American Journal of Computational Linguistics *Microfiche 58 : 78*

MATHEMATICAL MODELS OF LANGUAGE

SOVIET PAPERS IN FORMAL LINGUISTICS

EDITED BY FERENC KIEFER

Sprakförlaget Skriptor AB
Box
S 104 65 Stockholm
1973

284 pages SKr 60

REVIEWED BY DAVID B. BENSON

Computer Science
Washington State University
Pullmar 99163
(Currently visiting at the University of Colorado)

From the Preface:

"The volume is more or less a random sample of the great number of works done in the field of mathematical linguistics by Soviet scholars. The random character of the selection is due to the difficulties which an editor inevitably encounters in compiling an anthology like the present volume. If I were to start working on this volume now I would certainly choose more recent papers, perhaps ones in one or another aspect more representative than those included in this volume. Nonetheless, these articles are at least in one respect representative. They clearly testify to the breadth of interest and variety of approaches in Soviet mathematical linguistics. This anthology is intended to convince the reader who has not mastered Russian and is perhaps not familiar with works by Soviet "mathematical linguists" that they deserve much more attention than they have received up to now."

The papers in this volume are indeed beginning to show their age. From internal evidence, primarily the bibliography or notes at the end of each paper, these papers were written in 1967-1972. As a "random sample", the only way to review these papers is to take each in turn.

1. M. V. Arápov - E. N. Efimova
   On the Complexity of Government Trees.
   pages 3-36.

A government tree is a derivation tree deprived of its labels. Thus the complexity relates solely to the structure of the tree without regard to phrase-names (nonterminals), lexical considerations and the like.

> "On the one hand, the government tree contains information about the structure of the text which must be taken account of in any model. On the other hand, it is a comparatively simple object for which it is easier to develop a suitable mathematical apparatus."

The complexity depends on the internal arrangement of vertices, thus for a sentence of length n, there are government trees with n leaves which have minimal complexity. "Here we shall proceed from the assumption that those structures which have minimal or close to minimal complexity are realized in natural language." Very reasonable.

The complexity measure used is developed as follows: For each vertex let k_i be the out-degree of i, that is, the number of descendents of i, and let i* be the father of i. Let the root of the tree be node 0. The complexity of each vertex is defined as:

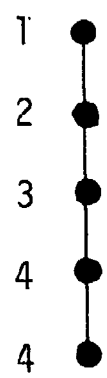
$$F(0) = k_0$$

$$F(i) = k_i + F(i^*) \text{ for } i>0.$$

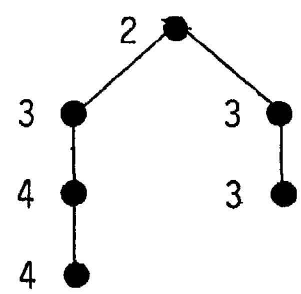
For tree Δ_N with N vertices, the complexity is:

$$F(\Delta_N) = \sum_{i=0}^{N-1} F(i)$$

For example:



$F(\Delta_5) = 14$



$F(\Delta_6) = 19$

Consider the set of all trees with N vertices, M_N and define $\Delta \varepsilon M_N$ to be the minimal if and only if

$$F(\Delta) \leq F(\Delta')$$

for all $\Delta' \varepsilon M_N$. Let $M_N \subset M_N$ be the set of minimal trees with N vertices and let $M = \bigcup_N M_N$ be the set of all minimal trees. Then for each minimal tree $\Delta_N \varepsilon M$ $\phi(N) = (\Delta_N)$ ϕ is the complexity measure studied. This does not directly find the complexity of the minimum tree with n leaves, which is a more interesting question given the paper's stated orientation toward linguistics. Nonetheless, the authors find several suggestive results about the structure of minimal trees.

Theorem I. If $\Delta\varepsilon M$, then $k_0 \leq 3$

Assuming that this notion of minimality is indeed a principle of economy, then no sentence has more than three main constituents.

