

## XVIII. LINGUISTICS\*

Prof. R. Jakobson  
Prof. A. N. Chomsky  
Prof. M. Halle  
Prof. Y. Isami  
Dr. G. H. Matthews  
Dr. Paula Menyuk

Dr. K. Wu  
T. G. Bever  
S. K. Ghosh  
Barbara C. Hall  
J. J. Katz  
R. P. V. Kiparsky  
S.-Y. Kuroda

D. T. Langendoen  
T. M. Lightner  
P. M. Postal  
C. B. Qualls  
J. J. Viertel  
A. M. Zwicky, Jr.

### A. SEMANTIC INTERPRETATION OF IDIOMS AND SENTENCES CONTAINING THEM

#### 1. Two Types of Idioms

In the present report we examine the semantics of idioms in natural languages. We show how a class of idioms may be treated in terms of the recently developed conception of the semantic component of a linguistic description.<sup>1-4</sup> Familiarity with this conception is assumed throughout.

The essential feature of an idiom is that its full meaning, and more generally the meaning of any sentence containing an idiomatic stretch, is not a compositional function of the meanings of the idiom's elementary grammatical parts. For example, the meaning of the idiom kicked the bucket cannot be regarded as a compositional function of the meanings kick, ed, the, bucket, regardless of the syntactic structure attributed to kicked the bucket by the structural descriptions of the sentences in which it appears. Hence the projection rules that a semantic theory provides to obtain the meaning of compound expressions and sentences as a compositional function of the meanings of their elementary parts cannot obtain the idiomatic meaning of an idiomatic stretch from the meanings of the syntactically atomic parts of that stretch. Therefore, the fact that no projection rules at all are employed in obtaining semantic interpretations for whole idiomatic stretches is the formal representation that a semantic component gives of the idiomatic status of such stretches.

Before showing the manner in which a semantic theory can provide such semantic interpretations for idiomatic stretches, it is necessary to differentiate two sorts of things that are traditionally referred to as idioms. The characterization of an idiom as any concatenation of two or more morphemes whose compound meaning is not compositionally derived from the meanings of the concatenated morphemes does not differentiate those idioms that are syntactically dominated by one of the lowest level syntactic categories, i.e., noun, verb, adjective, etc., from those whose syntactic structure is such that no single lowest level syntactic category dominates them. Let us call the

---

\*This work was supported in part by the National Science Foundation (Grant G-16526); in part by the National Institutes of Health (Grant MH-04737-03); and in part by the U. S. Air Force (Electronics Systems Division) under Contract AF19(628)-2487.

(XVIII. LINGUISTICS)

former type 'lexical idioms', the latter 'phrase idioms'. We shall be concerned primarily with the latter type.

The syntactic component of a linguistic description contains two parts, a set of syntactic rules and a lexicon. The lexicon contains all of the lexical morphemes. The grammatical morphemes are all introduced by the syntactic rules. These rules generate derivations whose final lines contain not lexical items but particular grammatically marked positions in which lexical items may be placed. There is then a substitution condition that permits lexical items from the lexicon to be substituted into derivations provided that the grammatical markings of the lexical item are compatible with those of the grammatical position into which it is substituted. Most of the entries in the lexicon will presumably be single morphemes, e.g., book, run, hate, big, etc., but a significant number will be compounds of two or more morphemes, e.g., tele+phone, bari+tone, etc. These are the entries for lexical idioms. They are marked as idioms by virtue of the fact that in the dictionary of the semantic component (not identical with the lexicon of the syntactic component) these sequences of two or more morphemes are directly assigned readings that represent their senses. These readings are not the result of projection rules amalgamating readings for individual morphemes listed separately in the dictionary.

It might be maintained that phrase idioms should also be handled simply by listing each of them as a single lexical item and assigning each a set of readings. This would mean eliminating the distinction between lexical and phrase idioms by also treating the latter in the lexicon as representatives of the lowest level syntactic categories. It would mean that an expression such as kicked the bucket in its idiomatic meaning of 'died' would be regarded as a compound intransitive verb. However, it is easy to show that not all idioms can be regarded as lexical idioms. This follows because there is a large class of idioms, like kicked the bucket, which have a compositional meaning, as well as an idiomatic one. These cannot be regarded as compound lexical items on a par with tele+phone, bari+tone, etc., because listing elements like kicked the bucket in the ordinary lexicon unnecessarily complicates both the syntactic and phonological components of a linguistic description. First, for every case of an idiomatic stretch  $x$  which also has a compositional meaning, regarding  $x$  as a lexical idiom requires that a new entry be added to the syntactic lexicon (e.g., intransitive verb = kicked the bucket). This addition is quite unnecessary, however, because the syntactic component must already generate  $x$  with its atomic parts as lexicon entries in order to provide the formal structure that bears the compositional meaning. Second, the phonological component operates on the syntactic structure of a sentence to assign it a phonetic shape.<sup>5-7</sup> Thus, in English the rules that assign stress patterns to sentences operate on the final derived phrase markers of these sentences. But in the case of idiomatic stretches like kicked the bucket both an occurrence with idiomatic meaning and one with compositional meaning have the

same stress pattern. Therefore all instances of both types of occurrence must have the same syntactic description. Kicked the bucket cannot have in one case the structure Intransitive Verb and in another the structure Verb + Noun Phrase. This latter treatment, if carried out consistently for all such cases, would lead to enormous complications of the phonological component.

Thus elementary considerations of grammatical simplicity suffice to show that at least some phrase idioms cannot be treated as lexical idioms, i. e., as members of one or another of the lowest level syntactic categories. Instead, at least the members of the class of idioms whose occurrences also have compositional meanings must receive the ordinary syntactic structure assigned to occurrences of the stretches with compositional meanings.

The previous considerations establish the fact that the semantic interpretation of sentences containing idiomatic stretches with compositional parallels will have to both account for the idiomatic meaning and mark the semantic ambiguity between this and the compositional meaning. A sentence such as

(1) the old lady kicked the bucket

must be assigned two readings, one of which attributes to it the meaning 'the old lady struck the bucket with her foot', while the other attributes to it the meaning 'the old lady died'. Moreover, the semantic interpretations of sentences containing phrase idioms will have to be properly related to the semantic interpretations of other sentences. Thus sentence (1) must be marked as a paraphrase (on a reading) of the old lady died; the sentence

(2) the man who has been dead for a week just kicked the bucket

must be marked as semantically anomalous on both of its readings, although for different reasons on each, etc.

## 2. Phrase Idioms

In order for the semantic component to handle phrase idioms properly, it is necessary to broaden the conception of the dictionary subpart of a semantic component presented in previous discussions of semantic theory.<sup>1-4</sup> In addition to having entries for unitary and compound lexical items, the semantic dictionary must also contain entries for the phrase idioms of the language, each such entry associating with its idiomatic stretch a set of readings to represent the sense of that stretch. The dictionary should now be thought of as having two parts, a lexical-item part and a phrase-idiom part. Entries in the latter part will have the form: first, a particular string of morphemes, the idiomatic stretch; next, some associated constituent that must dominate the idiomatic stretch in the phrase markers that are to be assigned the semantic information

(XVIII. LINGUISTICS)

associated with the pair of the idiomatic stretch and its dominating constituent, and, finally, this semantic information itself.

Access to the semantic information in the two subparts of the dictionary is obtained by different methods of assigning sets of readings associated with dictionary entries to the minimal semantic elements in the underlying phrase markers of sentences. In the case of minimal semantic elements that are lexical items, i. e., single morphemes or lexical idioms, the method of assigning readings is the following: Given such a minimal element e in an underlying phrase marker M, assign to e in M all and only the readings from e's dictionary entry which are compatible with the syntactic structure of e in M.<sup>8</sup> This method thus associates sets of readings with the lowest level or terminal elements of underlying phrase markers. In the case of minimal semantic elements that are not lexical items, i. e., phrase idioms, the method of assigning readings is the following: Given a string of morphemes t that is dominated by the constituent C in the underlying phrase marker M, assign to the set of readings associated with C in M those readings from the dictionary entry for t that have the form t → C → X, where X is the representation of the meaning of t provided by the dictionary of the semantic theory. This method thus assigns readings to higher level constituents in underlying phrase markers, not to terminal symbols. For example, in the case of the idiomatic stretch kicked the bucket, which has as part of its phrase-idiom dictionary entry the dominating constituent MV, the reading representing 'die' is associated not with any of the morphemes composing this stretch, but rather with the constituent MV that dominates occurrences of kick the bucket in underlying phrase markers. Note that it is this feature of assigning readings representing the meanings of phrase idioms directly to higher level constituents which is the aspect of the semantic theory's treatment of idioms which represents the fact that their meaning cannot be broken up into components and these parcelled out to the morphemes that make up the idiomatic stretch.

Of course, once readings have been supplied for all idioms in a sentence, the projection rules of semantic theory operate in the normal fashion, amalgamating readings drawn from sets of readings associated with constituents to form derived readings to be assigned to the constituent dominating them. A pair of readings, one or both of which is a reading for an idiom, amalgamate in exactly the same manner as readings that are not readings for idiomatic stretches.

Given this approach to the problems of idioms, linguistic theory provides an explanation of the difference between sentences whose meaning is wholly a compositional function of the meanings of its lexical items and sentences whose meanings are at least in part determined idiomatically (i. e., by phrase idioms). The meaning of a sentence is idiomatically determined if and only if its semantic interpretation is assigned partly on the basis of information obtained from the phrase-idiom part of the semantic dictionary. The phrase idioms of a language are just those expressions listed in the

phrase-idiom part of the dictionary of the semantic component of the optimum linguistic description of that language.

### 3. Further Justification for Our Approach

Our approach to phrase idioms is, of course, justified to a great extent by the considerations that originally motivated it, i. e., by the fact that for a large class of idioms, namely those with compositional parallels, it eliminates the otherwise required ad hoc additions and unnecessary complications to the syntactic and phonological components. But there are other more subtle justifications. These in turn can also serve to further justify some quite independently motivated features of the syntactic component, and by doing so further justify our approach to phrase idioms.

