with the exception of emphatic verbs).<sup>12/</sup>

It is not possible here to discuss in detail the word order regularities of emphatic sentences in Hungarian.<sup>13/</sup> It should be made clear, however, that some word order rules must apply obligatorily, others are optional like the rules that will account for the following changes

<sup>12/</sup> This is, by the way, a good proof for emphasis. It enables us even to differentiate between contrastive and emphatic stress. Take, for example, the sentence.

Peter is working in Vienna and John in Paris.

where "Peter-John" and "Vienna-Paris" receive so-called contrastive stress. In Hungarian if we take a verb with particle the order is Part V in case of contrastive stress but V Part in case of emphasis.

<sup>13/</sup> This has been done in Kiefer 1967 at some length.



- Tegnap Péterrel találkoztam az utcán.  
 yesterday Peter-with met-I the street-on  
 (20) Yesterday I met Peter in the street.  
Péterrel találkoztam tegnap az utcán.  
 Az utcán tegnap Péterrel találkoztam.  
 Tegnap az utcán Péterrel találkoztam.  
 etc.

The changes in (20) seem to be mere stylistic variants, their semantic-pragmatic interpretation remains apparently unchanged.

A similar situation seems to hold with respect to Com. In general the observation has been made that if the direct object is moved from the postverbial to a preverbial position it must be either emphatic or it is a Com.

- Az autódat láttam az utcán.  
 the car-yours saw-I the street-in  
 (21) Péter levelet irt anyjának.  
 Peter letter wrote-he mother-his-to

and also (16) are pertinent examples.

Here, too, some changes can be considered with good reason to be mere stylistic ones.

- Péter anyjának levelet irt.  
 Anyjának levelet irt Péter.

The obligatory changes in case of Com are, however, less clear than with emphatic sentences.

It seems safe to conclude that any work on word order must take into account both Emph and Com and that some of the word order rules will be obligatory and others will be optional. Before concluding our remarks on word order we must take up the following problem. In languages with free word order is there any distinguished word order that can be considered to be basic? For basic word order we may stipulate the following definition. A word order is referred to as basic if it can stand without any presupposition as to what should be considered as being already known. Thus, the word order represented



by either (20) or (21) cannot be considered to be basic. For (20) and (21) one would, quite intuitively, set up the following basic orders

- (22) Tegnap találkoztam Péterrel az utcán.  
Láttam az autódat az utcán.<sup>14/</sup>

Alongside (22), however, one may also have

- (23) Tegnap találkoztam az utcán Péterrel.  
Találkoztam az utcán Péterrel tegnap.  
Láttam az utcán az autódat.  
etc.

This may be partly due to the free positional status of some of the adverbials (in the first place of those of time and place). It can easily be shown that as soon as some other categories are involved in these changes, presuppositions about the linguistic context will emerge. The sentence

- (24) Az autódat láttam az utcán,

will be felt to be incomplete (without Emph or Com). We would ask "And what" - "és?" and a possible answer would be:

És megcsodáltam.  
and admired-I-it

Some changes will be felt to be ungrammatical:

Olvas levelet János.  
read letter John  
John is reading a letter.

Az autódat az utcán láttam.

Thus, except for some of the adverbials, it seems to be possible to establish a unique basic order which will be generated by the base component of grammar.

<sup>14/</sup> I have left out the third sentence deliberately. Intuitively, "Péter levelet irt anyjának" seems to be the basic order but here Com (or Emph) is obligatory. There are, at least in Hungarian, many sentence structures that trigger off Com or Emph. This indicates, that the definition for basic order must be modified in some way in order to cover these cases as well.



All changes will be carried out by transformational rules in the transformational component according to the principles put forward in the preceding discussion.

We may conclude this extremely sketchy discussion, that hardly touches upon all important questions of word order and leaves even many of the discussed problems open, that even in languages with so-called free word order word order is far from being free. This is a fact, that can only be brought to the fore by means of a more subtle analysis than has been undertaken so far.<sup>15/</sup>

### Bibliography

- M. BIERWISCH: Towards a Generative Semantics, Berlin 1968  
(mimeographed)
- N. CHOMSKY: Syntactic Structures, Mouton and Co., The Hague, 1957
- N. CHOMSKY: Aspects of the Theory of Syntax, The M.I.T. Press, Cambridge  
(Mass.), 1965
- FR. DANEŠ: A Three-Level Approach to Syntax, TLP 1 (1964), pp. 225-240.
- FR. DANEŠ: K otázce pořádku slov v slovanských jazycích, SaS 20 (1958), pp.  
1-10.
- L. DEZSŐ: Notes on the Word Order of Hungarian Simple Sentences,  
Computational Linguistics 4 (1965), pp. 3-59.
- L. DEZSŐ: Typologische Besonderheiten der ungarischen Wortfolge, Acta  
Linguistica Hung. 18 (1968), pp. 125-159.
- L. DEZSŐ - GY. SZÉPE: Adalékok a topic-comment problémához, Nyelvtudo-  
mányi Közlemények LXIX (1967), pp. 365-388.
- M. DOKULIL - FR. DANEŠ: K tzv. významové a mluvnické stavbě věty,  
O vědeckém poznání soudobých jazyků (1958), pp. 231-246.
- L. ELEKFI: Kriterien der aktuellen Satzgliederung in ungarischen Kernsätzen,  
manuscript

<sup>15/</sup> I will discuss the relation between topic-comment and presuppositions in a separate paper. Cf. Klefer, forthcoming.



- Ch. FILLMORE: The Case for Case, in: Universals in Linguistic Theory, Holt, Rinehart and Winston, forthcoming 1968 a
- Ch. FILLMORE: Types of Lexical Information, Paper presented at the Conference on Mathematical Linguistics, Balatonszabadi (Hungary), September 7-10, 1968 b
- J. FIRBAS: Notes on the Function of the Sentence in the Act of Communication, in: Sborník prací filos. fak. brněnské university XI, A 10 (1962), pp. 133-148.
- J. FIRBAS: On Defining the Theme in Functional Sentence Analysis, TLP 1 (1964), pp. 267-280.
- A. G. HATCHER: Theme and Unerlying Question, Supplement to Word 12 (1961).
- K. E. HEIDOLPH: Kontextbeziehungen zwischen Sätzen in einer Generativen Grammatik, Kybernetika 2 (1966), pp. 274-281.
- J. J. KATZ - P. POSTAL: An Integrated Theory of Linguistic Descriptions, The M.I.T. Press (1964).
- F. KIEFER: On Emphasis and Word Order in Hungarian, Indiana University, Uralic and Altaic Series Vol. 76 (1967).
- F. KIEFER: Topic-comment and Presuppositions, in preparation.
- S. - Y. KURODA: Attachment Transformations, M.I.T. (1968), (mimeographed).
- J.H.T. LU: Contrastive Stress and Emphatic Stress, The Ohio State University, Project Report No. 10 (1965).
- V. MATHESIUS: Zur Satzperspektive im modernen Englisch, Archiv für das Studium der neueren Sprachen und Literaturen 155 (1929), pp. 202-210.
- J. McCAWLEY: The Role of Semantics in a Grammar, Chicago University (1967), (mimeographed).
- P. NOVAK: On the Three-Level Approach to Syntax, TLP 2 (1966), pp. 219-223.
- K. PALA: O nekotorych problemach aktualnogo členenija, PSML 1, (1966) pp. 81-92.
- P. SGALL: Functional Sentence Perspective, PSML 2 (1967), pp. 203-225.
- J. F. STAAL: Some Semantic Relations between Sentoids, FL 3 (1967), pp. 66-88.







# THE TREATMENT OF NON-PROJECTIVE STRUCTURES IN THE SYNTACTIC ANALYSIS AND SYNTHESIS OF ENGLISH AND GERMAN

Jürgen Kunze

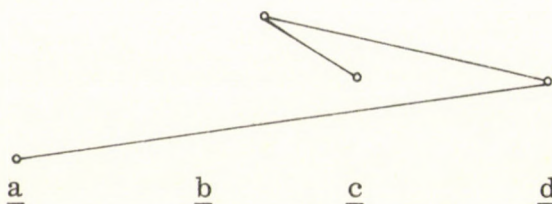
## 1. The concept of projectivity

For the purposes of this paper we represent the syntactic structure of sentences by means of a dependency grammar, and we do not take other types of grammar into account.

If we consider a particular language it is natural to ask which general constraints are satisfied by all the dependency trees occurring in the syntactic representation of the sentences of this language. This applies for example to the hypothesis of depth. The most important constraint of this kind is undoubtedly the demand of projectivity. It allows considerable simplification in syntactic algorithms. We shall return to this point later. However, this convenient simplification is confronted by the fact that in the languages we are considering there are sentences whose syntactic structures, according to an adequate concept of the (formal) term "dependency", are not projective. These sentences must be regarded as grammatically correct and do not represent any deviation from the grammar, for example:

<u>was</u>	<u>hat</u>	<u>er</u>	<u>getan</u> ?	( <u>what</u> <u>has</u> <u>he</u> <u>done</u> ?)
				( <u>what</u> <u>did</u> <u>he</u> <u>do</u> ?)

These three sentences have as a pure dependency tree the structure:





Since we can expect to find non-projective sentences in a scientific text it is necessary, at least in syntactic analysis, to account for them. However, since many sentences (of the examples above) can only have a word-order which leads to nonprojective structures, the analogous problem arises also for syntactic synthesis.

Before we turn to the exact definition of projectivity it is necessary to understand clearly the nature of this constraint. We first make the observation that strictly speaking we cannot talk about whether a sentence is projective or not. It is inadmissible for two reasons:

1. As is well-known there exists a certain amount of freedom in the conception of a particular dependency grammar for a particular language. This freedom concerns above all the (more heuristic) definition of dependency, i.e. one which determines consistently, adequately, and in a unified way which word is to be sub-ordinated to which other in a (syntactically unambiguous) sentence. Unfortunately no unique general criterion for dependency is yet known. It is therefore incorrect to speak of the dependency structure of a (syntactically unambiguous) sentence. Since there are widely differing conceptions of what dependency is, it is possible that one and the same sentence may be projective according to one conception and non-projective according to another. However if we are careful we can use this freedom to "make projective" a large number of sentences.

We adhere to the principle that projectivity is not a feature of sentences but of dependency trees. The definitive mapping:

(syntactically unambiguous) sentence                  dependency tree

is not wholly unequivocal. This should not be the case within a particular model.

2. If the definitive mapping is given then it still holds that in an adequate model we should be able to map several dependency structures onto a syntactically ambiguous sentence. It is then possible that one of these structures will be projective while the other is not. There are numerous examples of this. At the same time it is by no means the case that the projective structure is necessarily the more probable or the preferable syntactic interpretation of the sentence.



In order to be able to discuss projectivity we further need to consider the linear sequence of the nodes of the dependency tree. We can conceive of this sequence also as a component of the dependency tree. But there are a number of reasons for regarding the linear sequence and the dependency structure separately, i.e. for regarding the tree as unordered. The word order is actually one of the kinds of sentence relation which belong to a different linguistic level than do the syntactic structures. In this connection projectivity takes the form of a general principle of organisation for the sentence relation "linear sequence." From amongst all the possible sequences of nodes of the tree certain ones are selected of which only some are grammatically correct. This is analogous to the fact that (sufficiently abstract) syntactic relations in the sentence can be rendered in different ways, for instance certain objects either by case or by a prepositional adjunct.