Theorem II. If $\Delta\varepsilon M$, for each vertex $i>0$, $k_i^* \geq k_i$.

The deeper one goes in a minimal tree, the lower the out-degree.

"..the monotonous decrease in the number of arrows issuing from a vertex proportionate to its... distance from the root of the tree essentially agrees with the empirical facts. In fact, the number of completed valences for the verb-predicate (which are usually placed in the root of the tree) is on the average larger than for a noun which is subordinate to it, larger for the noun than for an adjective subordinate to the noun, and this number is more often than not equal to zero for an adverb governed by such an adjective. Of course, such a monotony is in reality only approximate."

Theorem III. For any $\Delta_N \varepsilon M$ with $N>81$, $k_0 = 3$.

Theorem IV. $\phi(N)$ is of the order $N \ell n N$.

The authors point out that a detailed comparison of minimal government trees with 'concrete' syntactic structures is without much meaning. Nonetheless, this is the first paper that I know of which broaches the notion of an economy of syntactic effort. Whether the theorems are indeed suggestive of linguistic reality is a matter for future research.

2. V. B. Borscev - M. V. Xomjakov

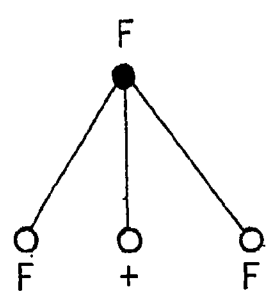
Axiomatic approach to a Description of Formalized Languages and Translation Neighborhood Languages.

pages 37-114.

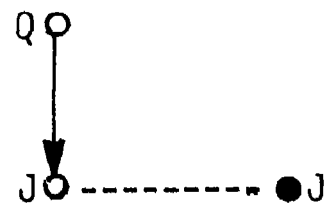
This lengthy contribution consists of four chapters of detailed development. The basic plan is an interpretation of P. M. Cohn's Universal Algebra (Harper and Row, New York, 1965) as relational systems to treat "texts" and grammars. While I enjoyed reading Cohn's excellent treatment of universal algebra, I did not enjoy this paper. It tends to wander, whereas I prefer papers which build to a definite climax. Further, most of the authors' ideas have been presented in the Western literature, so I found at most two new nuggets of wisdom. Nonetheless, here is the substance of the paper.

Chapters I and II build a notion of "text" and grammar via systems of relations. One has relations of "to the left of" and "below" in trees as well as other relations, such as "isomorphic subtree". Even the notions of terminal and non-terminal alphabets are treated as relations. This uniformity might offer some advantages for the abstract development about classes of sign systems, texts and grammars, but makes the concrete cases and examples hard to follow. In fact there are other uniform treatments, mentioned below, which are undoubtedly better for the particular cases in question.

The authors treat neighborhood grammars in these two chapters. A neighborhood of a vertex in a tree consists of some of the connecting arcs and nearby nodes. For example, a neighborhood of F is (page 53):



where the distinguished node whose neighborhood is in question is marked by Given a collection of neighborhoods, a tree is in the neighborhood language if all the neighborhood constraints specified in formula which constitute the grammar are satisfied all nodes of the tree. The major virtue of this approach is in enabling one to specify other connections between the nodes of trees other than the usual descendant relation. Thus