Consider the following sentences:

- (3) John kicked the bucket
- (4) the bucket was kicked by John

Clearly sentence (3) is semantically ambiguous. Thus, the semantic component must mark this ambiguity and assign to its terms the readings 'John died' and 'John struck the bucket with his foot'. We have explained the manner in which an extension of the dictionary component, in combination with the ordinary projection rule apparatus, can accomplish this. But the semantic theory must also account for the fact that sentence (4) is definitely unambiguous and is a paraphrase of sentence (3) on the latter reading but not on the former. Sentence (4) is, of course, 'the passive of' sentence (3). In earlier transformational descriptions of English it was assumed that sentences like (4) were derived by the action of the passive transformation on the identical phrase marker that underlies sentence (3). Corresponding actives and passives would thus always have identical underlying phrase markers. More recently, however, it has been suggested by Klima<sup>9</sup> that passive sentences have a different underlying phrase marker from active sentences, one that contains a Manner Adverbial constituent represented terminally by a passive morpheme. The passive transformation then substitutes by plus the subject NP of this underlying phrase marker for this passive morpheme, places the object NP where the subject originally was, and adds certain elements to the Auxiliary constituent. This treatment is dictated syntactically by formal considerations within the theory of transformational grammar, chiefly those having to do with the automatic assignment of derived constituent structure by simple, general, mechanical conditions. It also helps to account for certain selectional restrictions on passive constructions, in particular the fact that verbs that do not occur with Manner Adverbials do not have passive forms. The question naturally arises whether or not any external, in particular, any semantic, justification can be found for this way of treating the passive.

Consider again sentences (3) and (4). Taken together, the treatment of the passive

(XVIII. LINGUISTICS)

by means of a passive morpheme in underlying phrase markers and the conception of idioms presented above suffice to explain why sentence (3) but not sentence (4) is ambiguous. For note that the entry for kick the bucket in the phrase-idiom component of the semantic dictionary will be of the form: kick+the+bucket → MV → reading that represents the meaning 'die'. That is, in an underlying phrase marker in which the constituent MV dominates kick the bucket, the reading for 'die' is associated with MV. The underlying phrase marker for sentence (3) is

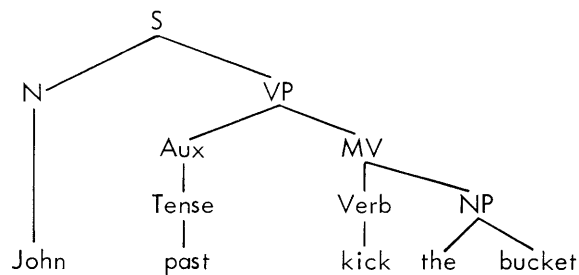


Diagram 1

This phrase marker is such that the conditions for access to semantic information in the phrase-idiom part of the semantic dictionary is met for the stretch kick the bucket, and hence the reading for 'die' is assigned to the constituent MV. But the underlying phrase marker for sentence (4), under the new treatment suggested by Klima, is

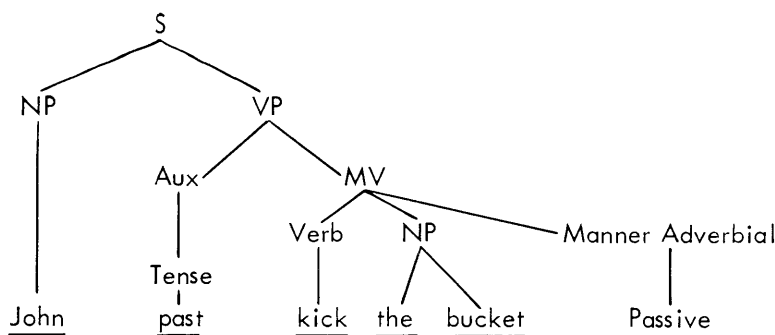


Diagram 2

But here the constituent MV does not dominate the string of morphemes kick+the+bucket as required by the syntactic part of the entry in the dictionary for the phrase idiom kick the bucket, but rather MV dominates kick+the+bucket+Passive. Hence the reading for 'die' cannot be assigned to MV in this phrase marker and thus the sentence that

Diagram 2 underlies, namely (4), is unambiguous although the active that has Diagram 1 as its underlying phrase marker is semantically ambiguous. We thus have the striking result that the treatment of the passive suggested by Klima and the present treatment of phrase idioms automatically explain the fact that the idiomatic meaning of kick the bucket is not found in the passive forms that correspond to this VP. This serves to justify both our treatment of idioms and Klima's syntactic description of the passive.<sup>10</sup>

Note, incidentally, that the lack of ambiguity of sentence (4) cannot be accounted for by claiming that idiomatic meanings are not carried over by transformations. Both sentences (5) and (6) are semantically ambiguous, having one sense that is due to a phrase idiom, and yet one is the result of the question transformation, the other of the imperative transformation:

(5) did John kick the bucket?

(6) kick the bucket!

#### 4. Syntactically Deviant Idioms

Our treatment of idioms leaves open the question of how a linguistic description is to handle idioms that are not syntactically well formed. Because the idiom kick the bucket is syntactically well formed, the syntactic component generates sentences containing occurrences of this idiom and the semantic component can assign them the readings found in their entries in the phrase-idiom part of the dictionary. But idioms that are not syntactically well formed, such as beat about the bush, will not appear as constituents of sentences that are generated by the syntactic component. If these idioms did appear in strings that are generated by the syntactic component, this component would not be empirically adequate because its output would contain some ungrammatical strings. Thus, there will be no occurrences of syntactically deviant idioms available for the normal process of semantic interpretation.

A suggestion of Chomsky offers a way of handling these idioms within the framework of the present paper.<sup>11</sup> He pointed out that sentences containing such idioms can be generated by the device that gives a syntactic description of the semisentences of the language.<sup>12</sup> This being so, if there is in the phrase-idiom part of the dictionary a section containing entries for these syntactically deviant phrase idioms, then these entries can be assigned to occurrences of such idioms in semisentences – so long as some provision is made for the semantic interpretation of semisentences.<sup>13</sup>

J. J. Katz, P. M. Postal

#### References

1. J. J. Katz and J. A. Fodor, The structure of a semantic theory, *Language* (in press).

(XVIII. LINGUISTICS)

2. J. J. Katz, Analyticity and contradiction in natural languages, Readings in the Philosophy of Language, edited by J. A. Fodor and J. J. Katz (Prentice-Hall, Inc., Englewood Cliffs, N. J., in press).
3. J. J. Katz, The nature of a semantic theory, Quarterly Progress Report No. 69, Research Laboratory of Electronics, M.I.T., April 15, 1963, pp. 181-193.
4. J. J. Katz and P. M. Postal, Towards an Integrated Theory of Linguistics Descriptions, Unpublished manuscript.
5. N. Chomsky, Explanatory models in linguistics, Logic, Methodology and Philosophy of Science: Proc. 1960 International Congress, edited by E. Nagel, P. Suppes, and A. Tarski (Stanford University Press, 1962), pp. 528-550.
6. N. Chomsky, The logical basis of linguistic theory, Proc. Ninth International Congress of Linguists (Mouton and Company, 's-Gravenhage, in press).
7. N. Chomsky and M. Halle, The Sound Pattern of English (Harper and Row, New York, in preparation).
8. The manner in which the notion 'compatible with' is to be explicated in syntactic terms requires an extensive statement of the form of the syntactic component which cannot be given here.
9. E. S. Klima, Unpublished work, 1960-1961.
10. The present treatment of idioms like kick the bucket provides considerable support for the treatment of the tense suffixes of the verb as part of the Auxiliary constituent in underlying P-markers. Clearly simplicity considerations require not separate phrase-idiom entries for kicked the bucket, kicking the bucket, kicks the bucket, etc., but rather only a single entry for kick the bucket. But this is only possible if MV does not dominate the tense morphemes in underlying phrase markers.
11. N. Chomsky, Private Communication, 1963.
12. N. Chomsky, Some methodological remarks on generative grammar, WORD 17, 219-239 (1961).
13. J. J. Katz, Semi-sentences, Readings in the Philosophy of Language, op. cit.

B. STRESS AND PITCH IN THE SERBO-CROATIAN VERB

The stressed syllable in a Serbo-Croatian word can have either rising or falling accentuation. The occurrence of rising and falling accentuation is restricted as follows:

- (a) Falling accentuation is possible only on the first syllable of a word; and
- (b) Rising accentuation is possible only in words of two or more syllables and can occur on any syllable except the last (thus, in particular, monosyllables can have only falling accentuation).

Suppose that a syllable is called high-pitched if it starts on a high pitch and low-pitched if it starts on a low pitch. Then every word has exactly one high-pitched syllable: if the word has falling accentuation, then the stressed syllable (i. e., the first syllable) is high-pitched, and if it has rising accentuation, then the syllable following the stressed syllable is high-pitched. Thus the location and type (rising or falling) of stress can be predicted from the location of high pitch. Moreover, high pitch can occur on any syllable of the word.<sup>1</sup>



Serbo-Croatian verbs are of two types: thematic and athematic. Thematic verbs are those that have a verbalizing suffix, for example, *igrati* = *igr* + *aj* + *ti*, where /*aj*/ is the verbalizing suffix. Athematic verbs are those that have no verbalizing suffix, for example, *trēsti* = *trēs* + *ti*. I assume that the infinitive and present tense forms of the thematic verbs have the following immediate constituent structure:

infinitive:	((stem +	verbalizing	) +	infinitive			
	suffix	ending	)				
example:	<i>igrati</i> =	(( <i>igr</i> +	<i>aj</i> )	+ <i>ti</i> )			
		present tense:	((stem +	verbalizing	+ present	) +	personal
		suffix	theme	ending	)		
example:	<i>igrām</i> =	(( <i>igr</i> +	<i>aj</i>	+ <i>ā</i> )	+ <i>m</i> )		

There are two types of thematic verbs: those that have the same stress in the infinitive and all forms of the present tense, and those that undergo what Schooneveld<sup>2</sup> calls the "cardinal intonational alternation." Schooneveld defines the cardinal alternation as an alternation between a "rising non-initial penultimate syllable in the infinitive and a rising accent on the preceding syllable in the truncated forms of the present or (if the penultimate syllable is at the same time the initial syllable), between a rising initial-penultimate syllable and a falling initial-penultimate syllable". Examples<sup>3</sup>:

Infinitive	<i>venčāvati</i>	<i>igrati</i>
1st sg.	<i>venčāvām</i>	<i>igrām</i>

Schooneveld speaks as if there were two separate alternations going on here: one a shift of rising intonation, the other an alternation between rising and falling intonation. However, if high pitch rather than stress is marked in the examples above, it will be noted that in both cases all that happens is that the high pitch is shifted one syllable to the left:

<i>venčāvati</i>	<i>igrati</i>
<i>venčāvām</i>	<i>igrām</i>

(the raised dot denotes high pitch).