We now proceed to give an exact definition of projectivity which is particularly suited to a generalisation to be undertaken later. Some explanation is necessary. If  $x$  is a node of a tree  $\mathcal{L}$ , then we call  $\underline{A}(x)$  the set of all nodes which are indirectly dependent on  $x$ , including  $x$  itself. In our example above the following holds:

$$\underline{A}(b) = \{\underline{a}, \underline{b}, \underline{c}, \underline{d}\}, \quad \underline{A}(d) = \{\underline{a}, \underline{d}\}, \quad \underline{A}(c) = \{\underline{c}\}$$

In an ordered tree  $\mathcal{L}$  let  $\underline{n}(x)$  be the position number of the node  $x$ . In our example  $\underline{n}(a) = 1, \dots, \underline{n}(d) = 4$ . A sub-set  $\underline{E}$  of the set  $\underline{B}$  of all nodes of  $\mathcal{L}$  is called a section if it contains together with two nodes all those which lie between. In other words if  $\underline{E}, \underline{y} \in \underline{E}, \underline{z} \in \underline{B}$  and  $\underline{n}(x) < \underline{n}(z) < \underline{n}(y)$ , then  $\underline{z} \in \underline{E}$ . In our example  $\{\underline{a}, \underline{b}, \underline{c}\}$  is a section and  $\{\underline{a}, \underline{d}\}$  is not.

An ordered tree  $\mathcal{L}$  is called projective if for all  $\underline{B}$  the set  $\underline{A}(\ )$  is a segment. In our example  $\underline{A}(a)$ ,  $\underline{A}(b)$  and  $\underline{A}(c)$  are sections, but  $\underline{A}(d)$  is not a section. It can be proved that this definition is equivalent to those of Fitialov, Hays and others. It implies that all constituents are continuous.



## 2. Simplifications resulting from projectivity

We consider here two possibilities of simplifying syntactic algorithms on the presupposition that all the sentences discussed have projective dependency structures.

First we consider the syntactic analysis. In the successive construction of a dependency tree during the application of the individual rules the words of the sentence are examined from the point of view of certain conditions, after the fulfillment of which a subordination is to take place. Assuming that by such a rule a pair  $\underline{a}, \underline{b}$  of words have been found of which one is to be subordinated to the other. Of course  $\underline{a}$  and  $\underline{b}$  are not usually adjacent, so that the problem arises of designating the possible strings of words between  $\underline{a}$  and  $\underline{b}$ . An enumeration of these strings in explicit or recursive form cannot normally be carried out and they would in many cases be too numerous. However, from projectivity we can easily obtain a simple, necessary condition: if the "finished" tree contains the subordination of  $\underline{b}$  to  $\underline{a}$  and if this tree is projective, then all word-forms  $\underline{c}$  between  $\underline{a}$  and  $\underline{b}$  must be contained in  $\underline{A}(\underline{a})$ . But this does not necessarily mean that by the time the step in the analytic process is taken by which  $\underline{b}$  is subordinated to  $\underline{a}$  the words  $\underline{c}$  already have to have been subordinated. On the other hand this can be achieved by applying the rules step by step in a particular sequence.

The rules can be grouped in such a way that the following two principles hold: let  $\mathcal{L}$  be the "finished" tree obtained after the syntactic analysis, while  $\mathcal{L}^*$  contains precisely those subordinations already undertaken by the analytic step we are considering, the tree  $\mathcal{L}^*$  is, so to speak, "part" of  $\mathcal{L}$ .

1. (Principle of "upward analysis.") A word  $\underline{b}$  is not subordinate to a word  $\underline{a}$  unless all words  $\underline{c}$  which are directly subordinate to  $\underline{b}$  in  $\mathcal{L}$  are directly subordinate to  $\underline{b}$  in  $\mathcal{L}^*$  also. The result of this is immediately obvious: a word  $\underline{b}$  only become subordinate to a word  $\underline{a}$  when all words contained in  $\underline{A}(\underline{b}) \setminus \{\underline{b}\}$  contained in are already subordinate in  $\mathcal{L}^*$ .

2. (Principle of "outward analysis.") A word  $\underline{b}$  does not become subordinate to a word  $\underline{a}$  unless all words  $\underline{c}$  occurring between  $\underline{a}$  and  $\underline{b}$  and directly subordinate to the word  $\underline{a}$  in  $\mathcal{L}$  are directly subordinate to the word  $\underline{a}$  in  $\mathcal{L}^*$  also.



Principles 1. and 2., assuming that  $\mathcal{L}$  is projective, allow as a logical consequence the following condition:

3. A word b only becomes subordinate to a word a if all words c occurring between a and b are already subordinate - in fact subordinate to words occurring between a and b (inclusive).

Clearly, while condition 3. is formally simple on the one hand, it is on the other hand still sufficient to avoid many wrong subordinations. For its application it is necessary that 1. and 2 hold, which again necessitates a certain ordering of rules. This ordering can however be conveniently arranged by directing the repetition of identical rules through suitable cycles. This enables the analysis to proceed flexibly, superfluous repetitions being avoided. An analysis which proceeds in this way is at the same time very convenient for the treatment of syntactically ambiguous sentences (i.e. those admitting more than one result of analysis). We shall not dwell longer on this problem, since it does not directly concern that of projectivity.

We shall now turn briefly to syntactic synthesis, in particular the problem of obtaining the word-order of the target language. In this connection the following holds: if all the sequences we wish to obtain are such that the corresponding trees become projective, then it is sufficient to determine the relative positions of words directly subordinate to each other. We thus only need to consider "bundles" which state in which sequence the words b occur which are directly subordinate to a word a, and in which position (before, between or after them) the word a can stand. In this way it becomes considerably simpler to produce the word order.

### 3. A possible generalisation of projectivity

There have already been some attempts to generalise projectivity ( 1 , 2 ). We do not wish to discuss them here since our generalisation bears no relation to them.



First we set up some conditions which a "reasonable" generalised projectivity must fulfill:

1. A generalised projectivity should have a simple formal definition, so that its validity for a given dependency structure can easily be tested.

2. It should be of such a kind that it is satisfied not only by all projective dependency structures but by many other (nonprojective) structures, without being satisfied by too many counter-intuitive structures. This means in particular that (considered extensionally) it should not be too weak.

3. It should as far as possible be capable of explaining the underlying reasons for and the immediate causes of the nonprojectivity of correct structures.

For the latter condition some explanation is first necessary. It is in general possible to state the underlying reasons of the non-projectivity of a correct structure. We shall deal here with a complex of reasons for many (probably even most) instances of nonprojectivity in German:

There are the compound predicates of the following types

(a) Formation of compound tenses

(aa) perfect: participle with one form of haben or sein

Das habe ich noch nicht getan (I haven't done that yet)

Für ihn ist kein Brief gekommen (No letter has arrived for him)

Heute hat ihn der Vater besucht (His father has been to see him today)

(ab) pluperfect: (corresponding exactly to the perfect)

(ac) future I: infinitive with a form of werden

Für ihn wird kein Brief kommen (No letter will arrive for him)

Morgen wird ihn der Vater besuchen (His father will come and see him tomorrow)

(ad) future II: (combination of (aa) and (ac))

(b) use of modal verbs: pure infinitive with forms of modal verbs

Das kann er nicht machen (He cannot do that)

Heute darf ihn der Vater besuchen (Today his father may come and see him)

(c) use of verbs demanding an infinitive with zu:

infinitive with zu with any form of these verbs:



Das braucht er nicht zu machen (He didn't need to do that)

Heute scheint ihn der Vater zu besuchen (His father seems to be coming  
to see him today)

(d) passive: participle with a form of werden:

Das wurde nicht besprochen (That wasn't discussed)

Anschliessend wurde über diesen Vorschlag im Beschluss gefasst

Then this resolution was passed

Dann wurde ein Brief an das Ministerium geschickt

Then a letter was sent to the ministry

(e) adjectives as predicative nouns: adjective with the copula:

Davon ist alles weitere abhängig (Everything else is dependent on it)

Jedoch war für eine Lösung des Problems ein grosser Aufwand nötig

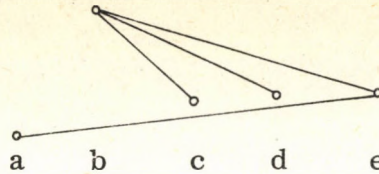
(Yet in order to solve the problem a greater effort was necessary)

We note that types (a) to (e) can be further combined. All the sentences given here besides many other instances, have non-projective dependency trees in a concept of dependency adequate to German.

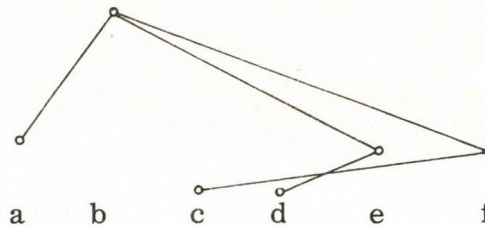
So far we have regarded the dependency trees  $\mathcal{L}$  as pure graphs. More precisely the trees were given by the set  $\underline{B}$  of their nodes and a one-place function  $\underline{m}$ , which maps onto every node of  $\underline{B}$  except one (the top of the tree  $\mathcal{L}$ ) its superordinated node. In our example at the beginning  $\underline{B} = \{\underline{a}, \underline{b}, \underline{c}, \underline{d}, .\}$   $\underline{b}$  is the top and  $\underline{m}(\underline{a}) = \underline{d}$ ,  $\underline{m}(\underline{c}) = \underline{b}$ ,  $\underline{m}(\underline{d}) = \underline{b}$ . We now enrich these structures by adding subordination features. The set  $\mathcal{U}$  of all subordination properties in a language forms a complete and sufficiently delicately categorised system of all the possible syntactic functions and relations in that language. A subordination property from  $\mathcal{U}$  is now mapped onto each node  $\mathcal{X}$  of the tree  $\mathcal{L}$ . This is property called  $\underline{C}(\mathcal{X})$  and it represents the syntactic relation between and the node  $\underline{m}(\mathcal{X})$  to which it is subordinated. For the top,  $\underline{z}$ , of the tree  $\mathcal{L}$  we proceed somewhat differently.

Since the exact structure of the system  $\mathcal{U}$  is irrelevant to the following considerations we shall not consider it further here. If the system  $\mathcal{U}$  is sufficiently delicate then the types (a) - (e) given below (besides others) form a certain sub-set  $\mathcal{H}$  of  $\mathcal{U}$ . This means that these relations are precisely accounted for by the set  $\mathcal{H} \subseteq \mathcal{U}$ . Consider the two examples:





$\alpha$  / Das habe ich nicht getan (I haven't done that) (perfect)



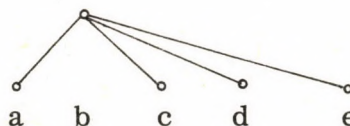
$\beta$  / Heute hat ihn der Vater besucht (Today his father has come to see  
him) (perfect)

The two corresponding structures are non-projective. The following ones have projective structures:

/ Das wusste ich nicht (I didn't know (that)) (simple past)

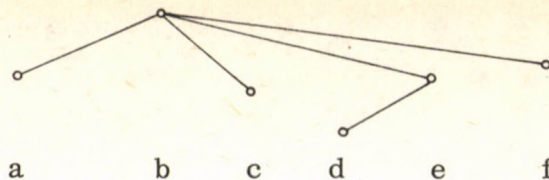
/ Gestern besuchte ihn der Vater (Yesterday his father came to see  
him) (simple past)

The structures of these two sentences are obtained by omitting the nodes e and f in / and / respectively, and by attaching to b all nodes dependent on them. But projective trees are also obtained by leaving all nodes in / and / intact and attaching to b all those nodes dependent on e and f in / and / respectively. Thus:



/ Das habe ich nicht getan





<sup>1</sup>/ Heute hat ihn der Vater besucht

The same result is achieved in all the other types mentioned and in any combination of the types (a) to (e). More generally this procedure can be formulated thus: if a node  $x$  is directly dependent on a node  $y$  for which  $C(y) \mathcal{H}$  holds, then  $x$  is directly subordinated to the first node  $w$  in the sequence  $y, \underline{m}(y), \underline{m}(\underline{m}(y)), \dots$  for which  $C(w) \mathcal{H}$ . In this way a projective structure is obtained.