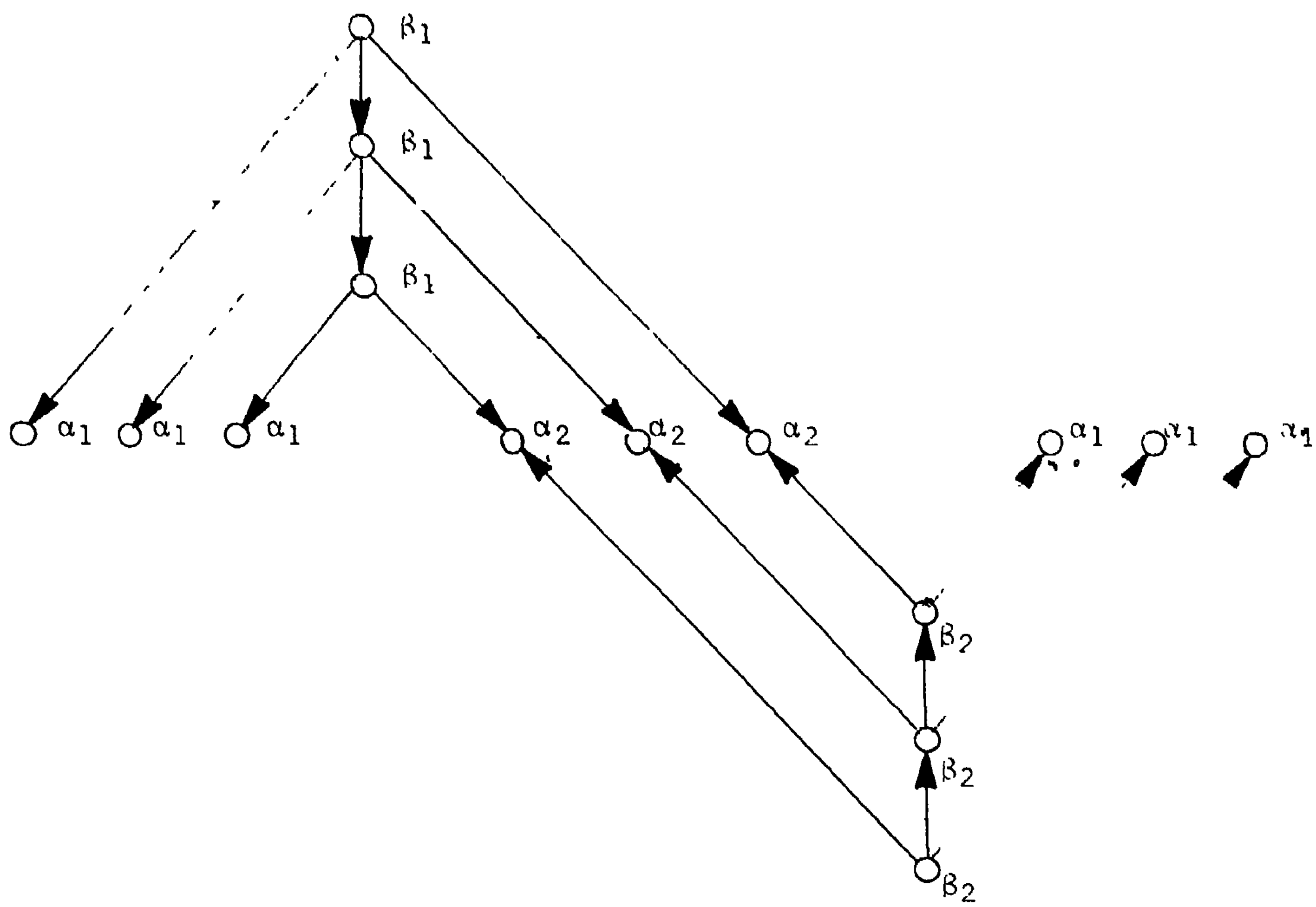


is a neighborhood specifying -- via other information too complex to describe here -- that both copies of J dominate isomorphic subtrees. This enables one to specify "syntactically" that every variable in a programming language must be declared before it is used.

However, there are better methods to handle these non-tree restrictions. For example, property grammars (Aho and Ullman, The Theory of Parsing, Translation, and Compiling: Vol. 2, Compiling, Prentice-Hall, Englewood Cliffs, N. J., 1973), macro-grammars (M. J. Fischer, Grammars with macro-like productions. Ph.D. Thesis, Harvard University, 1968), and mathematical semantics (R. D. Tennent, The Denotational Semantics of Programming Languages, Comm. ACM 19:8 (Aug.1976), 437-453.)