The infinitive and present tense forms of a typical verb that undergoes the cardinal alternation are

	phonetic form	underlying form
infinitive	<i>igrati</i> = <i>igrati</i>	(( <i>igr</i> + <i>aj</i> ) + <i>ti</i> )
1st sg.	<i>igrām</i> = <i>igrām</i>	(( <i>igr</i> + <i>aj</i> + <i>a</i> ) + <i>m</i> )
2nd sg.	<i>igrās</i> = <i>igrās</i>	(( <i>igr</i> + <i>aj</i> + <i>a</i> ) + <i>ś</i> )
3rd sg.	<i>igrā</i> = <i>igrā</i>	(( <i>igr</i> + <i>aj</i> + <i>ā</i> ) + <i>φ</i> )
1st pl.	<i>igrāmo</i> = <i>igrāmo</i>	(( <i>igr</i> + <i>aj</i> + <i>a</i> ) + <i>mo</i> )
2nd pl.	<i>igrāte</i> = <i>igrāte</i>	(( <i>igr</i> + <i>aj</i> + <i>a</i> ) + <i>te</i> )
3rd pl.	<i>igrājū</i> = <i>igrājū</i>	(( <i>igr</i> + <i>aj</i> + <i>ā</i> ) + <i>ū</i> )

## (XVIII. LINGUISTICS)

The verbalizing suffix /aj/ is deleted in all present tense forms except the 3rd pl. by rules that will be formulated later; note that in these forms the stem vowel is high-pitched. In the infinitive and the 3rd plural, in which the verbalizing suffix is not deleted, the verbalizing suffix is high-pitched.<sup>4</sup> These facts can be accounted for by saying that the underlying forms of these words have a high-pitched verbalizing suffix and that if the verbalizing suffix is deleted, the high pitch is shifted from it onto the preceding syllable. The appropriate rules for deleting the verbalizing suffix in the five present tense forms in which it is dropped and for deleting the theme vowel  $\bar{a}$  in the 3rd person plural seem to be a rule that deletes /j/ between like vowels followed by either a consonant or a word boundary, and a rule that deletes the first of a sequence of two vowels. I thus formulate the rules:

- $$3. \quad j \rightarrow \phi \quad \text{in env} \quad \left[ \begin{array}{l} + \text{ voc} \\ - \text{ cns} \\ \alpha \text{ grv} \\ \beta \text{ flt} \\ \gamma \text{ dif} \end{array} \right] \text{ ————— } + \left[ \begin{array}{l} + \text{ voc} \\ - \text{ cns} \\ \alpha \text{ grv} \\ \beta \text{ flt} \\ \gamma \text{ dif} \end{array} \right] \left\{ \begin{array}{l} C \\ \# \end{array} \right\}$$
- $$4. \quad V \rightarrow \dot{V} \quad \text{in env} \quad \text{ ————— } C_o \dot{V} V$$
- $$5. \quad V \rightarrow \phi \quad \text{in env} \quad \text{ ————— } V$$

These rules will be assumed to be part of a cycle, i. e., the rules are to be applied first to the innermost immediate constituents, then to the next innermost constituents, etc.

These rules will correctly generate the relevant forms of thematic verbs with cardinal alternation. How then should thematic verbs with fixed stress be handled? For verbs such as  $\check{v}ideti/\check{v}id\bar{m}$ , the solution is obvious: the /i/ of the stem is marked high-pitched. Since rule 4 can only affect verbs with a high-pitched verbalizing suffix, the high pitch would remain on the /i/ throughout the entire paradigm.

However, there are also verbs such as  $\check{c}itati/\check{c}it\bar{m}$  represented in my system as  $\check{c}it\grave{a}ti/\check{c}it\grave{a}\bar{m}$ . If the underlying form for  $\check{c}it\grave{a}ti$  were marked with a high pitch on the /a/, the incorrect form  $*\check{c}it\bar{m} = *\check{c}it\bar{a}\bar{m}$  would be obtained, since rule 4 would automatically shift the high pitch onto the first syllable. The solution that I propose is to treat the underlying forms for the paradigm of  $\check{c}it\grave{a}ti$  as having no high-pitched syllables whatever and having a rule that marks the last syllable as high-pitched if none of the preceding syllables are high-pitched:

- $$7. \quad V \rightarrow \dot{V} \quad \text{in env} \quad (X \text{ ————— } ), \text{ where } X \text{ contains no } \dot{V}.^5$$

For the rule to yield the correct forms  $\check{c}it\grave{a}ti$  and  $\check{c}it\bar{m}$ , it will be necessary to have the rule apply as part of the cycle. If the rule were not in the cycle, it of course would give the incorrect form  $*\check{c}it\grave{a}ti = *citati$ . However, if the rule is put in the cycle, it operates as follows:

		infinitive		
		((č̇it + aj) + ti)	((č̇it + aj + ā) + m)	((č̇it + aj + ā) + ū)
first pass through cycle	$\left\{ \begin{array}{l} \text{rules 3-5} \\ \text{rule 7} \end{array} \right.$	no effect	no effect	no effect
		č̇it + āj	č̇it + aj + ā̇	č̇it + aj + ā̇
		(č̇it + āj + ti)	(č̇it + aj + ā̇ + m)	(č̇it + aj + ā̇ + ū)
second pass through cycle	$\left\{ \begin{array}{l} \text{rule 3} \\ \text{rule 4} \\ \text{rule 5} \\ \text{rule 7} \end{array} \right.$	no effect	č̇it + a + ā̇ + m	no effect
		no effect	no effect	č̇it + āj + ā̇ + ū
		no effect	č̇it + ā̇ + m	č̇it + āj + ū
		no effect	no effect	no effect

and thus gives high pitch on the correct syllables.

An alternation between plain and palatalized or palatal consonants, known as transitive softening or jotovanje, takes place in Serbo-Croatian. I am at the moment unprepared to state the exact environment in which it takes place, other than to state the general shape of the rule: the change occurs when a consonant is followed by two vowels that satisfy some condition that I see no way of stating other than to list the combinations of vowels before which the change occurs. It occurs, for example, in the present tense of brisati: ((bris + a + ē) + m) → briṧēm. In any event, the simplest way to treat the phenomenon probably will be to say that a /j/ is inserted before the appropriate combination of vowels and that by later rules kj → č̇, tj → č̇, sj → ṧ, etc. Stating the rule in terms of the insertion of a /j/ allows an elegant treatment of the ova/uj alternation. Suppose that /v/ and /j/ are represented in the dictionary forms by /u/ and /i/ (which can be done, since the former segments occur only in postvocalic and initial prevocalic position). Thus the rules will have to contain somewhere a rule that states that high vowels become glides in postvocalic or initial prevocalic position. Suppose that kupovati is represented as ((kup + ou + a) + ti). The rule just mentioned will convert the /u/ into /v/ (actually, into /w/, which later becomes /v/), yielding kupovati. The present tense form kupujem has the expected underlying form ((kup + ou + a + ē) + m). Since a + ē is one of the vowel combinations before which transitive softening takes place, the latter form would be converted into ((kup + ouj + a + ē) + m). If this all precedes rule 5, that rule will delete vowels followed by vowels and yield ((kup + uj + ē) + m), i. e., the correct form.

The transitive softening rule would have to precede the rule for converting /u, i/ into /v, j/, since otherwise it would apply to the /ou/ in the present tense form and yield the incorrect form \*kupovjēm → \*kupovljēm. I shall assume that these two rules are part of the cycle. Thus rules 3-5 will be preceded by the rules

1. /j/ is inserted in env \_\_\_\_\_ + V + V, subject to some condition on the V's

## (XVIII. LINGUISTICS)

$$2. \left[ \begin{array}{l} - \text{cns} \\ + \text{dif} \end{array} \right] \rightarrow [- \text{voc}] \quad \text{in env} \quad \left\{ \begin{array}{l} \text{V} \\ \# \text{ \rule{1cm}{0.4pt}} \text{V} \end{array} \right\}$$

I maintain that the rules given thus far are sufficient to account for the stress shifts in all six classes of thematic verbs given by Meillet and Vaillant.<sup>6</sup>

There are several verbs that Meillet does not include in his six types but that behave like those treated above. Consider, for example, the verbs

Infinitive	m <sub>l</sub> ɛti	br <sub>ɔ</sub> ti	kl <sub>ɔ</sub> ti
1st sg.	mɛljēm	berēm	koljēm

Suppose that base forms mel + é, ber + á, and kol + á are assumed. The 1st sg. will then have the underlying forms ((mel + é + ē) + m), ((ber + á + ē) + m), and ((kol + á + ē) + m). In the first pass through the cycle, transitive softening will occur, the high pitch will be shifted one syllable to the left, and the verbalizing suffix will be eliminated, yielding mɛlj + ē + m, berj + ē + m, and kōlj + ē + m. On the second pass through the cycle, none of the rules apply. These are the correct forms for mleti and klati, and in the case of brati, the correct form is obtained by a later rule by which rj → r (/rj/ never occurs in Serbo-Croatian, having coalesced with /r/). The infinitive will have the underlying forms ((mel + é) + ti), ((ber + á) + ti), and ((kol + á) + ti). None of the rules in the cycle apply, so that the forms meléti, koláti, and beráti, i. e., \*meleti, \*berati, and \*kolati, would be obtained. However, if the grammar were provided with a rule that deleted low-pitched /e/ and /o/ in CVL stems, what would remain after the application of that rule is ml + é + ti, br + á + ti, and kl + á + ti, i. e., the correct forms. The incorporation of such a rule into the grammar appears to do no harm, since, as far as I can determine, all C<sub>o</sub><sup>e</sup>L stems work like this.

Consider now verbs with prefixes. In the thematic verbs of the six types that Meillet recognizes, a prefix does not affect the location of high pitch:

stress notation		high-pitch notation	
gr <sub>ɔ</sub> āditi	gr <sub>ɔ</sub> ādīm	gr <sub>ɔ</sub> āditi	gr <sub>ɔ</sub> ādīm
razgr <sub>ɔ</sub> āditi	razgr <sub>ɔ</sub> ādīm	razgr <sub>ɔ</sub> āditi	razgr <sub>ɔ</sub> ādīm

The stressed prefix in the present tense of these verbs is thus not a case of stress shift (as it is traditionally described), since the high pitch is on the same syllable regardless of whether or not there is a prefix.

