

## XXIII. LINGUISTICS\*

Prof. R. Jakobson	T. G. Bever	T. M. Lightner
Prof. J. Kurylowicz	J. L. Fidelholtz	P. S. Peters, Jr.
Prof. M. Halle	R. S. Glantz	C. B. Qualls
Prof. G. H. Matthews	J. S. Gruber	J. R. Ross
Prof. P. M. Postal	Barbara C. Hall	S. A. Schane
Prof. P. Colaclides	T. J. Kinzer III	R. J. Stanley
Dr. Paula Menyuk	R. P. V. Kiparsky	J. J. Viertel
Dr. W. A. O'Neill	S.-Y. Kuroda	A. M. Zwicky, Jr.

### RESEARCH OBJECTIVES

The research of the linguistics groups aims to develop a general theory of language which encompasses all that can be known about language. This theory attempts to reveal the lawful inter-relations existing among the structural properties of different languages and among the different levels of a given language. As regards subject matter, therefore, all aspects of language are of interest to our group. Work now in progress deals with the phonology, morphology, and syntax of a score of different languages and the abstract features of these linguistic levels, with language learning, language disturbances and speech perception, with linguistic change (syntactic as well as phonological), with semantics, the philosophy of language and the history of ideas concerning the nature of language, with the poetic use of language and the structure of literary works, with the mathematical and logical foundations of linguistic theory, as well as with the abstract study of symbolic systems similar to natural languages.

Since many of the problems of language lie in the area in which several disciplines overlap, an adequate and exhaustive treatment of language demands close cooperation of linguistics with other sciences. The inquiry into the structural principles of human language suggests a comparison of these principles with those of other sign systems, which, in turn, leads naturally to the elaboration of a general theory of signs, semiotics. Here linguistics touches upon problems that have been studied by philosophy. Other problems of interest to logicians – and also to mathematicians – are touched upon in the studies devoted to the formal features of a general theory of language. The study of language in its poetic function brings linguistics into contact with the theory and history of literature. The social function of language cannot be properly illuminated without the help of anthropologists and sociologists. The problems that are common to linguistics and the theory of communication, the psychology of language, the acoustics and physiology of speech, and the study of language disturbances are too well known to need further comment here. The exploration of these interdisciplinary problems, a major objective of this group, will be of benefit not only to linguistics; it is certain to provide workers in the other fields with stimulating insight and new methods of attack, as well as to suggest to them new problems for investigation and fruitful reformulations of questions that have been asked for a long time.

M. Halle

### A. NOTE ON THE MOTIVATION FOR USING TRANSFORMATIONAL RULES IN PHONOLOGY

In the development of liquid diphthongs in South- and East-Slavic,<sup>1</sup> one can account for the insertion of vowels by application of any of the three rules given below:

---

\* This work was supported in part by the National Science Foundation (Grant GP-2495), the National Institutes of Health (Grant MH-04737-04), and the National Aeronautics and Space Administration (Grant NsG-496); and in part by the U. S. Air Force (Electronic Systems Division) under Contract AF19(628)-2487.

## (XXIII. LINGUISTICS)

(1)  $\phi \rightarrow \begin{bmatrix} + \text{vocal} \\ - \text{cons} \\ - \text{tense} \\ - \text{diff} \\ \underline{a \text{ grave}} \end{bmatrix}$  in env:  $\text{---} \begin{bmatrix} + \text{vocal} \\ - \text{cons} \\ - \text{tense} \\ - \text{diff} \\ \underline{a \text{ grave}} \end{bmatrix} \text{ L C}$

(2)  $\phi \rightarrow \begin{bmatrix} + \text{vocal} \\ - \text{cons} \\ - \text{tense} \\ - \text{diff} \\ \underline{a \text{ grave}} \end{bmatrix}$  in env:  $\begin{bmatrix} + \text{vocal} \\ - \text{cons} \\ - \text{tense} \\ - \text{diff} \\ \underline{a \text{ grave}} \end{bmatrix} \text{--- L C}$

(3) Struct. Descr:  $V \underbrace{\text{L C}}_{1 \ 2}$

Struct. Change:  $12 \rightarrow 112$

In these rules we used the following abbreviations:

V for any  $\begin{bmatrix} + \text{voc} \\ - \text{cons} \\ - \text{tense} \\ - \text{diff} \end{bmatrix}$  segment

L for any  $\begin{bmatrix} + \text{voc} \\ + \text{cons} \end{bmatrix}$  segment

C for any  $[+ \text{cons}]$  segment

Rules (1) and (2) are context-sensitive rewriting rules, whereas rule (3) is an elementary transformational rule. Since all three rules derive doorg from dorg, beerg from berg etc., the question arises as to whether the context-sensitive rewriting method or the transformational method of describing this phenomenon is correct.