Chapter III, "concrete sign systems" develops phrase structure grammars and nominal neighborhood grammars. The type 0 phrase structure grammars produce "phrase structures", as generalizations of trees. These phrase structures have appeared in the Western literature in at least the following papers: J. Loeckx, The Parsing for General Phrase-Structure Grammars, Inform. & Control 16:5 (Jul 1970), 443-464, H. W. Buttelmann, On the Syntactic-Structures of Unrestricted Grammars, I. Generative Grammars and Phrase Structure Grammars, Inform. & Control 29:1 (Sept 1975), 29-80, D. B. Benson, Syntax and Semantics: A categorical view, Inform. & Control 17 (1970), 145-160, all three of which were originally written in 1969-1970. This development was clearly ripe at that time in Russia, Europe and the U. S.

The nominal neighborhood grammars are an extension of neighborhood grammars which allow fairly complex structures. For example, the following is taken from page 90:



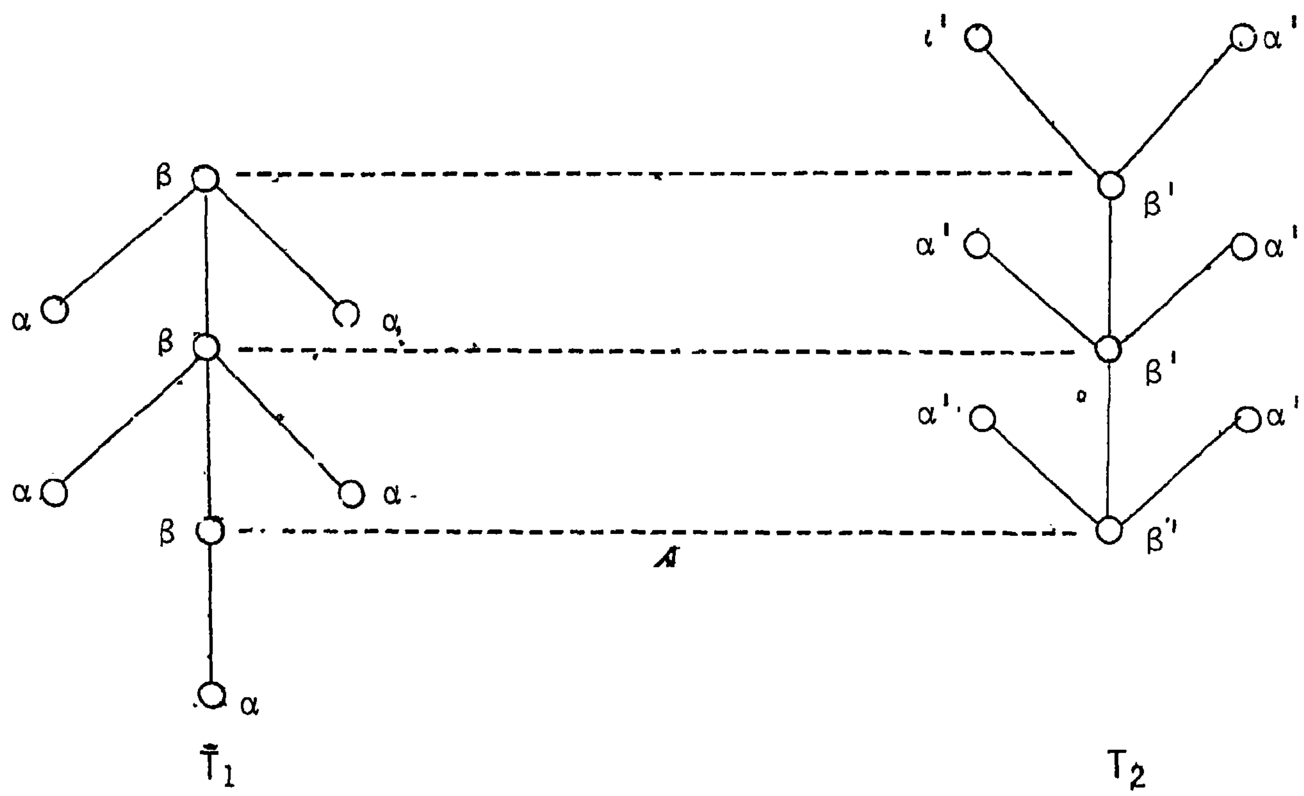
It appears that the language

$$\{\alpha_1^n \alpha_2^n \alpha_1^n | n \geq 1\}$$

can be generated by nominal neighborhood grammars in an essentially context-free manner. The nominal neighborhood grammars are new to me and appear to offer considerable generating power at the usual expense of a complex definition.