However, the prefix does become high-pitched in the following verb:

zr <sub>ɛ</sub> ti	zrēm	(=	zr <sub>ɛ</sub> ti	zrēm)
obazreti	obazrēm	(=obazr	ɛti	obazrēm)

I will disregard for the moment what happens to the stress and concentrate on

the segmental phonemes. The obvious underlying form for *zreti/zrēm* is *zr + ej* (i. e., a nonsyllabic stem). The present tense form *zrēm* arises from  $((zr + ej + \bar{e}) + m)$  by the truncation rules 3 and 5. However, it will be noted that this solution automatically produces the correct stress as well: given the underlying form  $((oba + zr + \acute{e}j + \bar{e}) + m)$ , rule 4 would shift the high pitch from the /e/ of the stem onto the /a/ of the prefix, and the correct form  $(ob\acute{a} + zr + \bar{e} + m)$  would be obtained.

There are a large number of other verbs besides *zreti* in which the prefix becomes high-pitched in the present tense, for example,

Infinitive	dōbiti	umrēti	prosūti	naḍūti	popēti
	dōbijēm	umrēm	prospēm	naḍmēm	popnēm

A further peculiarity of these verbs is that in the prefixless verbs (or at least, those that exist: many of the stems in question only occur with prefixes) the high pitch in the infinitive is on a different syllable than when there is a prefix (recall that with all verbs treated thus far, even *zreti*, the high pitch has always been on the same syllable in the infinitive regardless of whether or not there was a prefix). Example: *mrēti* but *umrēti* (i. e., *mrēti* but *umrēti*).

Suppose that the stems *mr*, *sp*, *dm*, *pn*, etc., are regarded as ending in a high-pitched segment (which in this case will be a consonant) and that the stress shift rule is modified so as only to require a high pitch (not necessarily a high-pitched vowel) in its environment:

4'.  $V - \acute{V}$  in env \_\_\_\_\_  $C_o[+ \text{high pitch}] + V$ .

Then whenever one of the stems *mr*, *sp*, *dm*, *pn* is followed by a vocalic ending, rule 4' will move the high pitch onto the prefix if there is one. This can also be made to work for *biti* by representing it with a high-pitched /j/ if rule 4' is modified still further so as to allow an optional vowel between the  $C_o$  and the high-pitched segment. However, the rule still would work correctly after the modification, since verbs of the *biti* type are the only words in which a vowel would ever be followed immediately by a high-pitched segment.

To generate the infinitives of verbs with consonantal stems, a rule will be needed which inserts a vowel before the infinitive ending and (in the case of stems ending in a true consonant) deletes the stem-final consonant (however, stem-final liquids are retained). To obtain the correct stress, it is necessary that this rule be in the cycle, specifically, between rules 5 and 7, and that rule 7 be modified so that it will not apply to the prefixless infinitives on the first cycle. The obvious modification is to require the constituent to which rule 7 is applying to consist of at least two morphemes, i. e., to say that a vowel becomes stressed in the environment  $(X + Y \text{ \_\_\_\_\_\_ })$ , where X and Y contain no  $\acute{V}$ . The infinitives *mrēti* and *umrēti* are generated as follows (using 6

## (XVIII. LINGUISTICS)

to denote the rule for modifying nonsyllabic stems):

			$((m\acute{r}) + ti)$	$((u + m\acute{r}) + ti)$
first pass	1-5	no effect	no effect	no effect
through	6	$m\acute{r}\bar{e}$	$u + m\acute{r}\bar{e}$	
cycle	7	no effect	$u + m\acute{r}\bar{e}$	
			$(m\acute{r}\bar{e} + ti)$	$(u + m\acute{r}\bar{e} + ti)$
second				
pass	1-6	no effect	no effect	no effect
through	7	$m\acute{r}\bar{e} + t\acute{i}$	no effect	
cycle				

Final result:  $m\acute{r}\bar{e}t\acute{i}$ ,  $um\acute{r}\bar{e}t\acute{i}$ .

Athematic verbs can be either nonsyllabic or monosyllabic. The rules formulated thus far generate correctly the relevant forms of all nonsyllabic athematic verbs. Let me now turn to the monosyllabic ones. Monosyllabic athematic verbs fall into three classes:

1) those that have rising intonation in the infinitive and all present tense forms:  $tr\bar{e}sti/tr\bar{e}s\bar{e}m$ .

2) those that have a short falling pitch in the infinitive but a long rising pitch in the present tense:  $gr\bar{i}sti/gr\bar{i}z\bar{e}m$ , and

3) three verbs  $le\acute{c}i$ ,  $re\acute{c}i$ , and  $mo\acute{c}i$ , which have a rising pitch in the infinitive but a falling pitch in the present tense.

I see no alternative to simply treating type (3) as exceptions.

The obvious solution for type (1) is to represent the stems as having no high pitch. The rule 7 would put the high pitch on the  $-ti$  of the infinitive and the theme vowel  $\bar{e}$  of the present tense. This gives the correct answer for all forms except one, namely the 3rd person plural. The expected underlying form  $((tr\bar{e}s + \bar{e}) + \bar{u})$  for the third person plural  $tr\bar{e}s\bar{u}$  of  $tr\bar{e}sti$  would come out of the first pass through the cycle with a high pitch on the  $\bar{e}$ . On the second pass, rule 4 would shift the high pitch to the preceding syllable, rule 5 would delete the  $\bar{e}$ , and the incorrect form  $*tr\bar{e}s\bar{u}$  would be obtained. The only solutions that I can think of to this difficulty are all somewhat ad hoc, and all are essentially equivalent to saying that the present tense does not have the IC structure  $((stem + theme) + person)$  but rather  $(stem + theme + person)$ . If such a structure is assumed, then the relevant forms will go only once through the cycle, and rule 7 will put the high pitch on the final vowel; in particular, in the 3rd person plural, it will put the high pitch on the  $\bar{u}$ . However, I am unable to find any independent justification for assuming that athematic verbs have a different IC structure from thematic verbs.

Verbs of group (2) will undoubtedly require some special rule, since I know of no

process of vowel shortening or lengthening or stress shift of which they could be considered a special case (note that in verbs of group (2) the stress shift is in the opposite direction from that in cardinally alternating thematic verbs). The simplest ad hoc rule that I can think of for this alternation is a rule that makes a high-pitched vowel long and low-pitched in a monosyllabic stem followed by a vowel. That would involve representing type (2) stems as having a short high-pitched vowel, which is possible, since there are no other monosyllabic athematic verbs which there would be any reason to represent with a high-pitched vowel.

The order of application of the rules arrived at above is the following:

1. Insert /j/ in env \_\_\_\_\_ + V + V, subject to some condition on the V's.
2.  $\begin{bmatrix} - \text{cns} \\ + \text{dif} \end{bmatrix} \rightarrow [- \text{vos}]$  in env  $\left\{ \begin{array}{l} \text{V} \text{_____} \\ \# \text{_____} \end{array} \text{V} \right\}$
3. /j/  $\rightarrow \phi$  in env  $\begin{bmatrix} + \text{voc} \\ - \text{cns} \\ a \text{ grv} \\ \beta \text{ flt} \\ \gamma \text{ dif} \end{bmatrix}$  \_\_\_\_\_  $\begin{bmatrix} + \text{voc} \\ - \text{cns} \\ a \text{ grv} \\ \beta \text{ flt} \\ \gamma \text{ dif} \end{bmatrix}$   $\left\{ \begin{array}{l} \text{C} \\ \# \end{array} \right\}$
4. V  $\rightarrow \dot{V}$  in env \_\_\_\_\_ C<sub>o</sub>(V) [+ high pitch] + V
5. V  $\rightarrow \phi$  in env \_\_\_\_\_ V
6. Modification of nonsyllabic stems before the infinitive ending (an  $\bar{u}$  or  $\bar{e}$  is added after the stem; if the stem ends in a true consonant, the latter is deleted).
7. V  $\rightarrow \dot{V}$  in env (X + Y \_\_\_\_\_), where X and Y contain no  $\dot{V}$ .

J. D. McCawley

#### References

1. The possibility of representing stress in Serbo-Croatian by marking the occurrence of high pitch was discovered independently by myself and E. Wayles Browne of Harvard University
2. C. H. van Schooneveld, Serbo-Croatian conjugation, *International J. Slavic Linguistics and Poetics* 1, 55-69 (1957).
3. Instead of the traditional notation, I will use the more suggestive notation of marking rising and falling intonation by / and \, respectively, beneath the vowel. My notation and the traditional notation are related as follows:

	my notation	traditional notation
short rising	a /	à /
long rising	ā /	ā /
short falling	a \	\a
long falling	ā \	^a

The notation used here is a variant of that used in H. G. Lunt, *On the study of Slavic accentuation*, *WORD* 19, 82-99 (1963).

(XVIII. LINGUISTICS)

4. It is interesting to note that in Russian the basic stress moving rule calls for stress to be shifted when a stressed vowel is deleted, much like this rule. However, in Russian the direction of shift is the opposite, namely to the right rather than to the left.

5. The parentheses denote the beginning and end of the constituent to which the rule is applying.

6. A. Meillet and A. Vaillant, Grammaire de la Langue Serbo-Croate (Champion, Paris, 2d edition, 1952).

C. SOME LANGUAGES THAT ARE NOT CONTEXT-FREE

In many frameworks for describing the grammar of natural languages it has been assumed, implicitly or explicitly, that the grammar will assume the form of what Chomsky<sup>1-3</sup> has termed "context-free grammar." We shall consider a subpart of English (and many other natural languages) which is demonstrably not context-free.

Consider the English cardinal numbers. There are an infinite number of them, and each is of the form

$$N T^n (, N T^{n-1}) \dots (, N T) (, N) \quad (1)$$

where the comma indicates a comma intonation, the parentheses indicate that the enclosed elements are optional,  $T^n$  abbreviates  $n$  occurrences of thousand, and  $N$  indicates a number between 1 and 999. Here, million is interpreted as thousand thousand, billion as thousand thousand thousand, etc. A closely related analysis of the cardinal numbers treats each of them in a full form

$$N T^n, N T^{n-1} \dots, N T, N \quad (2)$$

where  $N$  now indicates a number between 0 and 999. The expression of cardinal numbers in terms of powers of 10 is also of this form, with the comma indicating a comma intonation followed by plus,  $T$  abbreviating times ten, and  $N$  indicating a number between 0 and 9.