It is clear that rule (3) expresses the linguistic fact of reduplication in a more natural way than either rule (1) or rule (2) does. Thus the facts are that lax, nondiffuse vowels are reduplicated in position before liquid followed by consonant. In rules (1) and (2), however, it is necessary to make an additional specification for the feature gravity.<sup>2</sup> In terms of reduplication, then, we prefer the formulation of rule (3). It is possible, however, that in terms of the remaining rules of the grammar, one might find reason to prefer either rule (1) or rule (2) to rule (3), and it is this question that we examine here.

The question will concern the exact place of insertion: in rule (1), the inserted vowel is placed before the original vowel; in rule (2), after the original vowel; in rule (3), the original vowel is reduplicated without specification of position.

In arguing for the context-sensitive rewriting rules (1) and (2), one can bring the following facts to bear: Slavic liquid diphthongs with acute accent develop into Russian forms that show stress after the liquid (dorg  $\rightarrow$  doróg); grave accent, however, develops into Russian forms that show stress before the liquid (bèrg  $\rightarrow$  béreg).

If one interprets acute accented vowels as being marked [+ accent], and grave accented vowels as being marked [- accent], then one can propose that rule (1) is correct, and that it must be modified in the following manner:

$$(1') \phi \rightarrow \begin{bmatrix} + \text{vocal} \\ - \text{cons} \\ - \text{tense} \\ - \text{diff} \\ \alpha \text{ grave} \\ - \text{accent} \end{bmatrix} \quad \text{in env: } \underline{\hspace{1cm}} \begin{bmatrix} + \text{vocal} \\ - \text{cons} \\ - \text{tense} \\ - \text{diff} \\ \alpha \text{ grave} \end{bmatrix} \text{ L C}$$

The derivations of Russian doróg, béreg are now the following:

$$\begin{aligned} \text{dórg} &\rightarrow 1' \rightarrow \text{doórg} \rightarrow \text{metathesis} \rightarrow \text{doróg} \\ \text{berg} &\rightarrow 1' \rightarrow \text{beerg} \rightarrow \text{metathesis} \rightarrow \text{bereg} \end{aligned}$$

We now require a rule that places stress on the first vowel of forms none of whose vowels are accented:

$$\text{bereg} \rightarrow \text{béreg}$$

Application of this initial stress rule will correctly derive stress in forms like ná bereg (from na bereg). Derivation of forms like za béreg, however, will require us to formulate an additional rule because the underlying form za bereg would give \*zá bereg. If we postulate that the initial stress rule places stress on the first vowel after # if the form in question contains no accented vowel, then the underlying form #za#bereg# will give #zá#béreg#. The additional rule that we propose is that in forms with more than one accented vowel, only the final accented vowel is stressed:

$$\text{\#zá\#béreg\#} \rightarrow \text{za béreg}$$

In order to derive ná bereg, we mark the root berg for the idiosyncratic feature of dropping # after the preposition na:

$$\text{\#na\#bereg\#} \rightarrow \text{\#na bereg\#} \rightarrow \text{ná bereg}$$

Acceptance of rule (1') has forced us into the position of postulating two additional rules:

- (4) Place stress on first vowel after # if the form in question contains no accented vowels.
- (5) Place stress on the last accented vowel of any form.

The argument in favor of using rule (1) has now disappeared, for, given rules (4) and (5), the correct results can be obtained either from application of a modified form of rule (2) or from application of rule (3). The modified form of rule (2) is the following:

(XXIII. LINGUISTICS)

$$(2') \phi \rightarrow \left[ \begin{array}{l} + \text{vocal} \\ - \text{cons} \\ - \text{tense} \\ - \text{diff} \\ \alpha \text{ grave} \\ \beta \text{ accent} \end{array} \right] \quad \text{in env:} \quad \left[ \begin{array}{l} + \text{vocal} \\ - \text{cons} \\ - \text{tense} \\ - \text{diff} \\ \alpha \text{ grave} \\ \beta \text{ accent} \end{array} \right] \quad \text{--- L C}$$

The derivations of doróg, béreg, ná bereg and za béreg resulting from application of rule (2') are:

dórg →<sup>2'</sup>→ dóórg →metathesis→ dóróg →<sup>5</sup>→ doróg  
 berg →<sup>2'</sup>→ beerg →metathesis→ bereg →<sup>4</sup>→ béreg  
 na#berg →<sup>2'</sup>→ na#beerg →metathesis→ na#bereg →# deletion→  
     na bereg →<sup>4</sup>→ ná bereg  
 za#berg →<sup>2'</sup>→ za#beerg →metathesis→ za#bereg →<sup>4</sup>→  
     zá#béreg →<sup>5</sup>→ za béreg