Chapter IV treats syntax-directed translations and certain extensions thereof using the idea of neighborhoods Most of their development is now standard (Aho and Ullman, The Theory of Parsing, Translation, and Compiling:

Vol. 1, Parsing, Prentice-Hall, Englewood Cliffs, N. J., 1972) and has been advanced to truely elegant abstractions by Alagic (Natural State Transformations, J. Comp. Sys. Sci. 10: 2 (Apr 1975), 266-307.) However, the use of neighborhoods allows for the extension of syntax-directed translations in new directions, best indicated here by the authors' diagram of a translation from tree T_1 to tree T_2.



I can't think of any use for this order-reversal in carrying out the translation, but it is an interesting idea nonetheless.

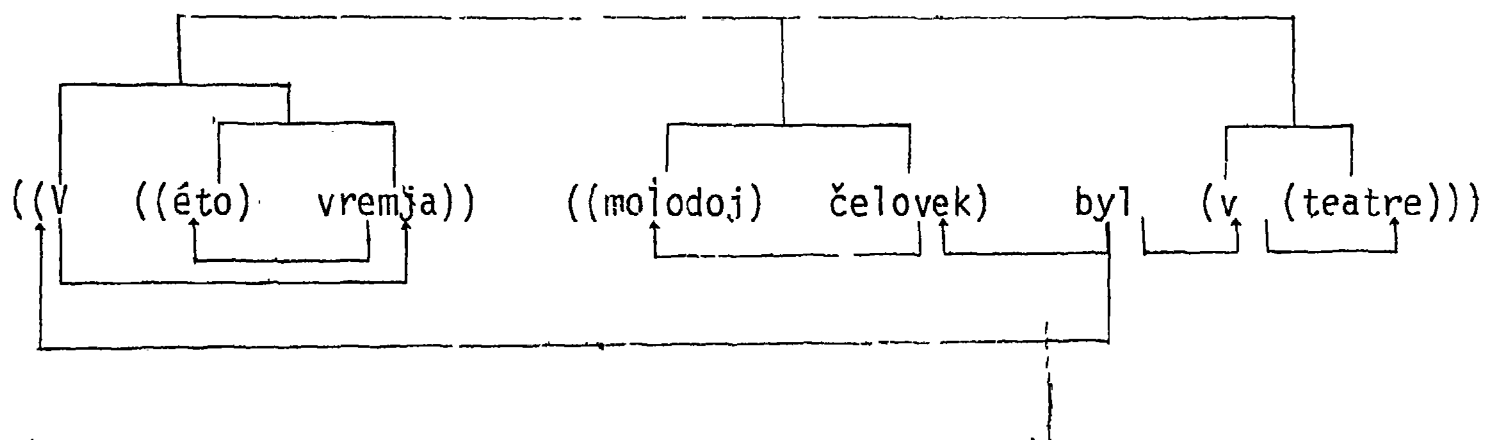
3. S.-Ja. Fitialov

   On the Equivalence of IC Grammars and Dependency Grammars

   pages 115-158.

   According to the author, both the direction and nesting of syntactic relationships should be accounted for in a sufficiently adequate and complete linguistic description. As dependency grammars handle direction and Immediate Constituent grammars handle nesting, the question of the relationship between the two descriptive mechanisms arises.