The set of cardinal numbers is of particular interest in a discussion of the form of grammar, since this set is the clearest example of an infinite subpart of natural languages. In the theory of transformational grammars, for example, there is some question as to whether the phrase-structure component of the grammar should generate a finite or infinite set of strings. It is interesting to note that there are quite natural ways of generating the cardinal numbers by means of transformations – in fact, there is a fairly simple set of transformations that generate the cardinal numbers in order.

To investigate whether the set of all English cardinals is CF (context-free) or not, we consider the following sets of strings:

$$P = \{x \mid x = b^n(ab^{n-1}) \dots (ab^2)(ab), \text{ where } n = 1, 2, 3, \dots\}$$



$$C = \{x \mid x = b^n ab^{n-1} a \dots ab^2 ab, \quad \text{where } n = 1, 2, 3, \dots\}$$

It is clear that  $P$  is CF if and only if the set of all strings of form (1) is CF, and that  $C$  is CF if and only if the set of all strings of form (2) is CF. Along with  $P$ , we consider the sets  $P_i$

$$P_i = \left\{ x \mid x = b^{n_i} ab^{n_i-1} a \dots ab^{n_0}, \quad \text{where } 1 \leq n_0 < n_1 < \dots < n_i \right. \\ \left. \text{and } n_0 = 1, 2, 3, \dots \right\}, \quad i = 0, 1, 2, \dots$$

Note that  $P = \bigcup_{i \geq 0} P_i$ . Along with  $C$ , we consider the sets  $C_i$ .

$$C_i = \{x \mid x = b^{n+i} ab^{n+i-1} a \dots ab^n, \quad \text{where } n = 1, 2, 3, \dots\}, \\ i = 0, 1, 2, \dots$$

It is easy to show that  $C_0 (= P_0)$ ,  $C_1$ , and  $P_1$  are CF.

We assume that any CF grammar  $G$  is such that each symbol in the vocabulary of  $G$  occurs in some  $S$ -derivation and is such that any  $S$ -derivation can be terminated by a finite number of applications of the rules of  $G$ . It can be shown that these two well-formedness conditions have no effect upon the set of all CF languages. Moreover, it is easy to show that any infinite CF language has a subset of the form  $F$ :

$$zx^m ty^m w \quad \text{for } m = 1, 2, 3, \dots$$

where  $z$ ,  $x$ ,  $t$ ,  $y$ , and  $w$  are fixed strings,  $t$  is non-null, and either  $x$  or  $y$  is non-null.

**THEOREM 1:**  $C_i$  ( $i \geq 2$ ) and  $C$  are not CF.

**PROOF 1:** Suppose that  $C_i$  is CF. Then  $C_i$  has a subset of form  $F$ . That is,

$$zx^m ty^m w = b^{n_m+i} a \dots ab^{n_m}$$

so that  $z = b^u$  and  $w = b^v$ , where  $u$  and  $v$  are fixed and greater than or equal to zero, and so that there must be an  $a$  in either  $x$ ,  $t$  or  $y$ . If there is an  $a$  in  $x$  or  $y$ , then there are at least  $m$   $a$ 's in  $zx^m ty^m w$ . But there are only  $i$   $a$ 's in  $b^{n_m+i} a \dots ab^{n_m}$ . By choosing  $m = i$ , we obtain a contradiction.

Hence, all of the  $a$ 's must occur in  $t$ ; that is,

$$t = ab^{n_m+i-1} a \dots ab^{n_m+i} a \quad \text{for } m = 1, 2, 3, \dots$$

For  $i \geq 2$ , as  $m$  increases without limit,  $n_m$  and hence the length of  $t$  also increase without limit. But  $t$  must be fixed. From this contradiction, we see that  $C_i$  ( $i \geq 2$ ) is not CF.

The proof that  $C$  is not CF is similar; again,  $x$  and  $y$  must be free of  $a$ , so that

$$t = ab^{n-1} a \dots ab^2 a \quad \text{for } m = 1, 2, 3, \dots$$

which is impossible, since  $t$  must be fixed.

The method of Proof 1 cannot succeed for  $P_i$  and  $P$  because  $P_i$  and  $P$  do have infinite subsets of form  $F$ ; for example, both  $P_0$  and  $P_1$  are subsets of  $P$ , and

$$\{x \mid x = b^m ab^2 ab, \quad \text{where } m = 3, 4, 5, \dots\}$$

is a subset of  $P_2$ . In fact, each  $P_i$  ( $i \geq 1$ ) can be expressed as the union of an infinite number of disjoint infinite sets, each of which is of form  $F$ . Then, since  $P$  is the union of the  $P_i$ 's, and the  $P_i$ 's are disjoint,  $P$  can be similarly expressed.

In the following proof we assume that for each nonterminal symbol  $A$  in a CF grammar  $G$ , there are infinitely many strings (hence, infinitely many terminal strings) that are derivable from  $A$ . Otherwise, whenever there is a rule  $B \rightarrow \phi_1 A \phi_2$  and only  $z_1, \dots, z_n$  are derivable from  $A$ , then  $A$  can be eliminated from the grammar by replacing the rule  $B \rightarrow \phi_1 A \phi_2$  by the rules  $B \rightarrow \phi_1 z_i \phi_2$  for  $1 \leq i \leq n$ .

Suppose that there is a CF grammar  $G$  for which  $L(G)$  is either  $P$  or  $P_i$  ( $i \geq 2$ ). We make the following definitions:  $V_1$  is the set containing  $b$  and every nonterminal symbol  $A$  for which every terminal derivative (i. e., every derivative consisting entirely of terminal symbols) of  $A$  consists entirely of  $b$ 's;  $T$  is the set of all strings of symbols in  $V_1$ ;  $V_2$  is the set consisting of every nonterminal symbol having at least one terminal derivative containing at least one  $a$ ;  $V_3$  ( $\subseteq V_2$ ) is the set consisting of every nonterminal symbol having an infinite number of terminal derivatives containing two or more  $a$ 's.

(a) In any  $S$ -derivative in  $G$ , no nonterminal symbol appears to the right of the symbol  $a$  or to the right of a symbol in  $V_2$ .

If there were such a nonterminal symbol, then there would be terminal  $S$ -derivatives not in  $P$  or  $P_i$ . Then

(b) If  $A \rightarrow \phi$  in  $G$ , then  $\phi$  is of the form  $\alpha Bx$ , where  $\alpha$  is null or  $\alpha \in T$ ,  $B$  is null or  $B \in V_2$ , and  $x$  is null or terminal.

(c) There is a rule of the form  $S \rightarrow \alpha_1 B_1 x_1$  in  $G$ , where  $\alpha_1$  and  $x_1$  have the same conditions as  $\alpha$  and  $x$ , respectively, in (b), and  $B_1 \in V_3$ .

PROOF OF (c): Consider all rules  $S \rightarrow \phi$  in  $G$ . By (b),  $\phi$  is of the form  $\alpha Bx$ . Suppose that in every such rule,  $B$  is either null or has only a finite number of terminal derivatives containing two or more  $a$ 's. Therefore, the terminal derivatives from each  $\alpha Bx$  are all of the form  $b^k y$  or  $b^{k_1} a b^{k_2} y$  (for some  $k, k_1, k_2 \geq 0$ ), where  $y$  is one of a finite number of fixed terminal strings. But then  $L(G)$  is properly included in  $P$  or  $P_i$  ( $i \geq 2$ ). Then (c) follows from this contradiction.

Consider all of the rules  $B_1 \rightarrow \phi$  in  $G$ . Again by (b),  $\phi$  is of the form  $\alpha Bx$ . Suppose that in each such rule  $B$  is null or has only a finite number of terminal derivatives

containing two or more a's. We can construct a CF grammar  $G_1$  in the following way: Replace the rules  $B_1 \rightarrow aBx$  in  $G$  by the rules  $S \rightarrow a_1 aBxx_1$  and the rules  $\overline{B}_1 \rightarrow aBx$ , where  $\overline{B}_1$  is a new symbol, and replace any occurrence of  $B_1$  in the remaining rules by  $\overline{B}_1$ . Then  $L(G_1) = L(G)$ , but there is no rule satisfying the conditions in (c). From the contradiction, there must be a rule of the form  $B_1 \rightarrow a_2 B_2 x_2$  in  $G$ , where  $B_2 \in V_3$ .

We can treat  $B_2$  as we have treated  $B_1$ , so as to obtain a  $B_3 \in V_3$ , etc. There are, however, only a finite number of vocabulary symbols in  $G$ , so that either  $S \Rightarrow \beta S y \Rightarrow \beta^n S y^n$  for some  $\beta$  (null or in  $T$ ), for some  $y$  (null or terminal), and for all  $n \geq 2$ ; or else  $S \Rightarrow \beta B_m y \Rightarrow \beta \gamma B_m z y \Rightarrow \beta \gamma^n B_m z^n y$ , with appropriate conditions on  $\beta$ ,  $\gamma$ ,  $z$ ,  $y$ , and  $n$ . Then  $x$  cannot be null, for if it were  $L(G)$  would be properly included in  $P$  or  $P_i$  again. Nor can  $x = a$ , since there is no string in  $P$  or  $P_i$  with two adjacent a's. Hence  $x$  contains at least one b. Now  $S$  and  $B_m$  must both be in  $V_3$ , so that they have terminal derivatives containing at least two a's. But then there must be strings in  $L(G)$  which are not in  $P$  or  $P_i$  ( $i \geq 2$ ). That is, any CF grammar either is insufficient to generate  $P$  or  $P_i$ , or else generates strings not in  $P$  or  $P_i$ . Therefore

THEOREM 2:  $P$  and  $P_i$  ( $i \geq 2$ ) are not CF.

A. M. Zwicky, Jr.

#### References

1. N. Chomsky, Context-free grammars and pushdown storage, Quarterly Progress Report No. 65, Research Laboratory of Electronics, M.I.T., April 15, 1962, pp. 187-194.
2. N. Chomsky, Formal properties of grammars, Handbook of Mathematical Psychology, Vol. 2, edited by R. R. Bush, E. H. Galanter, and R. D. Luce (John Wiley and Sons, Inc., New York, in press).
3. N. Chomsky, On certain formal properties of grammars, Information and Control 2, 137-167 (1959).

