

---

# The Influence of Speech Performance on Linguistic Structure

THOMAS G. BEVER

Bever, T.G. (1970). The influence of speech performance on linguistic structure. In W. Levelt & G. Flores d'Arcais (Eds.), *Advances in psycholinguistics* (pp 21 50), North Holland.

## INTRODUCTION AND SUMMARY

Recent theoretical developments in the study of language have depended on a strict distinction between linguistic "knowledge" and the implementation of that knowledge in speech performance. This distinction has been extremely useful, since it has allowed linguistic investigations of language structure to proceed without taking into consideration any behavioral properties of that structure as used in everyday speech. Linguists have reveled in the luxury of being able to ignore why we say what we say, how we say it and how others understand it. They have concentrated their efforts on the description of the 'formal' properties of language implied by the existence of particular well-formed sentences *in vitro*. The problem of understanding the behavior elicited by those sentences *in vivo* has been left to experimental psychologists.

Current study of the acquisition of language has depended upon the same distinction. It is generally assumed that language is learned in a series of "sub-grammars." Each subgrammar the child develops as he grows older takes into account the structure of more and more sentences, until finally his grammar is that of an adult. The virtue of such a view of language acquisition is that it allows the investigator to ignore general changes in the child's cognitive system during this period: the investigator's task is merely to isolate and write a grammar for each linguistic stage. Such a structuralist approach underlies the current

---

I am indebted to R. Hurtig, S. Reder, J. Langer and J. Grinder for assistance with the manuscript. This research was supported by PHS PO 1 GM 16737.

Source: G. B. Flores d'Arcais and J. M. Levelt, eds., *Advances in Psycholinguistics* (Amsterdam: North-Holland Publishing Company, 1970). Reprinted by permission of the publisher.

fascination with how much of language development is determined by innate 'linguistic' structures, rather than by structures that characterize all of human cognition and of which linguistic structure is merely one instance.

In this paper I should like to explore some of the drawbacks in our preoccupation with linguistic structure in the child and adult and its concomitant lack of attention to psychological mechanisms for the implementation of that structure. First, many intuitions about the grammaticality of potential sentences *in vitro* are influenced and sometimes determined by the properties those sentences would have in actual behavior: a sentence can sometimes be classified as ungrammatical even though closer inspection suggests that it is grammatically acceptable but behaviorally complex. Second, certain grammatical rules themselves may be shown to be structural accommodations to behavioral constraints. Thus certain universal structural properties of language may express general cognitive constraints rather than particular innate linguistic structures.

#### THE INFLUENCE OF PERCEPTION ON ACCEPTABILITY INTUITIONS

The linguist sets himself the task of finding a linguistic description of all and only those sentences that native speakers of a language feel are potentially grammatical, regardless of whether they have ever heard the sentences said or not. For example, speakers of English will agree that (1) is a possible grammatical sentence while (2) is not, even though (2) is perfectly comprehensible in many contexts:<sup>1</sup>

- (1) Phil discussed Sam in a sharp tone of voice.
- (2) \*Phil spoke Sam about in a sharp tone of voice.

Thus, a sentence may be comprehensible without being grammatical, nor does incomprehensibility imply ungrammaticality.

Consider a now classic example of a type of sentence which is incomprehensible but alleged to be grammatical by linguists, doubly-embedded sentences (cf. Chomsky and Miller, 1963). For example, (3) has a clause embedded within the main clause and is grammatical and comprehensible, but what about (4), in which the clause within a clause has a clause embedded within it?

- (3) The boy the girl left then slept.
- (4) The boy the girl the man left watched then left.

There is no natural way to block such double-embeddings within a grammar which allows single embeddings such as (3). One *could*, of course, argue that a "subordinate-clause-counter" be included as part of the grammar and that the

1. An asterisk indicates an ungrammatical example sentence; I use a question mark to indicate an unacceptable sentence without specifying the basis for its unacceptability.

counter be set at "1" as an upper limit. However, this will not do, since the restriction is not on the number of embedded clauses, but on the number of *self*-embedded clauses. For example, both sentences (5a) with one subordinate clause and (5b) with two such clauses are comprehensible and acceptable:

- (5a) The girl left the boy who then slept.
- (5b) The man watched the girl who left the boy who then slept.

An enterprising linguist could argue that there is a specific linguistic restriction on center-embedded clauses. But this would merely catalogue the restriction, not explain it.<sup>2</sup> From the standpoint of the structural linguist such a taxonomic exercise is sufficient, since he is free to claim that such a restriction is merely an example of an innate structural constraint on linguistic forms. We should not be satisfied with the linguists' desire to place such restrictions *hors de combat*. If we can find a plausible behavioral account for the difficulty of double embedded sentences we can avoid this obscure and scientifically stultifying claim about what is innate to linguistic capacity.