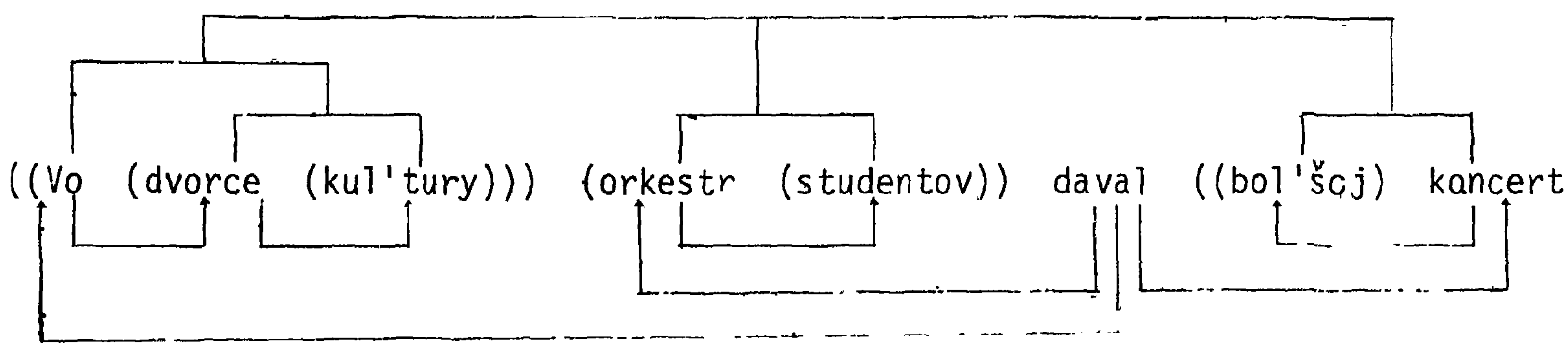
As the phrase names (non-terminal symbols) can not be determined from the dependencies, the IC structure considered consists solely of the tree. This is best illustrated by the following example. The element groups in the dependency structure are enclosed in parentheses, the dependent directions is shown below the sentence and the IC tree is shown above. From page 128:



(At that time the young man was in/at the theater)

Fig. 1.

Two dependency structures can give rise to a single IC structure. Compare with Fig. 1.



(The student orchestra gave a big concert in the Palace of Culture)

Fig. 2.

Furthermore, Fitialov gives examples in which the same sentence can have two different IC trees. Thus the "equivalence" is many-to-many. The author then sets up an algorithm to construct a dependency grammar from certain IC grammars. The IC grammar must have "finite degree", a technical concept that need not detain us. The final topic is carrying the idea of "degree of nesting" from IC structures with non-terminals over to dependency structures.

Much of this paper is apparently devoted to clarifying the ideas presented by Gaifman (Dependency Systems and Phrase-Structure Systems, Inform. & Control 8(1965), 304-337). I found little of interest in this selection.

4 A. Gladkij

An Attempt at the Formal Definition of Case and Gender of the Noun. pages 159-204.

With the recent interest in case in computational linguistics, this paper by the foremost Russian formal languages expert should appeal to those who wish to build logically coherent case structures. The mathematics is minimal but suggestive. The focus of the work is on a classification of (Russian) nouns. I am in no position to comment on the quality of the classification system proposed. Nonetheless, here is a sketch of the method.

Let V be the set of words. These are called segments by Gladkij to stress the graphical sense of word he is using. Each subset of V having "identical lexical meaning" is called a neighborhood. Thus:

{DOM, DOMA, DOMU, DOMOM, DOME, DOMA, DOMOV, DOMAM, DOMAMI, DOMAX} (house)

Let , a subset of V, be the set of nouns. "The set S should be a union of some neighborhoods." The next notion is subordination or dependency. Say that x (potentially) subordinates y if there is a sentence in which some occurrence of segment x "syntactically directly subordinates" some occurrence of segment y. Now let 0 be any neighborhood. Say that 0 subordinates y if y is subordinate to at least one segment in 0. Let N_0 be the set of all S-segments (noun words) which are subordinate to 0. A set N_0 is said to be <u>minimal</u> if N_0 is not empty and there are no non-empty $N_{0'}$ which are proper subsets of N_0. The minimal sets N_0 are said to be <u>cases</u>. "If two different neighborhoods 0 and 0' of the sets N_0 and $N_{0'}$ coincide, we will not consider N_0 and $N_{0'}$ to be different cases, but one and the same case."