#### D. NASAL DIPHTHONGS IN RUSSIAN

We define a nasal diphthong as any  $\begin{bmatrix} +\text{voc} \\ -\text{cns} \end{bmatrix} \begin{bmatrix} +\text{cns} \\ +\text{ns1} \end{bmatrix}$  cluster in the environment  $\text{_____} + \left\{ \begin{matrix} C \\ \# \end{matrix} \right\}$ . The existence of nasal diphthongs in Russian is

## (XVIII. LINGUISTICS)

shown in such forms as the following:

<u>Phonetic</u>	<u>Phonemic</u> <sup>1</sup>	<u>English</u>
žát,	(gím+tī) <sup>2</sup>	to press, squeeze
žm, ót	((gim+é)+tu)	he presses, squeezes
žát,	(gín+tī) <sup>2</sup>	to reap
žn, ót	((gin+é)+tu)	he reaps
nač, áť,	(na+=+kín+tī) <sup>2</sup>	to begin
nač, n, óť	((na+=+kin+é)+tu)	he begins
znám, ə	(znóm+en+∅)	banner (nom. sg.)
znám, in, i	(znóm+en+ī)	banner (gen. sg.)
znam, óně	(znóm+én+ō)	banner (nom. pl.)

Forms such as those listed above are sufficient to prove the existence of morpho-phonemic front nasal diphthongs in Russian.<sup>3</sup>

It is commonly considered that in contemporary Russian there are no back nasal diphthongs (except, perhaps, for the often quoted but entirely inconclusive forms [zvúk] 'sound' and [zvón] 'peal, ringing, chime'). In the Third Person Plural of verbs, however, we find a perfectly clear case for the existence not only of front nasal diphthongs but of back nasal diphthongs as well.

If we specify the Third Plural morpheme as

$$3 + \text{Pl} \rightarrow \underline{n} + \underline{tu}$$

then the phonemic transcription of some representative Third Plural verb forms will be as follows:

<u>Phonetic</u>	<u>Phonemic</u>	<u>English</u>
s, id, áť	((sīd+ē+ī)+n+tu)	they sit
gəvar, áť	((gouor+ī+ī)+n+tu)	they talk
n, isút	((nes+é)+n+tu)	they carry
znájut	((znóī+e)+n+tu)	they know
žmút	((gim+é)+n+tu)	they press, squeeze

In order to derive the phonetic transcriptions from the phonemic representations

we require the following rules (note that the rules given below represent simply an extension and a slight reordering of the rules given in [RMCR]; we have omitted all [RMCR] rules that play no part in the present report):

$$\text{C-1} \quad \begin{bmatrix} -\text{cns} \\ +\text{diff} \\ -\text{tns} \end{bmatrix} \rightarrow [-\text{vocalic}] \quad \text{in env: } \underline{\hspace{1cm}} \text{V}$$

$$\text{C-2} \quad \text{V} \rightarrow \phi \quad \text{in env: } \underline{\hspace{1cm}} + \text{V}$$

C-3 Erase parentheses and return to C-1. If there are no more parentheses, then proceed to P-1.

$$\text{P-1} \quad \begin{bmatrix} +\text{obst} \\ +\text{cmp} \end{bmatrix} \rightarrow \begin{bmatrix} +\text{str} \\ -\text{grv} \end{bmatrix} \quad \text{in env: } \underline{\hspace{1cm}} \begin{bmatrix} -\text{cns} \\ -\text{grv} \end{bmatrix}$$

$$\text{P-2} \quad \underline{e} \rightarrow \underline{o} \quad \text{in env: } \underline{\hspace{1cm}} [+nasal][+cons]^4$$

$$\text{P-3} \quad [+cons] \rightarrow [+sharp] \quad \text{in env: } \underline{\hspace{1cm}} \begin{bmatrix} -\text{cns} \\ -\text{grv} \end{bmatrix}$$

$$\text{P-4} \quad \left\{ \begin{array}{l} [+grave] \rightarrow \bar{u} \\ [-grave] \rightarrow \bar{o} \end{array} \right\} \quad \text{in env: } \begin{bmatrix} \underline{\hspace{1cm}} \\ +\text{voc} \\ -\text{cns} \end{bmatrix} [+nasal][+cons]$$

$$\text{P-5} \quad \begin{bmatrix} -\text{voc} \\ -\text{obs} \end{bmatrix} \rightarrow \phi \quad \text{in env: } \underline{\hspace{1cm}} + [+cons]$$

$$\text{P-6} \quad \begin{bmatrix} +\text{voc} \\ -\text{cns} \\ +\text{diff} \\ -\text{tns} \end{bmatrix} \rightarrow \phi^5$$

$$\text{P-7} \quad \underline{w} \rightarrow \underline{y}$$

(XVIII. LINGUISTICS)

We apply these rules to some of the phonemic representations given above:

sidjat: ((sĭd+ē+ĭ̇)+n+tu) →C-2→ ((sĭd+ĭ̇)+n+tu) →C-3→  
(sĭd+ĭ̇+n+tu) →C-3→ sĭd+ĭ̇+n+tu →P-3→ s,ĭd,+ĭ̇+n+tu  
→P-4→ s,ĭd,+ō̇+n+tu →P-5→ s,ĭd,+ō̇+tu →P-6→ s,ĭd,  
+ō̇+t → s,id,át

govorjat: ((gouor+ĭ̇+ĭ̇)+n+tu) →C-1→ ((gowor+ĭ̇+ĭ̇)+n+tu)  
→C-2→ ((gowor+ĭ̇)+n+tu) →C-3→ (gowor+ĭ̇+n+tu) →C-3→  
gowor+ĭ̇+n+tu →P-3→ gowor,+ĭ̇+n+tu →P-4→ gowor,+ō̇+n+tu  
→P-5→ gowor,+ō̇+tu →P-6→ gowor,+ō̇+t →P-7→ gowor,+ō̇+t  
→ gəvar,át

nesut: ((nes+é̇)+n+tu) →C-3→ (nes+é̇+n+tu) →C-3→ nes+é̇  
+n+tu →P-2→ nes+ō̇+n+tu →P-3→ n,es+ō̇+n+tu →P-4→  
n,es+ū̇+n+tu →P-5→ n,es+ū̇+tu →P-6→ n,es+ū̇+t →  
n,ɪsút

znajut: ((znō̇ĭ+e)+n+tu) →C-1→ ((znō̇j+e)+n+tu) →C-3→  
(znō̇j+e+n+tu) →C-3→ znō̇j+e+n+tu →P-2→ znō̇j+o+n+tu  
→P-4→ znō̇j+ū̇+n+tu →P-5→ znō̇j+ū̇+tu →P-6→ znō̇j+ū̇+t  
→ znájut

načala: (na+=+kin+l+ō̇) →C-3→ na+=+kin+l+ō̇ →P-1→ na+=  
+čĭn+l+ō̇ →P-3→ na+=+č,ĭn+l+ō̇ →P-4→ na+=+č,ō̇n+l+ō̇  
→P-5→ na+=+č,ō̇+l+ō̇ → nəč,ilá (Fem. Past)

načnut: ((na+=+kin+é̇)+n+tu) →C-3→ (na+=+kin+é̇+n+tu) →C-3→  
na+=+kin+é̇+n+tu →P-1→ na+=+čĭn+é̇+n+tu →P-2→ na+=  
+čĭn+ó̇+n+tu →P-3→ na+=+č,ĭn+ó̇+n+tu →P-4→ na+=  
+č,ĭn+ū̇+n+tu →P-5→ na+=+č,ĭn+ū̇+tu →P-6→ na+=+č,n+ū̇+t  
→ nač,nút

T. M. Lightner

Footnotes

1. For an explanation of the abbreviations used in our phonemic transcription, see T. M. Lightner, Remarks on the morphophonemic component of Russian (henceforth [RMCR]), Quarterly Progress Report No. 69, Research Laboratory of Electronics, M.I.T., April 15, 1963, pp. 193-199.

2. We require the presence of /i/ in the roots /kin/, /gim/, /gin/ not only to account for the shift of k and g to č, and z, but also to account for the tense /i/ in derived imperfectives such as nacínat' 'to begin,' otzimat' 'to wring out,' dozinat' 'to finish reaping,' and for the nondiffuse grave /o/ in derived nominals such as konec 'end.'

3. For further examples see T. M. Lightner, On obrazovat, and pon, at, type verbs in Russian, Quarterly Progress Report No. 67, Research Laboratory of Electronics, M.I.T., October 15, 1962, pp. 177-180.

4. The rule stated here is actually more general because e → o also before l followed by consonant. Thus, e.g., we shall derive žolob from /gelb/, and we shall account for the e/o alternation in Inf. mólot' but 3 Sing. melet by deriving both forms from the verb stem /mel/: Inf. (mel + tī), 3 Sing. ((mel + e) + tu). See Roman Jakobson, Remarques sur l'évolution phonologique du russe comparée à celle des autres langues slaves, III, Sec. 3, p. 21, Travaux du Cercle Linguistique de Prague, II (1929).

5. This rule, of course, will be preceded by a rule that lowers {u, i} in "strong" position. These two rules will account for such alternations as son (nom. sg.)/sna (gen. sg.), krasen (masc. short)/krasna (fem. short), etc. See Section XVIII-E.

#### E. NOTES ON THE VERBS čítat' AND -čest'

Both verbs čítat' and -čest' are formed from the root /kit/. The Imperfective requires the verb suffix /ōj/ and root vowel length: /kīt + ōj/. The Perfective requires no verb suffix and retention of the short vowel. The forms are thus as follows:

	<u>Imperfective</u>		<u>Perfective</u>
Inf:	<u>č, itát,</u> (kīt + ōj + tī)		<u>č, és, t,</u> (kīt + tī)
Masc. Past:	<u>č, itál</u> (kīt + ōj + l + os)		<u>č, ól</u> (kīt + l + os)
Fem. Past:	<u>č, itále</u> (kit + ōj + l + ō)		<u>č, lá</u> (kit + l + ō)
3 Sg:	<u>č, itájit</u> ((kīt + ōj + e) + tu)		<u>č, t, ót</u> ((kit + é) + tu)
3 Pl:	<u>č, itájut</u> ((kīt + ōj + e) + n + tu)		<u>č, tút</u> ((kit + é) + n + tu)
Imper:	<u>č, itáj</u> ((kīt + ōj + e) + ī + #)		<u>č, t, í</u> ((kit + é) + ī + #)

In addition to the rules already formulated in previous reports,<sup>1</sup> we shall require two more rules. One rule will lower "strong" {u, i}; the other rule will eliminate "weak" {u, i}. These two rules will account for both the presence of root vowel in the Masc. Past č, ól and the absence of root vowel in the Fem. Past č, lá. The two rules are as follows:

$$A: \left\{ \begin{array}{c} \underline{u} \\ \underline{i} \end{array} \right\} \rightarrow \left\{ \begin{array}{c} \underline{o} \\ \underline{e} \end{array} \right\} \quad \text{in env: } \text{-----} C_0 \left\{ \begin{array}{c} \underline{u} \\ \underline{i} \end{array} \right\}$$

$$B: \left\{ \begin{array}{c} \underline{u} \\ \underline{i} \end{array} \right\} \rightarrow \phi$$