In fact, several authors have attempted to formulate behavioral laws that would predict the difficulty of such sentences as (4). For example, Fodor and Garrett (1967) suggest that the difficulty of such sentences is due to their "density," i.e., to the number of underlying structure sentence units per word in the surface structure, which exceeds some critical threshold ("density" = 3/10 for (4); "density" = 3/12 for (5b)). This proposal is intriguing since it would suggest that at least one dimension of perceptual complexity is quantifiable in linguistic terms. However, that the proposal is incorrect is shown by the fact that the density of underlying sentences per word is even higher in (6) (3/9), yet (6) is entirely comprehensible and acceptable:

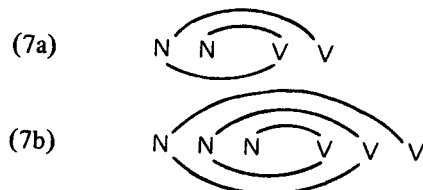
- (6) The man watched the girl leaving the sleeping boy.

The complexity of center-embedded sentences cannot be explained by appeal to this quantitative principle.

Chomsky and Miller (1963) have also attempted to define a perceptual principle that could account for the difficulty of center-embedded sentences. They argue that any perceptual rule may not interrupt its own operation more than once. In the case of a sentence like (3) (represented schematically in (7a)) the perceptual assignment of the 'actor'-'action' relation to the first noun and last

2. There is a current claim that all grammars include a set of "derivational constraints" which restrict the possible form of underlying/surface structure relations (cf. Postal, 1972; Lakoff, 1971). The restriction on doubly center-embedded sentences could be stated with such a device. Also it might be possible to account for the restriction on sentences like (4) in terms of a so-called "output restriction," which state constraints on the final form which surface phrase structures can have (cf. Perlmutter, 1971). However, the restriction appears to be on the underlying/surface relations rather than on the surface sequence itself, viz. examples (26)-(31) below.

verb is interrupted by the same assignment to the second noun and first verb. In (4) (represented in (7b)) the perceptual assignment of actor-action to the first noun and last verb is interrupted by the assignment of actor-action to the second noun and second verb, which is in turn interrupted by the assignment of the same function to the last noun and the first verb. (Upper lines in (7) represent subject-verb relations. Lower lines represent verb object relations.)

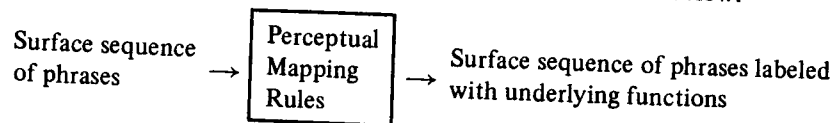


It is intuitively clear that a self-interrupting operation is more complex than one which does not interrupt itself. However, there is no theoretical motivation for one interruption being acceptable (as in (3)) and two interruptions being entirely unacceptable (as in (4)). Chomsky and Miller's argument, as in the case of the hypothetical clause "counter" on double embeddings, also serves to describe the restriction but not to explain it.

#### The Operation of Mapping Rules in Speech Perception

The failure to explain the restriction on self-interruptions is partly due to the fact that this phenomenon is not viewed in the context of a general theory of sentence perception. An explanation of the existence of any particular case of perceptual complexity must follow from a theory which explains how the normal perceptual processes are strained by those complex constructions.

Some recent studies have illuminated normal perceptual processes by considering the way in which listeners extract basic grammatical relations (e.g., 'subject, predicate, object, modifier') from the actual appearance of sentences.<sup>3</sup> (See Fodor, Bever, and Garrett (1968) and Bever (1968) for experimental reviews.) The basic claim of these investigations is that there is a set of perceptual rules which map surface sequences onto the corresponding internal relations. According to this view, speech perception proceeds as outlined below.



3. In this discussion I intentionally obfuscate the distinction between a semantic and a syntactic analysis of the underlying structure relations. That is, I shall not distinguish between a perceptual mechanism which extracts the syntactic 'logical subject, predicate, object' from one which extracts the semantic relations, 'actor, action, object,' or 'agent, action, patient.' This obfuscation does not vitiate the claims made about the general nature of speech perception. It is necessary in order not to make any claim about the form of underlying structure, which is a subject of current linguistic controversy.