These heuristic considerations may be stated exactly by the following definition: We define first inductively an operation  $\bar{x}$  for all nodes  $x \in B$ :

If  $z$  is the top of  $\mathcal{L}$ , then let  $\bar{z} = z$

If  $x$  which is not the top is a node of  $\mathcal{L}$ , then let

$$\bar{x} = \begin{cases} x & \text{if } C(x) \notin \mathcal{H}, \\ \underline{m}(x) & \text{if } C(x) \in \mathcal{H}. \end{cases}$$

By this definition  $\bar{x}$  is accounted for for all  $x \in B$ : if it is applied first for  $z$ , then for all nodes directly dependent on  $z$ , then for all nodes directly dependent on these, etc., then clearly in every case

$$x \in \underline{A}(\bar{x})$$

In example a/ since  $C(e) \mathcal{H}$  and  $C(a), C(b), C(c), C(d) \mathcal{H}$ , we obtain

$$\bar{a} = a, \bar{b} = b, \bar{c} = c, \bar{d} = d, \bar{e} = e.$$

The generalisation we proposed to make is now: if  $\mathcal{H}$  is any sub-set of the set  $\mathcal{V}$  of all subordination properties, then  $\mathcal{L}$  is called  $\mathcal{H}$ -projective if,



for every  $x \in \underline{B}$ ,  $\underline{A}(x)$  is a section. (Note that the operation  $\bar{x}$  is dependent on  $\mathcal{H}$ .) In example  $\mathcal{L} /$  we have:  $\underline{A}(a)$ ,  $\underline{A}(b)$ ,  $\underline{A}(c)$  and  $\underline{A}(d)$  are sections, and for  $x = e$ ,  $\underline{A}(b)$  (and not  $\underline{A}(e)$ !) must be a section. Thus this structure is  $\mathcal{H}$ -projective!

Since the set  $\mathcal{H}$  is a priori indeterminate we are dealing basically with a large number of generalisations: we obtain a different one for each set  $\mathcal{H} \subseteq \mathcal{V}$ . The essential thing however is that one can choose the set  $\mathcal{H}$  from a linguistic point of view in a specific way. It is clear that the following propositions hold: if  $\mathcal{H}_1 \subseteq \mathcal{H}_2$ , then, for every tree  $\mathcal{L}$ : if  $\mathcal{L}$  is  $\mathcal{H}_1$ -projective then  $\mathcal{L}$  is also  $\mathcal{H}_2$ -projective.  $\mathcal{L}$  is projective (in the usual sense) iff  $\mathcal{L}$  is  $\emptyset$ -projective (i.e.  $\mathcal{H}$  can then be selected as the empty set.) Furthermore any tree  $\mathcal{L}$  is  $\mathcal{V}$ -projective.

Finally we turn to the practical application of  $\mathcal{H}$ -projectivity. For the following let  $\mathcal{H}$  be a fixed, appropriately selected subset of  $\mathcal{V}$ . If an unspecified tree  $\mathcal{L}$  is given, then we construct a tree  $\tilde{\mathcal{L}}$  in the following way: let  $\mathcal{L}$  have the set  $\underline{B}$  of nodes and the function  $\underline{m}(x)$ , which states which node is superordinated to every  $x \in \underline{B}$  in  $\mathcal{L}$ . In  $\tilde{\mathcal{L}}$ , the operation  $\bar{x}$  is defined by  $\mathcal{H}$ . Let  $\tilde{\mathcal{L}}$  have the same set  $\underline{B}$  of nodes and let the function  $\tilde{m}(x)$  be determined as follows:

$$\tilde{m}(x) = \underline{m}(x)$$

If we regard as  $\mathcal{L}$  the structure given under  $\underline{B}$  and  $\underline{m}$  then  $\tilde{\mathcal{L}}$  is that given under  $\underline{B}$  and  $\tilde{m}$ . It is clear that  $\tilde{\mathcal{L}}$  is uniquely determined by  $\mathcal{L}$ . Conversely it can be shown that on certain conditions, which are simple and can always be fulfilled, concerning  $\mathcal{H}$ ,  $\mathcal{L}$  is also uniquely determined by  $\tilde{\mathcal{L}}$ . (The case of syntactic ambiguity however remains unaffected!) Because of this one-to-one correspondence it is basically of no consequence which of the two structures we operate with, whether with the "grammatically sound" tree  $\mathcal{L}$  or with the purely formally altered tree  $\tilde{\mathcal{L}}$ . On the basis of the following proposition it is however advantageous to work with the tree  $\tilde{\mathcal{L}}$  during analysis and synthesis. For every tree  $\mathcal{L}$  it holds that  $\mathcal{L}$  is  $\mathcal{H}$ -projective if and only if  $\tilde{\mathcal{L}}$  is projective in the usual sense.



For the normal process of syntactic analysis the following scheme has much to recommend it: the result of the algorithm is the tree  $\tilde{\mathcal{L}}$  which is then transformed into  $\mathcal{L}$ , which is relatively simple. In analysis with the result  $\tilde{\mathcal{L}}$  all the advantages of projectivity in the usual sense can be used. The transition from  $\tilde{\mathcal{L}}$  to  $\mathcal{L}$  need not necessarily be unambiguous in the case of syntactically ambiguous sentences. This means however that the analytic algorithm can be relieved of the task of finding certain (not all!) ambiguities, since the ambiguities do not become manifest until the transition from  $\tilde{\mathcal{L}}$  to  $\mathcal{L}$ .

For syntactic synthesis, in particular for the production of word-order, it holds correspondingly that the tree  $\mathcal{L}$  (which is not necessary projective, but only  $\mathcal{H}$ -projective) which is the input, i.e. first transformed into  $\tilde{\mathcal{L}}$  (which is obviously possible!) and then the rules are applied exploiting fully the advantages afforded by projectivity.

#### Literature:

- [1] SOLOMON MARCUS: Sur la notion de projectivité, Zeitschr. f. math. Logik u. Grundlagen d. Math., 11 (1965), 181-192
- [2] J. A. ŠREJDER: Svojstva proektivnosti jazyka, NTI, 1964, No 8, 38-41







# NOTES ON THE SEMANTIC INTERPRETATION OF VERBAL WORKS OF ART

S. J. Petőfi

## Contents:

1. Introduction	81
2. Some general remarks on the structural linguistic analysis	81-83
3. About the place of semantics in the structural linguistic analysis of 'works of art'	83
4. About the definition of the 'linguistic communication unit'	84-89
5. About the semantic interpretation of the 'linguistic communication units'	90-99
1. The establishment of the phrase-markers	90-93
2. About the semantic features of the lexical units	93-96
3. About the semantic characteristics of the communication units	96-99
6. About the 'composition units'	99-103
7. Concluding remarks	103-104
8. Notes	104-105







## 1. Introduction

Roman Ingarden in his much quoted work Das literarische Kunstwerk (1931) presents a phenomenological description of the verbal work of art considering it as a complex structure consisting of various strata. His work could not get at that time an adequate reception which was due to a large extent to the state of literature oriented linguistics in the early thirties, too. Linguists of that time were mainly preoccupied with the analysis of one or another aspect of verbal works of art.

The change that has taken place since then in the view of language - and the inquiries into general semiotics - render it possible now to raise the question of describing the whole of the linguistic structure in a homogenous way. This has been made certain by the fact that the intensional way of defining linguistic units, after that of the phonological elements, has entered into the domains of syntax and semantics, too.

In our present paper we should like to deal with some problems of the semantic interpretation arising within a full-fledged description of language - constructed out of intensionally defined elements.

Before discussing the real subject of our paper let us touch briefly on the question of the structural linguistic analysis of verbal works of art.

## 2. Some general remarks on the structural linguistic analysis<sup>1</sup>

A 'verbal work of art' (later on it will be referred to simply as 'work of art') can be regarded as a single sign with a particular structure.

The signans of this sign alone is also a complex structure. It represents a part of reality, formed in a given way, and is at the same time the vehicle



of different secondary structures (of poetical, ideological or some other kind of character), too. It is this complex structure which denotes the signatum, that is, the artistic message.

In our present paper we shall disregard the secondary structures in favor of the analysis of the linguistic structure which is of primary importance. Inside the linguistic structure it seems useful to distinguish two sign-components: a linguistic proper and a - say - musical one. The linguistic sign-component is composed of a syntactic and a semantic subcomponent while the musical one consists of a phonetic and a rhythmic subcomponent. Both sign-components may be structured in a hierarchical and a linear way.

Due to its individual character and the number of its components the linguistic structure can be analyzed and described only with the help of a complex model. The analysis and the structure-description may proceed, to our mind, on the following line.

The first task is to establish the primary elements of the structure in both the linguistic and the musical sign-component. Let us label them linguistic and musical communication units. Next we have to analyze the internal structure of these units. The aim of the analysis is to assign to each communication unit a proper set of markers revealing their general and specific properties. With the help of these markers it is possible for us to characterize the way in which the 'communication units' are organized into 'composition units', and these again into higher units until reaching the integrity of the linguistic and the musical sign-component of the 'work of art'. In such a way it will be possible to disclose the different parallelisms between different points on separate layers, too.

After having analyzed the linguistic and musical sign-components separately we have to establish the elements of the 'work of art' as 'a whole' which are correlatively constructed pairs of the 'communication units' of one of the sign-components and by the respective segments of the other one. (It should be noted that in these 'correlative pairs' now the linguistic now the musical communication unit is dominating, and the respective segment in the other sign-component is not necessarily a communication unit, too.) On the basis of the elements of the 'work of art' we have to unite the markers



referring to the single sign-components as well as the indices of the disclosed connections. This superposed set of markers renders it possible to discover the linguistic-musical semantic network of the linguistic structure.

The description of the 'linguistic structure' means the description of this semantic network starting out of the 'work of art' as a whole. This description should serve as an appropriate base for the analysis both of the reality represented in the given 'work of art' and the different secondary structures.

### 3. About the place of semantics in the structural linguistic analysis of 'works of art'

The linguistic sign-component of a 'work of art' may be conceived as consisting of the following strata:

the whole linguistic structure of the 'work of art'

composition units of different complexity

linguistic communication units

communication unit parts (immediate constituents)

words

The task of the semantic analysis is the semantic characterization of the units of these hierarchy-levels.

Since they are the linguistic communication units which constitute the starting-point of our analysis of the linguistic sign-component the first task will be to disclose this hierarchy-level.