The derivations with rule (3) used are identical to those given directly above. Since there is no external reason to prefer (1') or (2') to (3), we choose (3) for the reason outlined at the beginning of this report. Additional confirmation for the correctness of our decision can be drawn from the arbitrariness of the formulation of rules (1') and (2'). If one restricts the rules of the phonological component to simple rewriting rules, then there is no way to choose whether V should be inserted after V followed by L C or before V followed by L C — either formulation will derive correct representations. In the transformational description, however, there is no question as to whether the inserted vowel precedes or follows the original vowel.

T. M. Lightner

References

1. In this paper we assume familiarity with M. Halle and T. M. Lightner, On the phonology of tort, tolt, tert, telt in Old Church Slavonic and Russian, Quarterly Progress Report No. 75, Research Laboratory of Electronics, M.I.T., October 15, 1964, pp. 121-123.

2. There is reason to believe that in Russian, all lax vowels — regardless of their diffuseness — are reduplicated before liquid followed by consonant. In this case, then, the formulation of rules (1) and (2) requires two additional, unnecessary specifications, one for gravity and one for diffuseness.

B. ON THE DEVELOPMENT OF turt tirt tult tilt IN RUSSIAN

It is well known that Proto-Slavic turt tirt tult tilt developed into Modern Russian tort tert tolt tolt, respectively. In this report we shall examine evidence indicating

that this historical development did not involve a simple lowering of lax vowels before liquid followed by consonant.

Any discussion of the development of turt tirt, etc. in Modern Russian is complicated at the outset by the problem of those forms which have undergone secondary polnoglasié: sumerečnyj (cf. sumerki), dolog, polon, bestoloč (cf. tolk), verëvka (cf. verv'), zolovka etc. We shall not consider this problem here, but will confine our discussion to clear cases like torg, xolm, tvërdyj, pervyj, etc. Also, we shall omit from discussion the Church Slavic element in Russian.

Our proposal is that our earlier analysis of tort tolt tert telt be generalized to include all lax vowels.<sup>1</sup> Thus we propose the following set of rules:

(1) Struct. Descr:  $V \underbrace{L V}_{1 \quad 2}$

Struct. Change: 12 → 112

(2) SD:  $V \begin{matrix} L \\ 1 \end{matrix} \begin{matrix} C \\ 2 \end{matrix} \begin{matrix} \\ 3 \end{matrix}$

SC: 123 → 213

In these rules we used the following abbreviations:

V for any  $\begin{bmatrix} + \text{vocal} \\ - \text{cons} \\ - \text{tense} \end{bmatrix}$  segment

L for any  $\begin{bmatrix} + \text{vocal} \\ - \text{cons} \end{bmatrix}$  segment

C for any [+ cons] segment

Following are sample derivations:

torg: turg →<sup>1</sup> tuurg →<sup>2</sup> turug

cerkov': cirk →<sup>1</sup> ciirk →<sup>2</sup> cirik

mërznut': mirz →<sup>1</sup> miirz →<sup>2</sup> miriz

The later development included the following rules:

(3) Sharping of consonants before acute nonconsonantal segments.

(4) Desharping of c.

(5) Loss of jers. In particular, to strengthen jers before liquid-jer clusters, and weaken jers after jer-liquid clusters. We shall not make this rule precise as it involves obvious problems. Note, however, that a relaxation of the restrictions of this rule will permit us to account for the forms with secondary polnoglasié.

(6) Desharping of r before dentals.

(XXIII. LINGUISTICS)

- (7) Replacement of  $\underline{e}$  by  $\underline{\ddot{e}}$  before nonsharp consonants and in certain categories, such as loc. sg. ( $\underline{b, er, \ddot{e}z, e} < \underline{b, er, ez, e}$ ), short form adj. ( $\underline{t, \ddot{e}m, \ddot{e}n} < \underline{t, em, en}$ ), and in 2 pl. pres. ( $\underline{n, es, \ddot{e}t, e} < \underline{n, es, et, e}$ ).