For example, in English any sequence of several phrases of the same type (interrupted optionally with conjunctions), with a conjunction between the last two phrases, is a conjoint phrase of the same type as the compound phrases (e.g., 'noun phrase, verb phrase. . .'), (8).

(8) In "... x . . . y conjunction, z . . .", in which x, y . . . z are identical constituent types of type T then the entire sequence is a conjoint phrase of type T, each member of which has the same internal syntactic relation to other sentence constituents as the whole phrase.

(9a) The boy the girl and the man left.

(9b) The boy watched the girl, left the man and slept.

Thus the conjunct phrases in the sentences in (9) can be mapped directly onto their underlying syntactic relations. For example, after (8) has applied to (9a) it is labeled as in (10).

(10) The boy the girl and the man left.

(Conjoined noun phrase, each phrase being a separate subject of the verb.)

Such perceptual rules as (8) apply without reference to the full grammar. Rather than using the grammar in an analysis-by-synthesis recognition routine (cf. Halle and Stevens, 1964) or using an ordered series of "inverse transformations" each corresponding to a transformation to "detransform" the surface tree back to the underlying tree (Herzberger, 1966), many perceptual rules appear to provide *direct* mappings of the surface sequences onto the underlying syntactic relations.

The demonstration that this direct-mapping model is most appropriate for speech perception would be too time consuming for this discussion, particularly since much empirical work must still be done to test it adequately. However, if one accepts this view it is possible to subsume the unacceptability of double embedded sentences under a *general* conceptual restriction which simultaneously accounts for the perceptual difficulty of a number of other kinds of sentence constructions.

#### The Perceptual Basis for the Complexity of Double Embeddings

Given perceptual mapping rules such as (8), center-embedded sentences like (4) may be difficult because at first they may be misinterpreted as two compound phrases (e.g., (4) is misunderstood as "*the boy, the girl and the man left, watched and slept*"). Indeed Blumenthal (1967) found that subjects most often understand center-embedded sentences as though they were a compound noun followed by a compound verb. Furthermore, Fodor and Garrett found that sentences like (4) are a great deal easier to understand if the relative pronouns are included (e.g., "*the boy whom the girl whom the man watched left slept*").

Presumably this is partially due to the fact that the presence of the relative pronouns makes rule (8) inapplicable. Accordingly, part of the complexity of center-embedded constructions is that they present plausible sequences for inappropriate application of perceptual mapping rule (8). However, rule (8) does not explain why center-embedded sentences are difficult to process correctly; (8) applies to such sentences only after a correct interpretation cannot be found and explains only the attractiveness of the preferred inappropriate interpretation.

The action of another perceptual mapping rule does partially explain the initial complexity of double-embeddings. Consider the perceptual mapping rule that assigns the functions of underlying subject and object to noun phrases with relative clauses (11). This rule capitalizes on the fact that in clause initial (or postverbal) position two adjacent noun phrases followed by a verb other than *be* (with other noun phrases optionally intervening) are uniquely related such that the first noun phrase is the object (direct or indirect) of an underlying clause of which the second noun phrase is the subject. For example, the initial sequence of noun phrases in (3) and (4) would be assigned the appropriate relations by (11) as shown in (12) and (13):

(11) In a surface sequence ' $\dots NP_1 NP_2 (\neq who) (NP^*) V(\neq be) \dots$ ',  $NP_1$  is the object of a clause of which  $NP_2$  is the subject.

(12) The boy    the girl    left    ...  
Object    Subject

(13) The boy    the girl    the man    left    ...  
Object    Subject/  
Object    Subject

Clearly (11) must apply twice to double-embeddings to mark the middle noun phrase as both a subject and an object. This double marking by the same perceptual rule lies at the heart of the difficulty of center-embedded sentences.

There is a general restriction on the utilization of any conceptual dimension, (14),

(14) A stimulus may not be perceived as simultaneously having two positions on the same classificatory relation.

which interacts with the double application of rules like (11) to double-embeddings. (14) articulates the tautology that a stimulus cannot be perceived in two incompatible ways at the same time. For example, a noun phrase in the surface sequence cannot simultaneously be subject and object of the same verb. This principle, when applied according to the view of speech perception as a direct mapping of external sequences onto internal structures, will predict the difficulty of any sequence in which a phrase has a 'double function' in such a mapping operation. However, before applying (14) to explain the difficulty of center-embedded sentences, let us consider its application to some well-known facts.