(The generative theory of language in the sense of Chomsky will form the basis of our analysis. In semantics we intend to profit also from the trend initiated by Katz et alia.)



#### 4. About the definition of the 'linguistic communication unit'

We regard the 'linguistic communication unit' as the elementary unit of the linguistic sign-component, instead of 'words' or 'images' or the 'sentences of the authors'. Our intention was to come closer thereby to the 'elements' of the 'represented reality' and, at the same time, to provide an identical way of approaching the individual linguistic constructions of the various authors.

The definition of the 'linguistic communication unit' must be offered by a theory serving the analysis of the linguistic sign-component. At present we do not give an exact definition of this term. However, on the basis of breaking up the poems presented below<sup>2</sup> into communication units, we should like to indicate at least some viewpoints of the definition proposed by us.

Here we should like to note that the English translation must be considered as a background information, though our endeavor was to preserve the original characteristics of the poems as far as it has been possible. We have used square brackets to indicate extra information as compared to the original text, and essential differences in word-order are also referred to.

##### 1.

##### Hangtalan

/1/ Egy forrás-tiszta őszi csepp  
szaladja végig a fekete ágat  
/2/ hizik /3/ ragyog /4/ remeg  
/5/ szólni akar szinte akár a szemed  
  
ha könnybelábad  
/6/ Ugy tele lett  
hogyan lebukott  
Nem adva semmi szót  
semmi jelet  
/7/ Ezzel is kevesebb  
amit ma tudok

##### Soundless

/1/ A spring-bright autumn drop  
is running along the black bough  
/2/ swells /3/ glitters /4/ trembles  
/5/ wants to speak, it seems, like your  
eyes  
  
when filling with tears  
/6/ Became so full  
that [it] tumbled down  
Giving no word  
no sign  
/7/ What I know today  
is also lessened hereby



Esti dal

/1/ E mai napot is

Zsákmányát a vadász

Fuvarát a kocsis

Fáradalmát az arató

Hazahoztam

/2/ Az ablakfény mely az éjben vigyáz:

a Jó

a Rosszban

A Ház:

ahol valaki hazavár

Ez a helyem

/3/ A bérem mégis egy falat halál

/4/ Nem a szív: a lét lett istentelen

/5/ Aludj velem

Evening song

/1/ I have brought this day

The hunter his bag

The coachman his load

The reaper his fatigue

Home, too

/2/ The light of the window which keeps vigil  
in the night:

the Good

in the Evil

The House:

where somebody waits for me to come home

This is my place

/3/ My pay is nevertheless a bit of death

/4/ Not the heart: life became godless

/5/ Sleep with me



Lábnyomok

- /1/ A két napos hóban jól láthatók  
a lábnyomok.
- /2/ Itt jött a férfi, széles és nehéz  
léptekkel, akár az elrendelés.
- /3/ És szembe - könnyedén, szaladva tán -  
a lány.
- /4/ Itt találkoztak. /5/ Aztán - hol a nyom?
- /6/ A völgy felé, a hó-fuvatokon  
csak az a sulyos férfi-láb  
megy már tovább,  
a mély havat mélyebbre törve még.
- /7/ És Ő, a kicsi könnyűség?
- /8/ Oh, bár örökre-  
tartó boldogság kapta volna ölbe!

Footprints

- /1/ In the two day old snow clearly can be seen  
the footprints.
- /2/ Here came the man, with wide and heavy  
steps, like predestination.
- /3/ And towards [him] - lightly, running maybe -  
the girl.
- /4/ Here they met. /5/ Then - where is the print?
- /6/ Towards the valley, on the snow-drifts  
only that weighty man's-foot  
goes ever onwards,  
crushing the deep snow even deeper.
- /7/ And [she], tiny lightness?
- /8/ Oh, if only ever-  
lasting happiness had taken [her] on [its] lap!



Derengés

/1/ Hajnali szél; /2/ megigazitja  
 zilált kontyát az ifju meggyfa  
 /3/ Elcsöppennek a csillagok  
 /4/ Itt eltűnik, kibukkan ott;  
       a hegyre fut;  
 leleplezi tervét az ut  
 /5/ Csillan tü, /6/ megnyíló  
       ablak,  
 /7/ kakasfarok, /8/ eke,  
       /9/ harmat;  
 /10/ szemed, /11/ a tó -  
 /12/ Járni kezd a Nap fogaskereke.

Dawn<sup>+</sup>

/1/ Morning breeze; /2/ the young morello-tree  
 adjusts its tousled knot of hair  
 /3/ The stars drop away  
 /4/ The road - disappears here, emerges there;  
       runs uphill;  
 reveals its plan  
 /5/ a pin, /6/ an opening  
       window,  
 /7/ a cock tail, /8/ a plough  
       /9/ the dew;  
 /10/ your eyes, /11/ the lake - flash -  
 /12/ The cog-wheel of the Sun begins to work.

+ In connection with this last poem it should be mentioned that in the original Hungarian text

the subject of 'communication unit' 4 /'the road'/ is placed after  
 the predicates, while

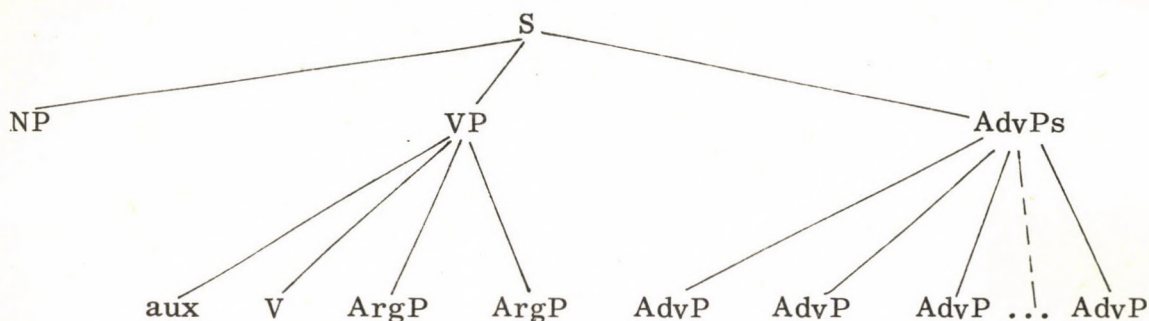
in 'communication units' 5-11 the predicate /'flash'/ stands before  
 the subjects.

These constructions are correct in Hungarian.



A 'linguistic communication unit' is - per definitionem - a continuous string of words of a given 'work of art', that is, the manifestation of a surface structure. (It should be noted that the surface structure, even in the case of one single sentence, can be the result of transformations carried out upon a set of underlying (deep) structures.)

That deep structure-pattern which can be considered as the most generally characteristic in Hungarian may be represented by the following tree-diagram (phrase-marker), down to the second level of its constituent structure analysis beneath the sentence-symbol /S/:



The symbol ArgP denotes the obligatory government of the verb /V/ while the AdvP symbols dominated by the AdvPs refer to the different adverbial complements.

When rewriting either the NP or the ArgP and AdvP symbols, the S symbol can reappear so, that the further rewritings can result in a 'complex phrase marker' possessing one single dominant S symbol.

Those NP, V, ArgP and AdvP symbols which are immediately dominated by the 'dominant S symbol' may be defined as the 'subject', the 'predicate', a certain 'government of the predicate' and a certain 'adverbial complement' of the sentence. Let us call them, under a collective designation, 'dominant immediate constituents'.

In the following, applying the terms introduced above, we should like to present some types of 'linguistic communication unit classes' remaining in the spheres of the poems analyzed by us.

1. One class of communication units may be conceived as being made up of those continuous strings of words to which can be assigned only one



single deep structure. (Such sentences are here in question which do not contain accumulated 'dominant immediate constituents' of the same character. - However, in the case of the subordinated part of a complex sentence this, too, is allowed.) For example, communication units 1.1, 1.6 (containing an accumulated 'object' in its 'adverbial complement of manner') and 3.6.

(Here we should like to note that all our remarks concerning the poems refer to the Hungarian text. The first number indicates the poem in question while the second one a communication unit of this poem.)

2. When breaking up into communication units a surface structure of such a kind which has been produced as a result of transformations carried out upon more than one deep structure the following main cases may occur.

2.1. If there emerge 'dominant immediate constituents' of the same character at the end of a continuous string of words that can be considered as a 'communication unit', each of these constituents, starting from the first repeated one, will be interpreted as a distinct communication unit. We have done so in the case of the string of words of the poem Dawn resulting finally in communication units 5-11. (Though not so clean-cut formulatable, the case is nearly the same with that string of words of the poem Soundless which has been decomposed into communication units 1-6.)

2.2. If the accumulation turns up at the beginning or in the middle of a string of words, the accumulating dominant immediate constituents will not constitute distinct communication units. That is the reason why, for example, 2.1 or 4.4 remain single communication units.

3. Finally, we have to consider as communication units those words or strings of words, too, which cannot be interpreted in the way outlined above and cannot be linked to one of the already disclosed communication units. This is the case, for example, with 4.1.

Thus, 'linguistic communication units' - even if consisting of one single word only - are 'units of meaning of full value' in a communication located between the limits of 'it begins' and 'it ends'. On the one hand they are linear, on the other, they express the simultaneity of the creator's vision and what he offers to be seen, too.



## 5. About the semantic interpretation of the 'linguistic communication units'

The semantic interpretation of the linguistic communication units is equal to the establishment and characterization of the relation-network of the lexical units given in terms of syntactic and semantic features. It is only possible if we have knowledge of the deep structure containing all relevant information. Thus, it is a primary task to discover those deep structures which belong to strings of words regarded as communication units.

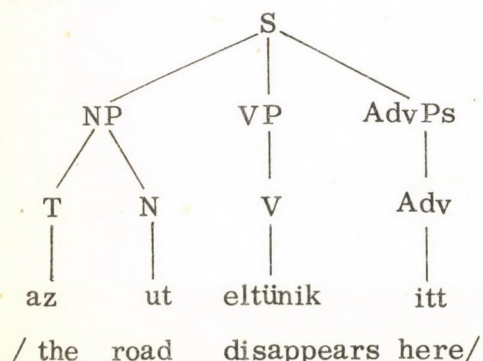
The disclosure of the deep structures may be carried out in two steps: first we establish the phrase-markers expressing the relations between the lexical units then we assign to them matrices containing their syntactic-semantic features.

### 5.1. The establishment of the phrase-markers

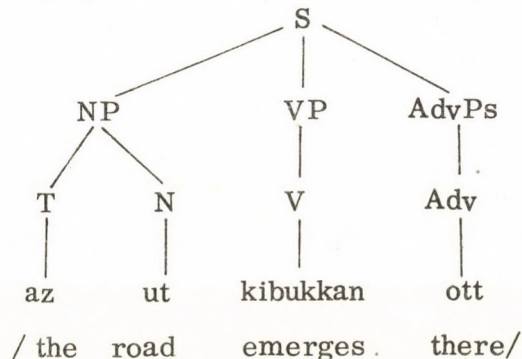
5.1.1. When establishing the phrase-marker in the case of such communication units as, for example, 1.1, 3.1, 3.6, 4.2, 4.3, or 4.12, there arises almost no difficulty.