Following are sample derivations:

měrznut': miriz  $\rightarrow 3 \rightarrow$  m, ir, iz  $\rightarrow 5 \rightarrow$  m, er, z  $\rightarrow 6 \rightarrow$  m, erz  $\rightarrow 7 \rightarrow$  m, ěrz

cerkov': cirik  $\rightarrow 3 \rightarrow$  c, ir, ik  $\rightarrow 4 \rightarrow$  cir, ik  $\rightarrow 5 \rightarrow$  cer, k<sup>3</sup>

T. M. Lightner

References

1. See M. Halle and T. M. Lightner, On the phonology of tort, tolt, tert, telt in Old Church Slavonic and Russian, Quarterly Progress Report No. 75, Research Laboratory of Electronics, M.I.T., October 15, 1964, pp. 121-123.

2. In the development given here we do not make any provision for stages of development intermediate between Proto-East-Slavic and Modern Russian. Thus, for example, it seems reasonable to assume that at some stage in the development of Old Russian the jer-liquid clusters were pronounced as syllabic liquids; for convincing argument in support of this position, see V. S. Golyšenko, K voprosu o kačestve plavnogo v kornjax, vosxodjaščix k \*turt, \*tirt, \*tult v drevnerusskom jazyke XII-XIII vv., Istoričeskaja grammatika i leksikologija russkogo jazyka (Moscow, 1962), pp. 20-28. Note also that the rules of development given here do not reflect chronological order. Thus, for example, sharpening of consonants (rule (3)) is clearly a chronologically earlier phenomenon than reduplication of vowels (rule (1)) or metathesis (rule (2)).

3. For discussion of the pronunciation of sharp [r,] in forms like cerkov', pervyj, etc., see L. A. Bulaxovskij, Istoričeskij komentarij k russkomu literaturnomu jazyku (Kiev, 1958), p. 98.

C. FINITE STATE REPRESENTATIONS OF CONTEXT-FREE LANGUAGES

1. We shall show the existence of a class C of regular (finite state) languages with the property that

(i) There is a fixed finite alphabet  $V_0$  from which each language in C is constructed;

(ii) Given any context-free (CF) language L there is a member  $R_L$  of C such that  $R_L$  is equivalent, in some sense, to L. More precisely, there is a single CF language K and a fixed homomorphic operation  $\phi$  (i.e., a one-state finite transducer  $\phi$ ) such that for each CF language L

$$L = \phi(K \cap R_L).$$

The language K and the operation  $\phi$  have a very simple and natural interpretation. Thus, while CF languages are a richer class than regular languages, each CF language L still has a simple and natural representation in terms of a regular language  $R_L$ .

We have referred implicitly above to "the class of all CF languages"; it is crucial that we be explicit about the vocabularies from which the terminal and nonterminal

symbols for these CF languages are drawn. Let  $V_T$  be the universal vocabulary from which we pick the terminal symbols of each CF language. The case for which  $V_T$  is finite is of most interest, and involves no real loss in generality. We thus assume in the following discussion that  $V_T$  is fixed and finite. If, however,  $V_N$  is the universal vocabulary from which we pick the nonterminal symbols of each CF language, it is apparent that we do lose generality if we let  $V_N$  be finite; for example, if  $V_N$  were finite, then the union of two CF languages need not be CF. Thus we shall assume that  $V_N$  is fixed and infinite.

Given this framework, we show the existence of the class C. Our method of doing this is essentially that of Chomsky,<sup>1</sup> i.e., we associate with each CF language L (more precisely, with some CF grammar generating L) a finite transducer  $T_L$  which takes finite strings over  $V_T$  as input. For any given input  $x$ , the output  $T_L(x)$  of  $T_L$  is the string  $x$  together with a bracketing of  $x$ . The set of all outputs  $T_L(x)$  is a language that, as we will show, has all of the desired properties of  $R_L$  given above. We cannot, however, use Chomsky's method of constructing  $T_L$  from L as it stands. The reason is that in his construction we need, for each nonterminal symbol "A" used in the grammar for the CF language L, two bracket symbols " $_A$ " and " $_A$ " in the output vocabulary of  $T_L$ . Thus the output vocabulary (call it  $V_O$ ) of the whole class of transducers  $T_L$  for all CF languages L must be infinite, since  $V_N$  is infinite. It is crucial, however, in showing that K is context-free that  $V_O$  be finite. We avoid this difficulty by showing that, notwithstanding the fact that  $V_N$  is infinite,  $V_O$  can be made to be finite. We do this by showing that there is a finite set S of bracket symbols (in fact, a set of 8 bracket symbols) such that each single symbol " $_A$ ", or " $_A$ ", for  $A \in V_N$ , can be represented as a (finite) sequence of symbols of S in such a way that the desired properties of the output languages of the transducers  $T_L$  still hold.