Thus the Masc. and Fem. Past will have the following derivations:

$$\begin{aligned} \text{Masc: } & (k\acute{i}t+l+os) \rightarrow k\acute{i}t+l+os \rightarrow \check{c}\acute{i}t+l+os \rightarrow \check{c},\acute{i}t+l+os \rightarrow \\ & \check{c},\acute{i}t+l+us \rightarrow \check{c},\acute{i}t+l+u \xrightarrow{-A} \check{c},\acute{e}t+l+u \xrightarrow{-B} \check{c},\acute{e}t+l \rightarrow \\ & \check{c},\acute{o}t+l \rightarrow \check{c},\acute{o}+l \rightarrow \check{c},\acute{o}l \\ \text{Fem: } & (kit+l+\acute{o}) \rightarrow kit+l+\acute{o} \rightarrow \check{c}it+l+\acute{o} \rightarrow \check{c},it+l+\acute{o} \xrightarrow{-B} \\ & \check{c},t+l+\acute{o} \rightarrow \check{c},+l+\acute{o} \rightarrow \check{c},l\acute{a} \end{aligned}$$

It can be seen, however, that no modification of rule A will predict the retention of root vowel in the Infinitive. I believe, therefore, that in addition to rule A one must also posit

$$A': \left\{ \begin{array}{c} \underline{u}' \\ \underline{i}' \end{array} \right\} \rightarrow \left\{ \begin{array}{c} \underline{o}' \\ \underline{e}' \end{array} \right\}$$

The Infinitive will now be derived in the following manner:

$$\begin{aligned} \text{Inf: } & (k\acute{i}t+t\bar{i}) \rightarrow k\acute{i}t+t\bar{i} \rightarrow \check{c}\acute{i}t+t\bar{i} \rightarrow \check{c},\acute{i}t+t,\bar{i} \xrightarrow{-A'} \check{c},\acute{e}t+t,\bar{i} \\ & \rightarrow \check{c},\acute{e}t+t, \rightarrow \check{c},\acute{e}s+t, \rightarrow \check{c},\acute{e}s,+t, \rightarrow \check{c},\acute{e}s,t, \end{aligned}$$

There is no difficulty in any of the Imperfective forms except for the rule that lengthens the root vowel of /kit/ to /k $\bar{i}$ t/. Although at present I can give no precise formulation of this rule, I think that the lengthening of this vowel must be accounted for by the same rule that lengthens {u, i} in derived Imperfectives like nazyvat' 'to call' (from /zuv/, cf. Perfective nazvat') and dobirat' 'to gather' (from /bir/, cf. Perfective dobrat').

T. M. Lightner

#### Footnotes

1. See Section XVIII-D, and T. M. Lightner, Remarks on the morphophonemic component of Russian, Quarterly Progress Report No. 69, Research Laboratory of Electronics, M.I.T., April 15, 1963, pp. 193-199.

#### F. THE SHIFT OF s TO x IN OLD CHURCH SLAVONIC VERB FORMS<sup>1</sup>

It is well known that Indo-European s preceded by i, u, r, k shifted to x in Proto-Slavic.<sup>2</sup> In this report we mention a few OCS verb forms in x (s before front vowels)



for which a synchronic analysis of OCS must posit phonemic s and thus reflect the historic s → x sound shift.

1. The 2 Sing. Pres. Tense ending:

<u>Phonetic</u>	<u>Phonemic</u>	<u>English</u>
dasi	(dōd+sī)	you give
molišī	((mol+ī+ī)+sī)	you beg
mъnišī	((min+ē+ī)+sī)	you think
glagolješī	((golgol+ō+e)+sī)	you speak
miluješī	((mīl+ou+ō+e)+sī)	you pity
etc.		

2. Aorist:

1 Sg: molixъ	((mol+ī+s)+u)
2 Pl: moliste	((mol+ī+s)+te)
3 Pl: molišę	((mol+ī+s)+in)
1 Sg: glagolaxъ	((golgol+ō+s)+u)
2 Pl: glagolaste	((golgol+ō+s)+te)
3 Pl: glagolašę	((golgol+ō+s)+in)
etc.	

3. Imperfect:

1 Pl: moljaaxomъ	((mol+ī+ēō+s)+e+mu)
2 Pl: moljaašete	((mol+ī+ēō+s)+e+te)
3 Pl: moljaaxę	((mol+ī+ēō+s)+e+on)
etc.	

In the Imperfect there are no telling forms with phonetic s (cf., however, 2 Sing. Pres. dasi and 2 Pl. Aorist moliste), but the analysis with s rather than x must be preferred because then the s will be common to both Past Tense forms. We find external confirmation for the correctness of this solution in historically younger forms that have -æasta/-æaste/-æaste in the 2 Dual/3 Dual/2 Plural Imperfect.<sup>3</sup>

The rule for the shift of s to x is as follows:

(XVIII. LINGUISTICS)

s → x in env: [-obstruent] + \_\_\_\_\_ [-obstruent]

This rule explains the retention of s in the 2 Sing. Pres. of athematic verbs (/dōd+sī/) but not in the 2 Sing. Pres. of thematic verbs (/...+ī+sī/ or /...+e+sī/), and the retention of s in the 2 Dual, 3 Dual, 2 Plural Aorist (/...+s+ta/ or /...+s+te/) but not in the other Aorist forms (/...+s+u/, /...+s+in/, etc.). Furthermore, this rule explains the otherwise inexplicable "reappearance" of s in the younger -æasta, -æaste (opposed to the older -æaseta, -æasete).

We shall not dwell on the already well-known fact that the environment of this rule must be expanded to include pre-s-velars because of (to mention but one example) Aorist forms like raxъ, raste, ræše from the root /rek/.

T. M. Lightner

Footnotes

1. This report is extracted from a larger work on OCS morphophonemics presented at Linguistics Seminars, Research Laboratory of Electronics, M.I.T., March 13, 1962 and July 15, 1962.

2. See, e.g., Roman Jakobson, Comparative Slavic phonology, Selected Writings (Mouton and Company, 's-Gravenhage, 1962), p. 413.

3. See, e.g., H. G. Lunt, Old Church Slavonic Grammar (Mouton and Company, 's-Gravenhage, 1959), sec. 9.12, p. 86. According to G. Nandriš, Old Church Slavonic Grammar (Athlone Press, London, 1959), p. 147, Sava's Book (Savvina kniga) and the Ostromirovo Evangelie use the -æasta/-æaste forms exclusively, whereas the Codex Zographensis and the Glagolita Clozianus use only the -æseta/-æsete forms.

G. ALTERNATION OF RULES IN CHILDREN'S GRAMMAR

1. Problem and Method

The process by which language is acquired has been postulated to be stimulus-response-reinforcement associations<sup>1</sup> or drive-cue response-reinforcement associations.<sup>2</sup> Presumably, the environment creates the drive or need to imitate a model. This drive at some stage in development is cued by particular utterances. These utterances are imitated and the response is reinforced either by reiteration, some kind of response from the model or internal gratification. The data (language production or preception) are analyzed according to the particular theoretical point of view held. They are also described simply in terms of traditional linguistic labels. Language is divided into tact and mands.<sup>1</sup> Language is analyzed in terms of cue-response-reward situations.<sup>2</sup> From the purely descriptive point of view the acquisition of phonemes by the infant and young child<sup>3</sup> and the usage of differently structured sentences at various age levels have been described.<sup>4</sup>

Essentially the same conclusions are reached from research undertaken from the

theoretical viewpoints mentioned above and studies using labelling procedures. As the child matures the proportion of usage of more complex language increases. He proceeds from the simple to the complex. At the level of meaningful utterances, the proportion of usage of more complex sentences increases, and, therefore, sentence length increases. Complexity is usually intuitively defined as more difficult and, in a circular fashion, seems, in large part, to be dependent on an expanding lexicon and increasing sentence length.

This study was undertaken to attempt to formalize the notion of increasing complexity in grammar as children mature and in this way to examine further the hypothesis of imitation in language acquisition. The model of grammar used in this study is a generative or transformational model. It is hypothesized that the perceiver or child has incorporated both the generative rules of the grammar and a heuristic component that samples an input sentence and by a series of successive approximations determines which rules were used to generate this sentence.<sup>5-7</sup> Instead of memorizing every sentence that he has been exposed to and imitating these sentences, he uses a set of rules to generate not only the sentences that he has heard, but also other possible examples.

The population in this study was comprised of 159 children, ranging in age from 2 years, 10 months to 7 years, 1 month. They were homogeneous in socio-economic status and I.Q. Language was elicited and recorded in various stimulus situations: 1) responses to a projective test, 2) conversation with an adult, and 3) conversation with peers. The last two situations took place both in controlled and free, that is, classroom, environments. The language sample produced by each child was analyzed by using the transformational model previously described. A grammar was written which included all of the postulated rules used to generate the sentences in the total language sample.<sup>8</sup>

It was found that all of the basic structures used by adults to generate their sentences which we have thus far been able to describe are in the grammar of children from 3 to 4 years of age. Three developmental trends were observed. The first was the very rapid acquisition of some phrase-structure rules that were excluded by some children under three. The second was that increasing numbers of the children used certain transformations as an increasingly older population was observed. The third was the use of alternate rules at all three levels of the grammar.<sup>9</sup> That is, in terms of this model of a grammar, the children generated their sentences from rules that produced completely grammatical structures and simultaneously from rules that did not. These latter rules have been termed rules restricted to a children's grammar. This report deals with this last developmental trend.

Three facts should be kept in mind about the use of these restricted forms. Some of these rules can be found in the grammar of some adults, although it seems unlikely that they are in the grammar of the parents of these children, since all parents have

(XVIII. LINGUISTICS)

occupations that fall within the upper 24 per cent of a middle class population. The sentences produced from these rules were deviations from complete grammaticalness but were not completely outside a set of possible sentences. For example, the child might generate the sentence 'You can't put no more water in it.' from these rules but not the nonsentence 'More water no put can't you in it.' or any such departures. Finally, these restricted forms occur infrequently in the total language sample. However, I believe that a close examination of the use of these rules does bring clarification and greater formality to the notion of children using an increasingly complex grammar as they mature.