5.1.2. Communication units containing more than one 'dominant immediate constituent' of the same character must be provided with as many phrase-markers as it is the number of these. Thus, for example, communication unit 4.4 needs four phrase-markers:

/i/

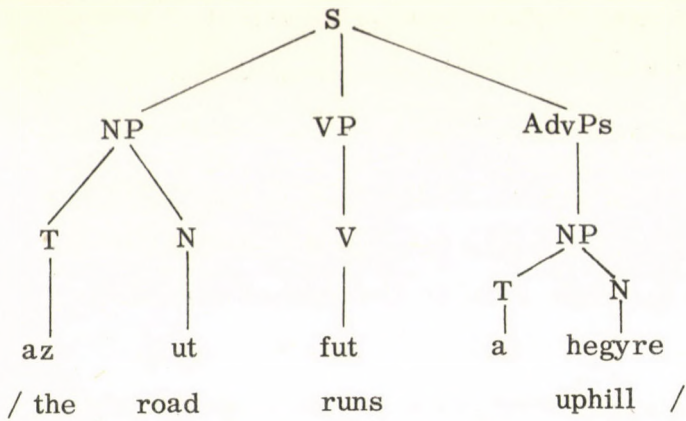


/ii/

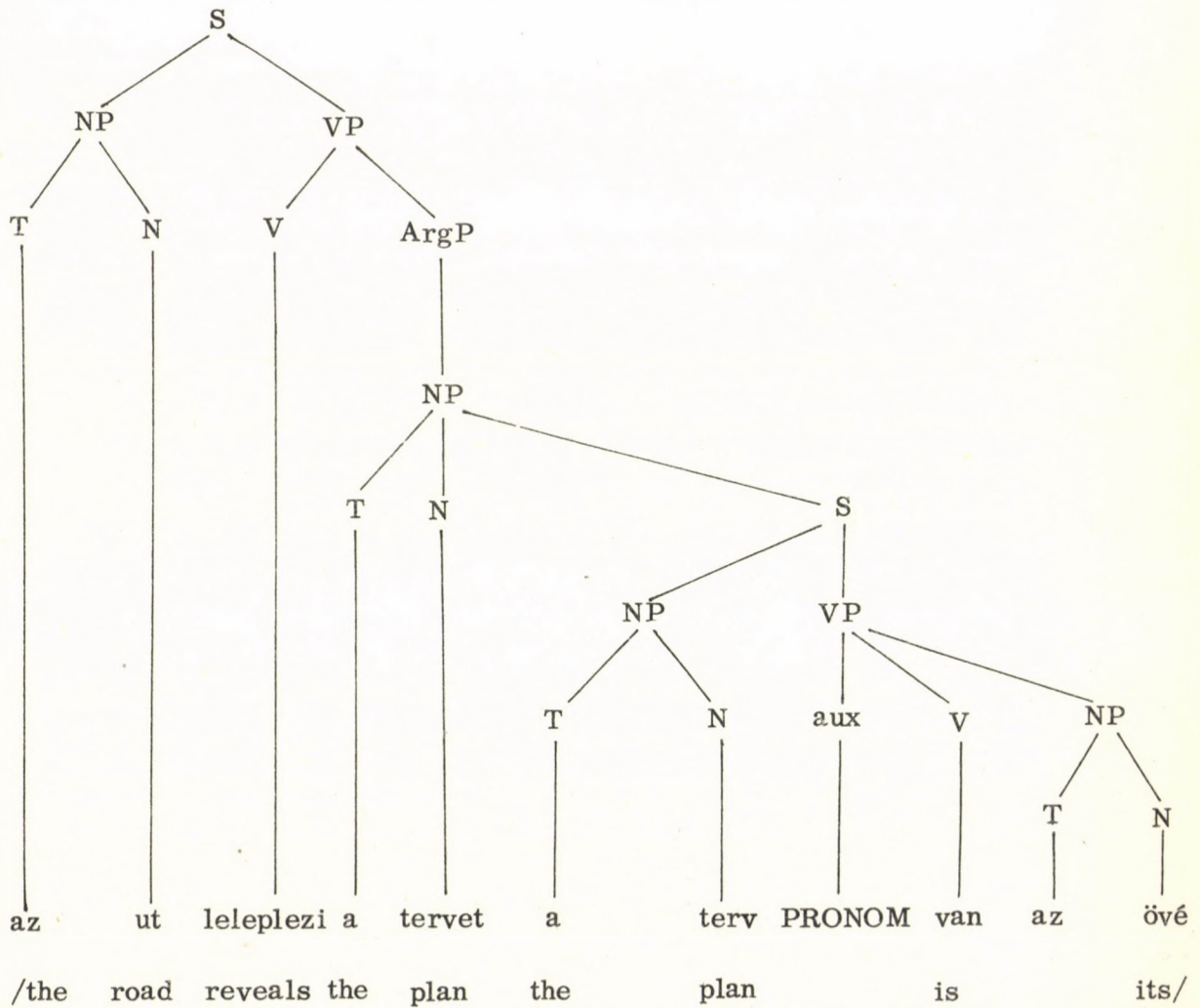




/iii/



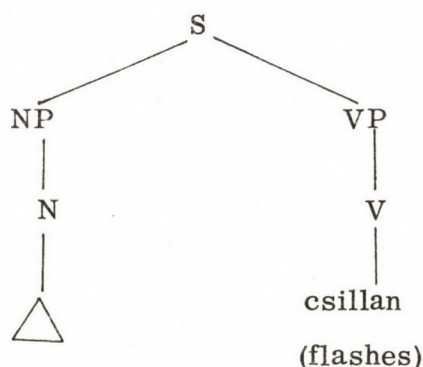
/iv/





5.1.3. Phrase-markers of communication units which are defective as to their interpretability must be completed if their context allows it.

a/ One of the well-definable classes of 'defective communication units' is made up of dominant immediate constituents separated as independent communication units. These must be 'lifted in', one by one, to the appropriate place of the phrase-marker of the basic communication unit and these now complete phrase-markers will be interpreted. This is the case, for example, with communication units 4.6-4.11. They are merely the subjects of the predicate ('flashes') of their basic communication unit, namely, of 4.5, and thus inapt for direct interpretation. In order to be able to interpret them as communications of full value we have to make them, one by one, the nominal part of the following phrase-marker:



b/ Further classes of 'defective communication units' may be constructed out of communication units.

- in which one for the 'dominant immediate constituents' is a pronoun, for example 3.7. (It should be noted here that in the case of such languages as Hungarian, in which the personal pronoun of 3rd person singular is not gender-indicating, the problem is even greater.);

- in which the 'lacking dominant immediate constituent' is not indicated by a pronoun but it occurs in the text - maybe in an unsuitable form for supplementing the communication unit in question directly (for example in a different morphological form). Such are for example

3.4: Here they met (namely the man and the girl),  
and 3.8: Oh, if only ever -



lasting happiness had taken her on its lap!

(her, namely the girl);

- which are 'absolutely lacking', that is, their supplement does not occur literally in the text but it can be rendered probable on the basis of their relation to the other communication units, for example

3.7: And she, tiny lightness?

(That is: what has become of her?);

- which are mere indications (those of a situation, a mood, etc.), as for example 4.1;

-etc.

## 5.2. About the semantic features of the lexical units

The vocabulary-component must be built up in such a way that the feature-matrices of the lexical units should contain all the features necessary for the interpretation. In our present paper we should like to confine ourselves only to the question of the semantic features and omit the syntactic ones.

5.2.1. All those universal features which are necessary to reveal a semantic deviation from the standard language<sup>3</sup> must be included in the set of the semantic features of the lexical units. (The so called 'petrified' tropes which have been taken over by the standard language must be stored, at the same time, separately in the vocabulary.)

To point out the specific character of these features let us examine, for example, the following word groups<sup>4</sup> each one containing elements pertaining to a perception-field.

1. 'sweeter than all lights' ('minden féynél édesebb'),
2. 'silent fragrant' ('néma illat'),
3. 'and is sobbing at a bruised fragrant'  
( 'és zokog egy felhorzsolt illaton' )
4. 'the silence roars into my ears' ('üvölt a csend a fülembé')
5. 'and the song of the skylarks twangs'  
( 's peng a pacsirták éneke' )



6. 'oh brooklet-voiced siskins' ('ó csermelyhangu csizek')
7. 'and the vine-shoot-fire laid on the top of the hill begins to croon secretly'  
( 's dudolni kezd titokban a hegytetőn rakott venyigetüz')
8. 'the quiet-voiced wind is sobbing' ('zokog a halkszavu szél')
9. 'the tasty wind may babble' ('az izes szél gagyoghat')
10. 'itching little winds are hissing' ('viszkető kis szelek sziszegnek')

To interpret these word groups it is necessary to know the followings about those lexical elements which pertain to a field of perception:

a/ To which field of perception they belong.

b/ Which constituent of the communication is indicated or referred to by them.

Namely, every single perception can be conceived as a taking part in a special kind of communication. We can differentiate the following constituents inside these communication processes: (See Table 1)

<u>Emitter</u>	<u>Encoding</u>	<u>The message</u>	<u>Decoding</u>	<u>Receiver</u>
		(proceeding along) the channel		
i.	emits light	the light (is spreading)	the light is seen	
ii. sy sg	emits sound	the sound (is spreading)	the sound is heard	
iii.	emits smell	the smell (is spreading)	the smell is smelled	
iv. -	-	-	the taste of sg is tasted	sy by sg
v. sg	emits heat	the heat (is spreading)	the heat is perceived	
-	-	-	the temperature, the surface, the shape, the weight of sg is perceived	

Table 1



This table also shows the different ways of perception: sight, hearing; smelling and tasting as being pure types, and skin-perception as being a mixed one.

Under the 'decoding' constituent we have listed some of the basic verbs expressing the receiver's turning toward the message. Considering it from the angle of the message, decoding can start only then, when the communication has reached the receiver. In this phase all kinds of perception can be considered as equals in so far as light, sound, smell and that something which has a certain taste, temperature, surface, shape and weight get into direct touch with the respective organ of sense. (This serves as one of the bases for changing the terms that indicate the different fields of perception.)

c/ Being constituents of given character of the communication, we have to know their place on the scales of values referring to them.

Namely, those elements which can be considered as synonyms can be arranged on different scales of values inside the primary classification shown above, for example: fragrant - smell - stink; stroke - touch - bruise; whispering - speaking - shouting. These may be conceived, approximately, as different realizations of the following scale of values:<sup>5</sup>

+ pleassant

0

- pleasant

d/ In the field of the hearing-perception, beyond what has been mentioned so far, the constituents of the communication can be put into the following classes according to the emitters (Table 2).

In the following table (Table 3) we shall characterize the oppositions contained by the 'groups of words' quoted above with the help of these main semantic characteristics (a-d). These will be referred to as 'perceptual', 'communicational', 'effect' and 'hierarchizational' characteristics.

By the above examples we only wanted to demonstrate the particular character of a small class of semantic characteristics necessary for the interpretation. (When revealing the deviations we have, of course, to take into consideration also the syntactic relations of the examined lexical units.)



<u>Inanimate:</u>		metals	twang, clang	
		water		
		(brooklet)	splashes	
		fire	crackles	
		wind	roars, whistles	
<u>Animate:</u>	plant		swishes	
			rustles	
	animal		buzzes	
		(snake)	hisses	
			twitters	
			barks	
			growls	
	human being	sings	speaks	sobs
		croons	babbles	cries
		.....	.....	.....