Gladkij gives examples of all these concepts, including the distinction between minimal and non-minimal neighborhoods. He goes on to show that the cases are not necessarily mutually disjoint, and then uses the development to explicate the "special position" occupied by the second prepositional and second gentive cases in Russian grammar. Gladkij then shows, to no one's surprise, that there are instances in which meaning, even the meaning of the prior several sentences, must be taken into consideration to determine the case of certain words. If one's purpose is to understand the text, then in these instances the case structure won't help. In most sentences however, it will clarify the relationships of the segments in the sentence and thus aid understanding. Whether Gladkij's formulation is more useful than unaided intuition and knowledge of the language is for others to judge.

The last sixteen pages of the paper develop a similar formalism for the concept of <u>coordinated</u> <u>class</u>, apparently as an aid to arriving, at the very end of the paper, in a definition of <u>gender</u>. The mathematics is very easy, but the Russian examples are not--for this reviewer,

5. Ju. K. Lekomcev

On Models for a Syntax with Explicitly Differentiated Elements (D-Syntax). pages 205-239.

This paper is, by Western standards, fussy and pedantic. One must suppose that the editor's selection was rather more random than less. Despite the following quotation from the introduction--

> "Concerning the characteristic of a D syntax model, it should be noted that our model is a continuation of the glossematic variant of the Saussurian trend, partly complemented by Russian and American concepts. The notions of syntagmatic-paradigmatic relations and of distinctive features lie at the heart of the concept."

--I was disappointed. The mathematical model, stated in the complete formality of first-order predicate logic, actually says very little. The foundation of the paper's development is a notion of differentiation system (DS). A DS is basically a system of lists of the values of attributes. Thus two element (i.e., lists) differ if some value of some common attribute differs. Actually the paper develops somewhat more complex differentiation systems, but the additional complexities are obvious, not requiring such an overly formal development.

This notion is then applied to the question of generating (resp., analyzing) words from phonemes, in a fashion that would have produced more insight if it had been treated in automata-theoretic terms. The concluding remarks--on applying DS to semantics--seem to this reviewer to be irrelevant, or else more clearly presented elsewhere.

6. Ju. A. Srejder

   On the Contrast between the Concepts 'Language Model' and 'Mathematical Model'

   pages 241-267.

> "The concept "language model" is widely used in structural and mathematical linguistics. In a certain sense, this concept is the cornerstone of these branches or linguistics, where so called formalized or precise methods have taken root. It is of some use, therefore, to gain an understanding of just what is meant with the words "language model"."

The author, evidently a mathematician, contrasts the notion of a mathematical theory and a language model. In the terms of mathematical logic, a <u>theory</u> consists of names for relations or functions, names for variables, a method of contructing well-formed formulae (wff), the system of formal deduction to be used, and the axioms of theory. A <u>model</u> of a theory is a system of sets and relations such that "if relation R_i is compared to every name of relation $\mathbb{R}_i$ in such a way that if variables x,y,z,... are explained as elements of set $\mathbb{M}$ all formulae of the given Theory are true."

After several examples of mathematical theories and models--the theory of partial orders and a model of it in the natural numbers is one--the author gives a fairly strong argument that what many linguistics "call a model is in mathematics known as a thoery." He then gives some examples of 'language models' to give substanence to this thesis. The most interesting is an all-too-brief discussion of the poem "Eugene Onegin" in which the <u>theory</u> is

> "An accented syllable can be located only on an even-numbered place from the beginning of the line."

for which presumably a standard edition of "Eugene Onegin" stands as a model. He continues by giving a short neighborhood grammar as the axioms of the theory. In an appendix he shows that Chomsky's "generative model of context-free grammars is in fact a particular mathematical theory.