Our task, then, is to define a particular CF language K, a homomorphism  $\phi$ , and an algorithm for associating with each CF language L a transducer  $T_L$  such that the relation above holds. We proceed in two steps. First, we construct a normalized grammar for each CF language. Second, we show how this normalized grammar can be used to define the instructions for the desired transducer. Except for the above-mentioned modification our proof is exactly that of Chomsky.

The details of the proof may be found in Stanley<sup>2</sup> and in Chomsky.<sup>1</sup> In this report we describe only the form of the normalized grammar, which is of interest in itself.

2. Consider a context-free grammar G that meets the following conditions.

- (I) Every nonterminating rule of G is of the form  $A \rightarrow BC$ , where B and C are single symbols of  $V_N$ .
- (II) Every terminating rule of G is of the form  $A \rightarrow a$ , where a is a single symbol of  $V_T$ .

(XXIII. LINGUISTICS)

(III)  $G$  has no rules of the form  $A \rightarrow BB$ .

(IV) If  $A \rightarrow BC$  is a rule of  $G$ , then  $D \rightarrow CE$  is not a rule of  $G$  for any  $D, E$ .

Conditions (I)-(IV) allow us to adopt the following notational convention in writing rules of  $G$ . Let the members of  $V_N$  be represented as  $L_1, L_3, L_5, L_7, \dots$ , and  $R_2, R_4, R_6, R_8, \dots$ , called L-symbols and R-symbols, respectively, where  $L_i \neq R_j$  for all  $i, j$ . Then we can write each rule of  $G$  in one of the four following forms:

$$L_i \rightarrow L_k R_\ell \quad R_j \rightarrow L_m R_n$$

$$L_i \rightarrow a \quad R_j \rightarrow a.$$

(V)  $G$  contains no pairs of rules of the form

$$L_j \rightarrow L_i R_k, \quad R_m \rightarrow L_i R_n.$$

That is, an L-symbol  $L_i$  cannot be dominated in  $G$  by both an L-symbol ( $L_j$ ) and an R-symbol ( $R_m$ ).

(VI)  $G$  contains no pairs of rules of the form

$$L_j \rightarrow L_i R_k, \quad R_m \rightarrow L_n R_k.$$

This is simply the analog of (V) for R-symbols.

Conditions (V) and (VI) say that any given symbol in  $G$  is dominated either by L-symbols or by R-symbols, but never by both.

(VII) If  $G$  contains the pair of rules

$$L_i \rightarrow L_j R_k, \quad L_p \rightarrow L_j R_s,$$

then, whenever  $G$  has a rule  $L_i \rightarrow L_m R_n$ ,  $G$  also has the rule  $L_p \rightarrow L_m R_n$ , and conversely. Similarly, if  $G$  contains the pair of rules

$$R_i \rightarrow L_j R_k, \quad R_p \rightarrow L_j R_s,$$

then, whenever  $G$  has a rule  $R_i \rightarrow L_m R_n$ ,  $G$  also has the rule  $R_p \rightarrow L_m R_n$ , and conversely. That is, any symbol that dominates a given L-symbol  $L_j$  in  $G$  dominates exactly the same pairs as any other symbol dominating  $L_j$ . (Note that we do not attempt to impose on  $G$  a condition on R-symbols analogous to this one on L-symbols.)

A context-free grammar meeting conditions (I)-(VII) will be called an  $\alpha$ -normalized grammar. We can show, by means of a long but fairly straightforward proof, that every content-free language can be generated by an  $\alpha$ -normalized grammar. Chomsky has pointed out that there is a mistake in his article<sup>3</sup> and that when he refers to a modified



normal grammar he actually needs a stronger form of normalization. This stronger form is precisely what I have called  $\alpha$ -normalization.

3. The relation  $L = \phi(K \cap R_L)$  is an important result in the theory of CF languages. It not only shows that CF languages can represent aspects of symmetry which cannot be represented in regular languages (which, perhaps, is intuitively clear), but it also shows that this symmetry property is the only property that CF languages have and regular languages do not. This follows from our relation which shows that every CF language  $L$  is homomorphic to just those members of some regular language which possess certain properties of symmetry. The properties of symmetry are precisely what characterizes sentences of  $K$ , a fact that becomes clear in a complete proof of our relation.

R. J. Stanley

#### References

1. N. Chomsky, Formal Properties of Grammars, in R. R. Bush, E. H. Galanter, and R. D. Luce (eds.), Handbook of Mathematical Psychology (John Wiley and Sons, Inc., New York, 1963); see especially Sections 1.6 and 4.2.
2. R. Stanley, Finite State Representations of Context-free Languages (submitted for publication to Information and Control).
3. N. Chomsky, Formal Properties of Grammars, op. cit., see p. 374.