Table 2

5.2.2. Beyond the universal semantic features each lexical unit must be provided with references to its synonyms as well as the denominations or indices of those 'thesaurus classes', thematic groups to which the given lexical unit belongs. (It is necessary, of course, to give a clear-cut definition for synonymy and its different degrees, respectively.)

The knowledge of the synonyms is necessary for the interpretation of the communication units while that of the thesaurus classes is necessary for the establishment of those higher units which are constructed out of communication units.

### 5.3. About the semantic characteristics of the communication units

The procedure of the semantic interpretation is, in general, determined by the semantic component of the theory of language.

However, the interpretation of communication units may differ in several respects from that of single sentences.



	<u>perception</u>	<u>communication</u>	<u>effect</u>	<u>hierarchization</u>
1. sweeter than light	tasting sight			
2. silent fragrant	hearing	property of the <u>emitter</u> <u>message</u>		
3. bruised fragrant	touching smelling	property of the <u>receiver</u> <u>message</u>	-pleasant +pleasant	
4. the silence  roars		<u>message</u> or rather the lack of it action of the <u>emitter</u>	+pleasant  -pleasant	
5. the song of the skylarks twangs		<u>message</u>  action of the <u>emitter</u>		+animate (animal) -animate (metal)
6. brooklet-voiced siskins				-animate (water) +animate (animal)
7. the vine-shoot-fire croons				-animate (fire) +animate (human being)
8. the quiet-voiced wind is sobbing			+pleasant  -pleasant	-animate (wind) +animate (human being)
9. the tasty wind babbles				-animate (wind) +animate (child)
10. itching wind are hissing		property of the <u>receiver</u> having the character of a <u>message</u>		-animate (wind) +animate (animal)

Table 3

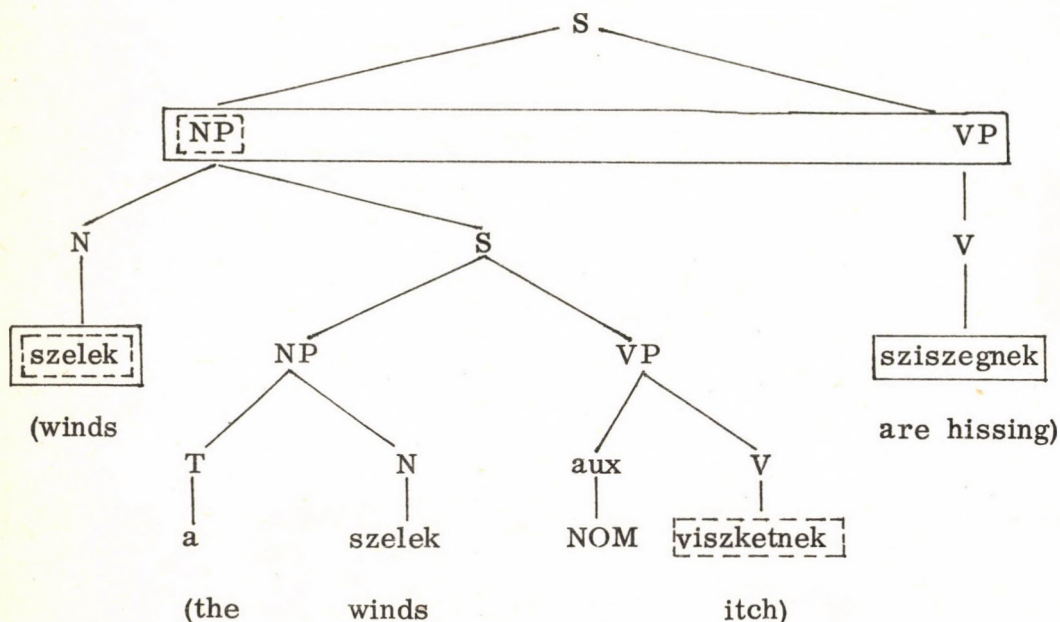


5.3.1. Such relations as, for example, the 'subject' or the 'object' of the communication unit will necessarily refer, in many cases, outside the interpreted communication unit (of. what has been said about supplementing defective phrase-markers). The 'nature of this referring-outside', as a semantic characteristic, must be assigned to the communication unit in question.

5.3.2. Those defective communication units which can be supplemented relying upon their context, must be provided with the indication of the nature of their supplementation, as a semantic characteristic. (Cf. what has been said about supplementing defective phrase-markers in 5.1.3.b.)

5.3.3. If a communication unit contains a 'semantic deviation', both the semantic and the syntactic character of this deviation must be indicated. In connection with the syntactic character we have to establish, for example, whether the deviation occurs inside a 'dominant immediate constituent' or in the relation of two (or more) dominant immediate constituents.

For example:



In the communication unit 'itching winds are hissing', applying the terms we have used in 5.2.1, the following semantic deviations can be found:

- i. a hierarchizational deviation in the relation NP      VP
- ii. a communicational deviation in the constituent      NP



5.3.4. Beyond the establishment of those relations discussed above which might be called 'syntagmatic relations', the 'paradigmatic relations' are also of great importance.

The establishment of these relations is, however, a rather intricate task. As we have seen so far, a communication unit is a specifically arranged string of lexical units being in definite syntactic relations with each other. In other words, it is the representation of a surface structure. If we interpret the different degrees of synonymy with regard to the 'meanings' of the communication units, too, on the basis of these degrees the communication units can be classed into synonym-classes. Thus the elements of these classes will be the representations of all the possible surface structures of those 'related' deep structures which are allowed by the synonym-relations.

Concerning the 'paradigmatic aspect' we have to reveal the relation of the communication units to their synonym-classes. (In this respect, to our mind, it will be necessary to rely on the generative transformational grammar as well as on the theory of semantics elaborated by Melchuk and Zholkovsky.)

## 6. About the 'composition units'

With knowledge of the syntactic-semantic structure of the linguistic communication units we have to establish the higher units of the 'work of art' in question.

6.1. In order to demonstrate the various types of 'linking together' of the communication units we shall confine ourselves to the brief analysis of the four poems presented above from this respect only. (The higher units formed out of communication units will be indicated by Roman numerals.)

### 1. Soundless

I : 1-6

The first communication unit is complete in itself while communication units 2-6 are only 'predicates' referring to the first one.



## II: 7

Communication unit 7 is connected to I (primarily to 5-6) by the adverb 'hereby' ('ezzel')

5: 'wants to speak, it seems, ...' ('szólni akar szinte...')

6: '...tumbled down	('... lebukott
Giving no word	Nem adva semmi szót
no sign'	semmi jelet')

7: 'What I know today	('Ezzel is kevesebb
is also lessened <u>hereby</u> '	amit ma tudok')

Thus, the connection is primarily of syntactic character.

2. Evening song

## I: 1-3

The possessive personal suffix of the nominal part of the 'predicate' in the second communication unit ('helyem': 'my place') refers to the first person (the speaker) indicated by a verbal suffix in the first communication unit ('hazahoztam': 'I have brought home').

This way of referring is to be found in the third communication unit ('bérem: 'my pay'), too, but here is also another connecting link, namely, the conjunction 'mégis' ('nevertheless').

## II: 4

The fourth communication unit has no formal connection with the other ones.

## III: 5

The pronominal adverbial complement of the communication unit ('velem: 'with me') indicates again the speaker while the imperative verb refers to a certain second person never mentioned before in the poem and not even revealed in this line.



Thus, the poem is completely open. The communication units - including 4 between I and III - are integrated into the poem by the monologue-like communicative situation revealing itself in the poem.

### 3. Footprints

The communication units of this poem are united into the poem by the interpenetration of the thesaurus-like semantic connections, the grammatical relations and the symmetry of the construction. (A graphic demonstration seemed to be an appropriate substitute for a detailed description. The continuous underlining indicates the thesaurus-like connections)

- I:1 clearly can be seen / the footprints
- II:2 Here came the man ... heavy / steps
- 3 And towards him - lightly, running... - the girl
- 4 Here they met (namely the man and the girl)
- III:5/1' / Then - where is the print?
- IV:6/2' / only that weighty man's-foot / goes ... onwards
- 7/3' / And she ... lightness?
- (that is: what has become of her,  
namely, of the girl?)
- 8/4' / On, if only ever-/lasting happiness had taken her on its lap!

The figurative level of communication unit 8 becomes of full content only in this full connection-network of the poem, this 'content' having been already prepared - on a material level - by the adverbial complement (of mood of communication unit 6 ('crushing the deep snow even deeper')).



#### 4. Dawn

In this poem there is only one higher unit having between its communication units an explicit connection. Communication units 6-11 are namely the subjects of the predicate of communication unit 5' ('csillan': 'flash'), being independent communication units in character.

The relatedness of the communication units is conveyed by their references: 'Morning...'; 'The stars drop away'; '... flash'; 'The cog-wheel of the Sun begins to work'.

We should like to underline, that our remarks were intended to serve only as a contribution to the enlightenment of a viewpoint, and by no means can they be considered as full analyses of these poems.

6.2. In order to disclose the composition units the following tasks are to be fulfilled:

To determine the possible syntactic and/or semantic connection-types of the communication units and the arrangement of these connection-types according to their 'strength-degree'. Synonymous and thesaurus-like connections will also be considered as semantic ones;

To characterize the ways in which the composition units constructed directly out of communication units are organized into higher units and these again into higher ones.

6.3. It will be necessary for the syntactic-semantic characterization of composition units to give as characteristics, the 'type of connection' by way of which the smaller units (let us label them 'compositional immediate constituents') are organized into the composition unit in question, as well as the semantic deviations to be found in it. These semantic deviations, again, may be revealed either inside a 'compositional immediate constituent' or in a specific relation of them.

(Thus it will also be possible to interpret all the manifestations of any degree of complexity of the so called 'figaruitive speech' in a homogeneous way.)



## 7. Concluding remarks

In our paper we have discussed some questions of the linguistic semantic analysis of 'verbal works of art'. Our intention has been to point out how the methods and results of the most recent syntactic-semantic investigations can contribute to building up a model that is aiming at an all-embracing structural linguistic analysis of verbal works of art.

The basic conception of this model has been that a full structural linguistic description is only possible by way of characterising intensionally the elements and units of both the linguistic and the musical sign-component.

As for our present paper, we have dealt exclusively with the 'linguistic sign-component' of the model confining ourselves primarily to the 'layer of communication units'. Concerning the 'layer of composition units' we only wanted to indicate that the linguistic analysis of composition units of different degrees of complexity can be carried out - if it is possible at all - in an analogous way with that of the analysis of communication units. Beyond the intensional characterization of the different units the following analogies underlies underlie this identical way of analysis:

	layer of <u>communication</u> units		layer of <u>composition</u> units
units	communication units	composition units constructed directly out of communication units	composition units constructed out of elements of different complexity (here belongs the whole of the work of art, too)
constituents	communication unit parts (dominant immediate const.)		composition unit parts (composition and/or communication units)
elements	words	communication units	composition and/or communication units)



These analogies contain essential differences as well.

Linguistic methods will always enable us to explore and describe the syntactic-semantic structure of communication units while concerning the analysis of composition units there may arise difficulties. It may happen that starting out from communication units we may arrive at groups of communication units which will be considered by our 'intuition' as composition units despite the lack of any kind of syntactic-semantic connection between them. This type of connection may occur again and again on different layers until we reach the work of art itself.

If not even the joint analysis of the linguistic and musical sign-component can explain the nature of these connections we can say, that the dominant structure-organizing factor is not of linguistic character. It may be the 'represented reality' or one of the 'secondary structures'.