However, linguistics is not mathematics and the models (i.e., the actual utterances or texts) fail to satisfy all details of the theory. Thus the author suggests

"1) The quasi-model of a Theory, i.e., the set in which the theory is almost fulfilled (for this it is necessary to introduce a measure onto the Theory) and;

2) The measure on a class of quasi-models of a given Theory, which allows us to say that the Theory can be fulfilled for almost all quasi-models."

Unfortunately, these fine ideas are not developed. Nonetheless, this paper does help explain the terminological differences between mathematicians and linguists.

.7. E. D. Stockij

Generalized Grammars and Their Properties,

pages 269-284.

"Let us assume that in grammar Γ not all derivations are permissible, but only those which can themselves be generated by another grammar Γ', which is working as the device for programming of the derivation. We shall investigate the question of how the selection of a strategy of phrase generation in grammar

Γ (in other words, selecting grammar Γ') affects the generative capacities of grammar Γ."

Let $X_0 \Rightarrow X_1 \Rightarrow \ldots X_n$ be a derivation in grammar Γ. Let V' be a set of names for the rules of Γ. Thus the derivation corresponds to a word $\rho_1\rho_2\ldots\rho_n \varepsilon (V')^*$ where ρ_i is the name of the rule doing the rewriting $X_{i-1} \Rightarrow X_i$. Each word over V' that corresponds to a derviation is called a control word. In general there is no one-to-one correspondence between the derivations and their control words.

A generalized grammar is a pair of grammars (Γ,Γ') such that the second grammar is used to control the first. Specifically, $X_0 \Rightarrow X_1 \Rightarrow \ldots \Rightarrow X_n$ is an allowable derivation of Γ if and only if there is at least one control word corresponding to it in $L(\Gamma')$. The language of the generalized grammar is that subset of $L(\Gamma)$ for which each word has at least one allowable derivation. Note that there is no requirement that the derivations be canonical (left-most).

Grammars in this paper are classified by the usual Chomsky Hierarchy into types 0, $\bar{1}$, 2, 3. Then generalized grammars have type (i,j) where i is the type of the language-producing grammar and j is the type of the controling grammar. Let D_{ij} be the class of languages generated by all generalized grammars of type (i,j), and D_i be the class of languages generated by all (ordinary) grammars of type i.

The main portion of the paper presents, without proof, the relationships known among the D_{ij} as of April, 1969, excluding some American work such as Ginsburg and Spanier's (Derivation bounded languages, J. Comp. Sys. Sci. 2:3(1968), 228-250). Most of the references cited--which contain the proofs--are to Stockij's own work on these questions. Example results are:

$D_{3j} = D_j$, j = 1,2,3.

$D_{13} = D_1$

$D_{23} \subset D_1$

$D_2 \subset D_{23}$

$D_{00} = D_{01} = D_{02} = D_{03} = D_0$

$D_{10} = D_{20} = D_{30} = D_0$ (without null words)

This last is a consequence of Stockij s disallowing rewritings to the null word in grammars of type 1, 2 and 3.

Now consider the set of control words, $P(\Gamma)$, of an ordinary uncontrolled grammar, Γ. Let $tP(i)$ denote the type of the language $P(\Gamma)$ for grammar Γ of type i. The following are representative results.

$tP(3) = 3$

$tP(2) = 1$

$tP(0) = 1$

The study of controlled grammars arises from the psycholinguistic idea that derivations are controlled by a "generation program" which determines the semantics of the phrases and their grammatical structure. Thereby, the results presented here presumably explicate the potential grammatical structures which such a generation program could possibly produce. Whether or not one accepts the "generating program" hypothesis, these are nice results in formal language theory.