2. Results

As was stated before, the use of alternate rules occurs at all three levels of the grammar. At the phrase-structure level almost all of the children use verbs, nouns, articles, prepositions, and particles correctly. At the same time, some of the children omit these parts of speech or use them redundantly. Also, substitution of verbs, articles, and prepositions takes place. As examples, verbs are omitted in 'This green.', verbs are substituted in 'I put them to the doctor's.', and verbs are used redundantly in 'He'll might get in jail.'

At the transformational level, having optionally chosen a transformation, the children sometimes do not observe all of the rules that are attendant on the use of this transformation. As an example, at the beginning of the age range when the verb phrase is being expanded from just the main verb to be + present participle + verb and the contraction transformation is applied, as it always is except in emphatic sentences, the children sometimes omit the contracted form and produce 'I going.' Later on, when the verb phrase is further expanded to have + perfect participle + be + present participle + verb and the contraction transformation is applied, the same result occurs and 'I been thinking about that.' is produced.

The most commonly used general transformation is conjunction, in which one sentence is added to another sentence. Sometimes the children use the conjunction transformation without using the rule that verb tense in the second sentence must agree with verb tense in the first sentence. An example of this is 'They mixed colors and pour buckets.'

At the morphology level of the grammar, again, omissions of rules, redundancies in rules, and substitutions take place. These occur with verbs, nouns, adverbs, pronouns, and possessive forms. As examples: noun endings are omitted in 'She has lots of necklace.' and 'He's next to a few stone.' Noun stems are substituted in 'Those are wolfs.' and 'We have childs in this school.', and noun endings are used redundantly in 'There's furnitures.' and 'Where are the peoples?'

Some quantitative statements can be made about the use of these restricted forms.

In general, the use of alternate rules gradually declines from the beginning of the age range to its end. There are significant differences found when one compares the percentages of children using these forms at the beginning of the age range and at its end. However, this decline is not asymptotic in nature, but, rather, fluctuating throughout the age range. Indeed, the specific structures that are formulated by alternate rules and the particular alternate rules used by a sizable number of the children at any given age period change as an increasingly mature population is observed.

When we divide the age range into 4-month periods and average the age periods in which there is peak usage of restricted forms, that is, all types of omissions, redundancies, and substitutions and all types of nonobservations of rules needed to produce simple and general transformations, and we then average the peak percentages of children using these forms, we see the following:

At the phrase-structure level the type of restricted form that peaks earliest and highest is omission. Then comes substitution, and finally redundancy. At the transformation level we see that nonuse of rules with simple transformations peaks earliest and highest. The peak for nonuse of rules with general transformations occurs later, when, in fact, more of the children are using some general transformations. At the morphology level, again, omissions peak earliest, then substitutions, and later, redundancies. At this level redundancies peak highest.

These trends can be seen in Figs. XVIII-1, XVIII-2, and XVIII-3. The real percentages of children using these forms at the beginning of the age range (from 2 years, 10 months to 3 years, 2 months) and at its end (from 6 years, 10 months to 7 years, 1 month) are also given.

I believe that there are qualitative, as well as quantitative, statements that can be made. The children's usage of grammar, and the word usage is stressed, did become increasingly complex over the age range observed. This complexity is not related simply to increasing sentence length or proportion of usage of what has been termed compound and complex sentences or, in the terms of this model of a grammar, general transformations. These changes seem to be extensions of behavior without additional rules in the grammar. Conjoining two sentences needs the same application of rules as conjoining three or four. To delete and substitute as in relative clauses needs the same application of rules whether we do it once in ten minutes or three times. This increasing complexity also seems to be very dependent on the child's ability to proceed from the application of the most general rule to the application of increasingly differentiating rules to produce a particular syntactic structure. In accordance with the model used, all instances of the use of a restricted form represent the use of an elementary rule, without or with some of the additional steps required. That is, rules for addition, deletion, permutation, and substitution were applied but without observation of ordering in some instances and in other instances without using the combination of these

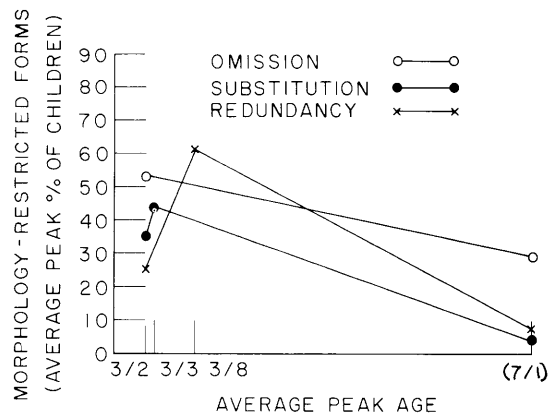


Fig. XVIII-1. Ages for peak usage of restricted forms at the phrase-structure level.

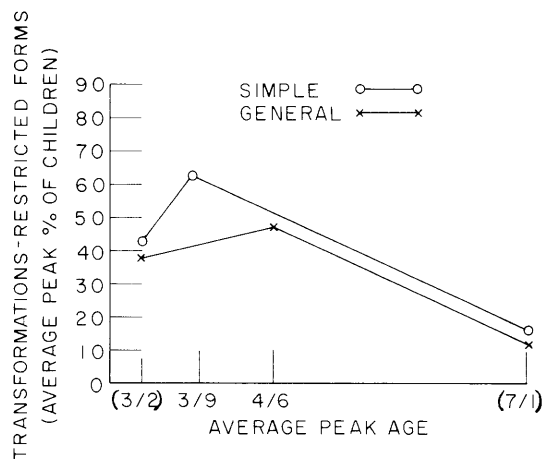


Fig. XVIII-2. Ages for peak usage of restricted forms at the transformation level.

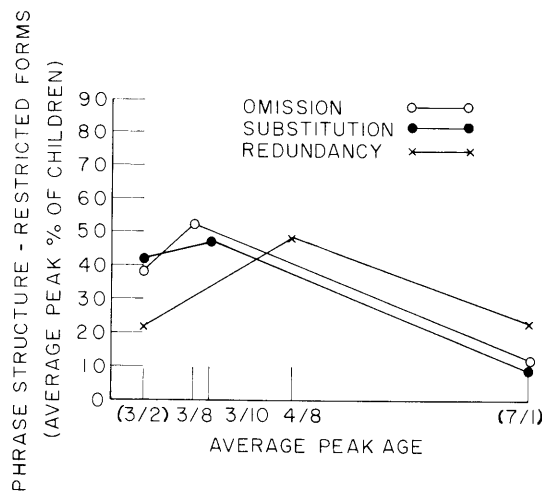


Fig. XVIII-3. Ages for peak usage of restricted forms at the morphology level.

elementaries required to produce the completed form of these structures.

Varying degrees of generalization take place in this process from greatest generalization to increasing differentiation to complete differentiation and, I believe, new organization. This process seems to reflect the hypothesized model, that is, the child determining by a series of successive approximations the rules used to generate a sentence. For example, omission would represent the application of the most general rule, then substitution, and redundancy before complete differentiation. One aspect of this new organization is differentiation between nonterminal and terminal rules, for example, to formulate that the use of a modal, such as can, may or will, + verb is a terminal rule and that two modals are mutually exclusive in the same context. We may say, 'He might get into jail.' or 'He will get into jail.' but not 'He'll might get in jail.' Other examples are to formulate that the question transformation is terminal and cannot be substituted to produce new sentences (we say, 'What is that?' but not 'I know what is that. '), and that certain endings of verbs and nouns are also terminal. We say pushed and people, but not pushted and peoples.

Perhaps we can say that younger children's usage of grammar is simpler than older children's or adults' usage of grammar because proportionally more of the younger children use an incomplete set of rules to produce some syntactic structures and because increasing levels of differentiation in the use of rules are found, going toward complete differentiation, as older children are observed. The word usage is again stressed because, although in some instances a child applies the elementary rule or an incomplete set in the formulation of a syntactic structure, in other instances he applies the complete set of ordered rules to the formulation of this same structure. He seems to display competence although this is not always realized in performance.

We now come to the question of what these data mean in terms of an hypothesis of motivated imitation in language acquisition. The limitations of the nervous system for memorizing all instances of sentences heard and storing them for later use seem to obviously negate the hypothesis of language acquisition as primarily an imitative function. However, there are some theorists who divide language learning into a twofold process: Early learning is the establishment of stimulus response associations with responses learned by imitation, and later learning (the first meaningful utterances) is accomplished by cognitive processes. From this point of view one would predict that the sounds and words produced first would represent complete mastery of articulation, but this is not the case. Children produce sentences long before articulation has been mastered. Finally, if language acquisition is an imitative function, then children should be producing, first, sentences with omissions because of the limitations of memory, and then complete sentences. One might assume that the other types of restricted forms produced by children are a result of imitation of peers. In that case one would expect a very random production of these restricted forms. The results of this study seem to indicate that the

(XVIII. LINGUISTICS)

process is neither random nor one of remembering to put in more of the missing parts of sentences as the child matures. Rather, the restricted forms produced beyond omission, such as substitution and redundancy, reflect the child's improved ability to generate particular structures from increasingly more differentiated sets of rules as he matures.

Paula Menyuk

References

1. B. F. Skinner, Verbal Behavior (Appleton-Century-Crofts, Inc., New York, 1957).
2. J. Dollard and N. E. Miller, Personality and Psychotherapy (McGraw-Hill Book Company, Inc., New York, 1950).
3. O. C. Irwin, Infant speech, J. Speech and Hearing Disord. 16, 159-162 (1951).
4. D. McCarthy, Language Development of the Preschool Child (University of Minnesota Press, Minneapolis, 1930).
5. M. Halle and K. Stevens, Speech recognition: A model and a program for research, Trans. IRE, Vol. IT-8, pp. 155-159, February 1962.
6. G. H. Matthews, Analysis by Synthesis of Natural Languages, Internal Memorandum, Mechanical Translation Group, Research Laboratory of Electronics, M. I. T., 1961.
7. G. A. Miller and N. Chomsky, Finitary models of language users, Handbook of Mathematical Psychology, Vol. 2, edited by R. R. Bush, E. H. Galanter, and R. D. Luce (John Wiley and Sons, Inc., New York, in press).
8. N. Chomsky, Syntactic Structures (Mouton and Company, 's-Gravenhage, 1957).
9. P. Menyuk, Syntactic rules used by children from preschool through first grade, J. Child Development (in press).