Thus we use the term 'full linguistic structural analysis' in the sense of disclosing all those structural connections inside a work of art which can be explored by linguistic means.

To reveal these connections - and at the same time to state the limits of the structural linguistic analysis - the most important task to be fulfilled appears to be the semantic analysis of the composition units.

### Notes

- + The present paper was discussed by the Symposium on Semiotics in Warsaw, August 1968.
- 1. Here we have summed up briefly our model concerning the structural linguistic analysis of 'poetic works of art' which has been exposed in its outlines at the 10th International Congress of Linguists (Bucarest. 1967). A detailed version of it appeared in Computational Linguistics VI (Budapest). Here you will find further bibliographical references.
- 2. Poems by the Hungarian poet GYULA ILLYÉS. (Poharaim, Összegyűjtött versek, Budapest, 1967.)



3. The term 'standard language' is used in the sense of that 'form of expression' which may be produced by the generative model of the given language or is contained in a supplementary list to the model.
4. Lines of verses taken from poems by MIKLOS RADNOTI and ÁRPÁD TOTH.
5. Scales of values are of wide-spread use in papers on semantics and lexicology. For example, of. J. LEVY. The Meanings of Form and the Forms of Meaning, in: Poetics Poetyka Poetika II. Warszawa, 1966 pp. 45-61.







## SYNTACTIC ANALYSIS ON THE BASIS OF CONTINUOUS JUNCTION OF SYNTACTIC UNITS<sup>+</sup>

D. Varga

In the last number of Computational Linguistics I gave the outlines of the automatic system of analysis that has been under development at the Computational Centre of the Hungarian Academy of Sciences since 1966 [1]. The present article gives a short description of the principles of the algorithm for carrying out syntactical analyses.

The algorithm produces all those structures of the analyzed symbol string that meet the requirements of the grammar applied. The grammar can be changed or modified; the analyzing algorithm itself processes matrices, a special input program has been elaborated to build up these matrices from the defined grammar.

Our starting point was B. Dömölki's algorithm for the syntactic analysis of formal languages [2]. The discovery of the syntactic structures of natural languages did demand, however, the elaboration of a more sophisticated algorithm. The input program is the work of É.B. Szöllősy and Zs. Varga.

The basic idea of the algorithm is somewhat similar to the idea of dynamic programming [3]. Dynamic programming aims at the step by step, gradually extended optimisation of the part processes instead of attempting to optimize the whole of the process in question. In our case the basic requirement is the formation of a single continuously organized structure out of all the syntactic units of the whole sentence. The algorithm works by constantly

---

<sup>+</sup>The present paper is part of the lecture delivered by the author in September 1968 at the Balatonszabadi Conference on Mathematical Linguistics.



checking whether the condition of continuity is fulfilled on the step by step extended string of syntactic units as related to the whole structure uncovered so far.

Continuity may be realized by the simultaneous application of a series of different rules.

The algorithm checks the fulfillment of the rules simultaneously, always taking into consideration those possibilities that have already fulfilled the previous checks for continuity.

The linguistically most important part of the algorithm is a matrix of hierarchization. The vectors that make up the rows of the matrix correspond to the breaking up of syntactic units on the basis of their syntactic behaviour. This "syntactic spectrum" indicates which position of which syntactic rule may be filled by the unit in question. (The algorithm employs a more condensed form in the storage and in the processing but for the sake of clarity I would prefer not to discuss it in detail.)

Let us examine the following grammar as an example:

$$(i) \quad c + d \Rightarrow G$$

$$(ii) \quad \begin{Bmatrix} a \\ c \end{Bmatrix} + \begin{Bmatrix} b \\ d \end{Bmatrix} \Rightarrow F$$

$$(iii) \quad a + \begin{Bmatrix} a \\ b \\ c \end{Bmatrix} + c \Rightarrow F$$

$$(iv) \quad F + G + e \Rightarrow K$$

$$(v) \quad \# + K + \Delta \Rightarrow \Sigma$$

In the matrix of hierarchization the following rows correspond to the syntactic units:



$$\begin{array}{lcl}
 a & \longrightarrow & (0, 10, 110, 0, 0) \\
 b & \longrightarrow & (0, 01, 010, 0, 0) \\
 c & \longrightarrow & (10, 10, 011, 0, 0) \\
 d & \longrightarrow & (01, 01, 0, 0, 0) \\
 e & \longrightarrow & (0, 0, 0, 001, 0) \\
 F & \longrightarrow & (0, 0, 0, 100, 0) \\
 G & \longrightarrow & (0, 0, 0, 010, 0) \\
 K & \longrightarrow & (0, 0, 0, 0, 010) \\
 \# & \longrightarrow & (0, 0, 0, 0, 100) \\
 \triangle & \longrightarrow & (0, 0, 0, 0, 001)
 \end{array}$$

This means that the syntactic unit a does not occur in the first rule, it can be first element in the second syntactic rule, first and second element in the third rule, etc. In the course of the analysis the possibilities of connecting symbols of the string under analysis are examined with the help of vectors that are assigned to the symbols.

Let us enlarge the symbol string in question during the analysis by adding such a symbol of complex syntactic unit to it which - according to the inverted Polish designation without brackets - expresses the syntactic structure accepted hypothetically. If for example the symbol string  $\# a b c d e \triangle$  is analyzed and during the analysis of elements  $a$  and  $b$  it comes to light that the elements  $a$  and  $b$  may be contracted according to a rule  $a + b \Rightarrow F$ , then a new symbol string  $\# a b F \dots$  corresponds to the analysis so far. (As we apply also rules other than binary ones an index "2" may be attached to  $F$  signifying the number of elements on the left hand side of the rule.)

When applying the rules we adhere to the principle of precedence i.e. as soon as we are convinced of the total applicability of a rule, we accept its lawful employment as long as the hypothesis shows no contradiction with the principle of the continuity of the structure. Thus by employing the principle of precedence the last hypothesis that proved incorrect may be corrected every time discarding the results obtained previously. Therefore the analysis may be continued in every case deleting a certain number of the hypotheses as counted from the end if there is a possibility to "join together" a continuous structure



at all. Going further, all possible structures may be disclosed by this very same principle.

Let us designate the  $t^{\text{th}}$  element of the enlarged string with  $x_t$ , and let  $H[x_t]$  be the vector string assigned to the syntactic unit from the matrix of hierarchization designated by the index  $x_t$ .

Dömölki gives the following definition of the vector function series expressing momentary states of the analysis:

$$Q_0 = O$$

$$Q_{t+1} = \left( \frac{1}{2} Q_t \vee B \right) \wedge H[x_{t+1}] ,$$

where  $B$  indicates the Boolean vector representing the initial position of the syntactic rules.

The meaning of this expression is as follows. The next in turn element of the symbol string under analysis may be attached to those positions of the already examined  $x_1 \dots x_t$  structured symbol string which positions have ties through the right hand side of either of the "live" positions of the  $Q$  vector of state, that is

$$\frac{1}{2} Q_t \wedge H[x_{t+1}] \neq O ,$$

or the symbol  $x_{t+1}$  will start a new syntactic rule, consequently

$$B \wedge H[x_{t+1}] \neq O .$$

Accordingly,  $Q_{t+1}$  may only be the continuation of the structure so far if

$$Q_{t+1} \neq O \text{ where } Q_{t+1} = \frac{1}{2} Q_t \wedge H[x_{t+1}] \vee B \wedge H[x_{t+1}]$$

The opposite of this statement, however, is not true, therefore the condition is not necessary only sufficient.



Dömölki's result may be made more potent if we employ instead of  $B$  such a  $B^+$  vector that designated the initial position of only those syntactic rules that may make up directly or indirectly such syntactic units that may be the continuation of the structure completed so far.

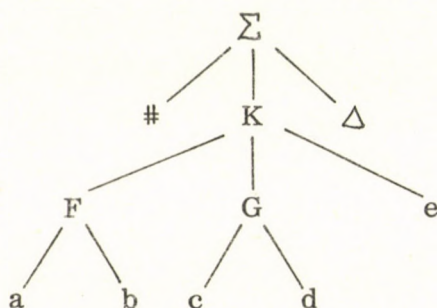
Let  $N$  be the vector to indicate the positions of the complex syntactic units where in the coordinates of  $N^i$  that are not zeros the complex syntactic units of  $S_{ij}$  may stand. The  $S_{ij}$  units define the individual subsets  $M_{ij}$  of the  $M$  syntactic system of rules by the following recursive definition:

A rule is an element of  $M_{ij}^+$  if and only if the first element of the right side is either  $S_{ij}$  or such an element that may be the first element of the left hand side of the rule that belongs to  $M_{ij}^+$ .

Let  $B_{ij}^+$  be that vector of position characteristic of the  $M_{ij}^+$  set, which assigned a 1 bit to the initial positions of the rules belonging to  $M_{ij}^+$ . With the help of these vectors  $B^+$  is easily obtained:  $B^+$  is the disjunction of the vectors  $B_{ij}^+$  assigned to the  $N^{i_k} \neq 0$  coordinates of the vector  $\frac{1}{2} Q_t \wedge N$ .

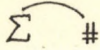
Applying the above apparatus the determination of the structure from the string  $\# a b c d e \triangle$  may be carried out as follows, applying the above grammar:

Let the structure under examination correspond to the following tree diagram:



First of all the possibility of continuous junction between the starting element of the string  $\#$  and the element  $\Sigma$  regarding as the starting point of all structure is examined by the vector  $B^+$ . As there is a rule that starts with  $\#$  and results in  $\Sigma$ , the condition of continuity is fulfilled, i.e.





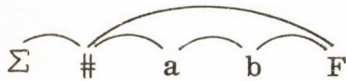
It is likewise fulfilled for the next element a, as for, continuing the rule  $\# + K + \Delta \Rightarrow \Sigma$  the necessary K element may be produced by a rule that begins with a symbol a. Formally this may be deduced from the fact that the rule  $a + b \Rightarrow F$  is an element of the  $M^+$  set of rules represented by  $B^+$  vector. Therefore the symbol string may be enlarged:



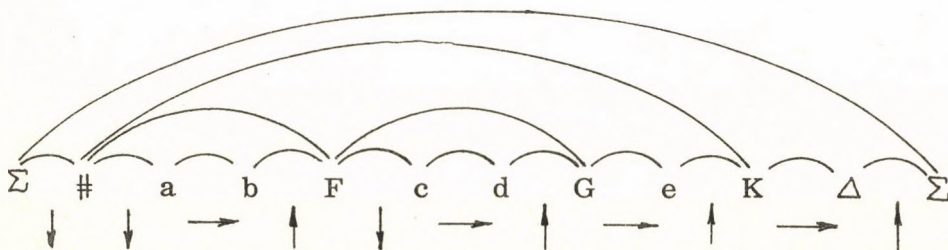
The condition of continuity may be fulfilled when the next element b is attached as b may continue the rule  $a + b \Rightarrow F$

At the same time one of the conditions for applying the rule is fulfilled: we controlled the whole left hand side of the rule.

At the stage we add the compound syntactic unit F to the symbol string under examination according to the principle of precedence defined for the analysis of structures if the second condition of continuity is also met. According to this condition the element in question should also be continuously joined to the syntactic element preceding the substituted element. In other words the new element should also fit into the structure "on the higher levels". This condition is also met, because the rule beginning with F is also part of the  $M^+$  set related to K. Therefore according to our designation



The analysis progressed in this manner on the basis of further continuity checks until the complete structure is formed:





Essentially by this method we "walk around" the structure to be analyzed. Our method unites in itself the advantages of both the analysis that progressed from top to bottom and the analysis in the reverse direction, from the bottom to top [4]. While in Dömölki's method the analysis progressed purely from bottom to top, thanks to the introduction of the  $B^+$  vector progressing from top to bottom has also become a controlled process. This saves us from getting into many unnecessary blind alleys.

### References

- [1] VARGA, D. Towards a new system of automatic analysis, Computational Linguistics VI, Budapest, 1967, 123-136.
- [2] DÖMÖLKI, B. An algorithm for syntactic analysis, Computational Linguistics III, Budapest, 1964, 29-46.
- [3] BELLMAN, R. Dynamic programming, Princeton University Press, Princeton, New Jersey, 1957.
- [4] Cf. BORSCHEV, V. B., JEFIMOVA, J. N. O sokraschenii perebora pri sintaksicheskom analize, Nauchno-texnicheskaja informacija, Ser. 2. 1967, 10. 27-33.







REVIEW







Ю. Д. Апресян: Идеи и методы современной структурной лингвистики / Краткий очерк /. Издательство "Просвещение". Москва, 1966. 300 стр.

Когда халифа Омара спросили, следует ли разделить книги, найденные в покорённом городе, среди правоверных вместе с другой добычей, он так ответил: "Если в этих книгах говорится то, что есть в коране, они бесполезны. Если же в них говорится что-нибудь другое, они вредны. Поэтому и в том и в другом случае их надо сжечь." Долгое время — пишет автор в предисловии — приблизительно так же относились и к структурной лингвистике. Её новые идеи считали вредными, а все остальные положения — давно известными истинами. В ложности этого "омарского взгляда" читатель книги Ю.Д. Апресяна скоро убеждается, так как рассказывается ему о закономерностях возникновения структурной лингвистики и о её интересных, новых результатах.

В СССР издавали уже много книг, занимающихся структурной лингвистикой, но ни одна из них не даёт такого полного представления о ней и в такой общепонятной форме, как книга Ю.Д. Апресяна / т.к. по тематике они или более широкие <sup>1</sup>, или более узкие <sup>2</sup> и рассчитаны скорее на читателей-специалистов по лингвистике /.

---

<sup>1</sup> См., напр. : В.А. Звегинцев, История языкознания XIX и XX веков в очерках и извлечениях. I — II. М. 1960.

<sup>2</sup> Напр. : И.И. Ревзин, Модели языка М. 1962 ; Метод моделирования и типология славянских языков М. 1967.



Автор данной книги не ставил себе целью систематического изложения структурной лингвистики, а лишь в форме очерка ввести читателя в её проблемы и подготовить к чтению специальной литературы. Эту задачу он прекрасно выполнил и снова показал свою способность писать работу обзорного характера <sup>3</sup>.

Первая часть книги посвящается истории структурной лингвистики. Среди причин возникновения структурной лингвистики различаются внешние и внутренние стимулы. Благодаря созданию электронных вычислительных машин появились с одной стороны т.н. информационное дело / МП, поиск информации и т.д. /, с другой стороны возможность механизации трудоёмких лингвистических работ, одинаково поставившие перед лингвистикой требование более точного описания языка. Более важную роль сыграли внутренние стимулы. С конца XIX века традиционная грамматика подверглась критическому и конструктивному пересмотру. Направили острiot критики в первую очередь против неточности понятий, определённых на основе интуитивных семантических критериев, против преимущества точки зрения слушающего, против эмпиризма. В ходе этого пересмотра вырабатывался структурный подход к языку.

В дальнейшем суммируются основные идеи непосредственных предшественников структурной лингвистики, И. А. Бодуэна де Куртенэ и Ф. де Соссюра и кратко / на 40

---

<sup>3</sup> См. Ю.Д. Апресян, Современные методы изучения значений и некоторые проблемы структурной лингвистики: Проблемы структурной лингвистики М. 1963. 102-150.



страницах / излагаются три классические школы структурной лингвистики: пражская / функциональная лингвистика /, копенгагенская / глоссематика / и американская / дескриптивная лингвистика /. Несмотря на различия между ними, они в той или иной мере усвоили учение Ф. де Соссюра, одинаково понимают язык, как объект лингвистики и характеризуются такими общими методологическими требованиями, как простота, полнота, последовательность, объективность, формальность. Сознательно широкое понимание структурной лингвистики позволяет автору с одной стороны считать вышеупомянутые школы разновидностями структурализма / но уже не причислять к ним Московскую, Лондонскую, Женевскую школы / и с другой стороны дать в дальнейшем широкий обзор новых учений, развивающих идеи названных классических школ.

Современную структурную лингвистику можно определить как науку о моделях языка. Вторая часть книги занимается лингвистическим моделированием<sup>4</sup>. Выводятся и анализируются следующие основные свойства моделей: / 1 / они имитируют функцию объекта, и отвлекаются от его физической природы; / 2 / являются некоторой идеализацией объекта; / 3 / обычно оперируют не понятиями о реальных объектах, а конструктами; / 4 / должны быть формальными; и / 5 / обладать свойством объяснительной силы.

---

4

За исключением одной главы, в которой объясняются те элементарные математические понятия из области теорий множеств, теории графов и теории вероятности, которые понадобятся читателю при чтении книги.



Построение модели начинается с фиксирования фактов, требующих объяснения, потом выдвигаются гипотезы, которые реализуются в виде моделей, не только объясняющих исходные факты, но и предсказывающих новые, ещё не наблюдавшиеся факты, а наконец следует экспериментальная проверка моделей.

Автор знакомит читателя с одним возможным способом классификации моделей, в которой модели отличаются друг от друга по характеру рассматриваемого в них объекта. Модели первого типа используют в качестве объекта конкретные языковые явления. Эти модели имитируют речевую деятельность человека / Пражская школа сделала первый серьёзный шаг в их разработке /. Они подразделяются на несемантические /имитирующие способность человека понимать и строить грамматически правильные фразы / и на семантические / имитирующие способность человека понимать и строить осмысленные предложения /. Среди них различаем модели анализа и модели синтеза; порождающие модели занимают между ними промежуточное место. В моделях второго типа в качестве объекта рассматриваются процедуры, ведущие лингвиста к обнаружению языковых явлений. Эти модели имитируют исследовательскую деятельность лингвиста. / Американские дескриптивисты сделали первый серьёзный шаг в их разработке. / Их назначение — объективно обосновать выбор понятий, используемых при изложении моделей первого типа. Исследовательские модели делятся в зависимости от того, какая информация рассматривается в них в качестве исходной. Исходная информация может быть или текст, или кроме текста и множество правильных фраз, или кроме них ещё и множество семантических инвариантов. В моделях третьего типа рассматриваются в качестве объекта уже готовые лингвистические описания.



Эти модели являются метатеориями. / Глоссематики сделали первый серьёзный шаг в их разработке. / Их назначение — выработать систему оценок, с помощью которой становится возможным сравнение различных моделей и выбор наилучшей из них. Модели построены в форме исчисления или алгоритма.

В последних трёх частях книги подробно излагаются названные типы моделей и иллюстрируются удачно выбранными конкретными исследованиями. Приведённые примеры являются по возможности простыми, часто взятыми из самых последних исследований / таким образом некоторые из них в известной мере могут служить вместо рецензии /. Наряду с исследованиями признанных авторитетов, получили здесь место и значительные результаты молодых учёных.

В связи с исследовательскими моделями отдельно рассматриваются модели дешифровки и экспериментальные модели, сходные по постановке задачи / перевод от текста к "системе" /, но отличающиеся друг от друга в методе её получения и в характере исходной информации. В качестве примеров для модели дешифровки представлены алгоритмы Б.В. Сухотина и З. Харриса и здесь излагается коротко модель Ю.Д. Апресяна, в которой описывается семантика русского глагола по его синтаксическим свойствам<sup>5</sup>. Экспериментальные модели используют разные примеры: добавление, опущение, перестановку элементов, субституцию, трансформацию, перевод. Модели этого типа иллюстрируются работами З. Харриса, А.А. Зализняка / морфология /, З.М. Волоцкой, П.А. Соболевой / словообразование /,

---

<sup>5</sup> Недавно вышла в свет новая книга Ю.Д. Апресяна, в которой он подробно излагает свою теорию: Экспериментальное исследование семантики русского глагола М. 1967.



в связи с синтаксисом излагаются метод НС и ТМ.

Круг моделей речевой деятельности настолько широк, что автор вынужден был ограничиваться некоторыми синтаксическими и семантическими моделями. Рассматриваются / 1 / порождающие модели / модели порождения по НС в тесной связи с гипотезой Ингве; трансформационная модель <sup>6</sup> ; аппликативная порождающая модель С.К. Шаумяна <sup>7</sup> ; / 2 / синтаксические модели анализа / последовательный анализ, примером служит модель И.А. Мельчука ; предсказуемый анализ; поиск опорных точек; метод фильтров, иллюстрируемый моделью И. Лесерфа / ; / 3 / семантическая модель анализа, которая представлена работой ЛМП МГПИИЯ / А.К. Жолковского, Н.Н. Леонтьевой, Ю.С. Мартемьянова, В.Ю. Розенцвейга, Ю.К. Щеглова и других исследователей / 8.

последняя часть книги посвящена метатеории. Показывается модель И.А. Мельчука, формализующая критерии оценки /полноту, адекватность и т.д. / на базе теории множеств, при помощи которой экспериментально можно сравнивать модели и иллюстрируется возможность теоретического сравнения моделей примером, взятым у Хомского.

---

<sup>6</sup> С тех пор теория порождающих моделей во многом развивалась дальше.

<sup>7</sup> Новый вариант этой модели: С.К. Шаумян, Структурная лингвистика М. 1965.

<sup>8</sup> С тех пор появились: А.К. Жолковский – И.К. Мельчук, О возможном методе и инструментах семантического синтеза: НТИ 1965. № 6; 23-28; О системе семантического синтеза I. НТИ 1965. № II, 48-55 и 2. НТИ 1967. № 2, 17-27.



Книга Ю.Д. Апресяна является образцом научно-популярной работы / тираж: 35 тыс. экземпляров ! /. Популярность и научность оказываются вполне совместимыми / что бывает довольно редко /, одна не вредит другой. Автор умеет чрезвычайно ясно и одновременно сжато писать, проникая в суть проблем. Термины постепенно вводятся и их число небольшое. В случае наличия более принятого термина, менее употребляемый термин заменён более известным / напр. : ассоциативное отношение Ф. де Соссюра заменяется термином парадигматического отношения, ссылаясь на замену /. Указывается на совпадение объектов, по разному наименованных отдельными лингвистами и на случаи, когда одни и те же термины по-другому использованы ими. Каждое утверждение сразу же подтверждается наглядным примером, богатые библиографические указания помогают читателю в дальнейшей ориентации.

Заметно, что автор относится к структурной лингвистике с большой любовью, из разных исследований старается выделять самые ценные идеи и результаты, из которых к концу "романа структурной лингвистики" складывается привлекательная картина о ней.

Рецензент надеется на переводные издания книги, в их числе и на её венгерское издание, тем более, что знает о готовящемся венгерском переводе / по нашим сведениям готовится и немецкое издание /.

П. Санто