Miller and Selfridge (1950) found that sequences with low-order probability approximations to English are difficult to perceive, e.g., a sequence like (15) is more difficult than (16):

(15) he went to the newspaper is in deep (2nd-order approximation)

(16) then go ahead and do it if possible (7th-order approximation)

(a '2nd-order approximation' is generated by giving a subject a word (e.g., 'went') and asking him to produce the next word of a sentence ('to'); the next subject is given the last word of the sequence ('to') and produces the next word ('the') and so on. A '7th-order approximation' is generated by giving each subject the last six words of the sequence and asking for the next word.) The relative ease of perceiving sentences as they increase in order of approximation was taken by Miller and others as evidence for the organizing role of syntactic structure at levels higher than a single word. For example, in (16) the sequence forms a sentence, while in (15) it does not. However, reference to the 'organizing properties of syntax' does not explain why low orders of approximation are behaviorally difficult in the first place. In fact, if forming a sentence makes word strings easy to perceive, it might be predicted that sequence (15) should be psychologically simpler since it simultaneously forms *two* sentences (as shown in (17) and (18)):

(17) he went to *the newspaper*

(18) *the newspaper* is in deep

The real basis of the psychological difficulty is clear: the italicized portion of the sequence is vital to both sentences, that is, it has a 'double function'. Psychological complexity results whenever such double functions appear, even in nonlinguistic stimuli. Consider, for example, the representation of the two adjacent squares in Figure 1. The line labeled 'y' is simultaneously shared by the right and left squares. As a result, Figure 1 is generally perceived as a divided rectangle rather than two adjacent squares. Often such double functions can produce 'impossible' figures from the combination of two 'possible' figures, e.g., Figure 2. The general psychological principle that governs these visual examples is a special case of (14): *in a closed system a component of a stimulus cannot serve two opposite functions at the same time*. For example, in Figure 1 line y cannot both end one square *and* begin another; or in Figure 2 the segment labeled

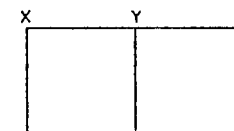


Figure 1. Figure most easily seen as a rectangle with one division at 'y', rather than two squares joined at 'y'.

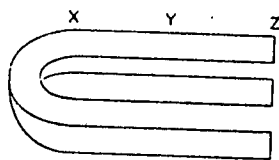


Figure 2. Figure that is 'impossible' because of combination of 2- and 3-dimensional projection at point 'y'.

'y' cannot both end one kind of figure (the 3-dimensional 'u' opening right in the segments labeled x-y) and begin the other (the three poles in the segments labeled y-z).

There is a related explanation for the psychological difficulty of 'center-embedded' sentences. Phillips and Miller (1966) noticed that in a sentence like (4) the second noun is the subject of one clause and the object of another (see 7) and suggested that this might be involved in the complexity of such sentences. According to our view, understanding a sentence involves labeling each phrase with its logical function in the underlying structure by means of perceptual mapping rules. The second noun in (4) could be interpreted as having a 'double function' with respect to perceptual rule (11) since the first noun is labeled by the same rule as the object of a verb of which the second noun is the subject. With respect to the preceding noun, the second noun is a subject, while it is an object with respect to the following noun (see 13).

The general double-function hypothesis for speech perception following from (14) is this:

- (19) In a sequence of constituents x, y, z, if x has an internal relation  $R_i$  to y and y has the same internal relation to z, and x, y, and z are superficially identical in construction type, then the stimulus is relatively complex, due to y's double function in the perceptual mapping rule,  $r_i$ , in which y is both a p and q. (Note that  $R_i \neq$  conjunction)

$$pq \xrightarrow{r_i} p R_i q$$

Thus doubly embedded sentences are complex because they involve the second noun phrase in a double function with respect to the perceptual mapping rule (11).

It should be emphasized that the difficulty implied by (19) does not refer to all cases in which a single phrase has more than one role in the deep structure. For example, in (3) "the boy" is both *object* of "left" and *subject* of "sleep," with no attendant perceptual complexity. Similarly in (20) "the boy" is simultaneously the underlying *subject* of "sleep," the *object* of "like," and the *indirect object* of "give." Yet the sentence is quite simple:

- (20) The girl liked the sleeping boy Sam gave the sandwich to.

Principle (19) applies only to these cases in which exactly the same perceptual mapping rule is used to assign different functions to the same phrase. That is, it is not the 'double function' of the middle noun phrase in the underlying structure that makes such sentences complex. It is, rather, the double application to it of perceptual rule (11). In applying the rule the listener first categorizes the middle phrase as the right-hand member of a pair related by rule (11) and then must characterize it immediately as the left-hand member of the same kind of pair defined by the same rule. In the terms used in (19) the middle noun phrase of (4) is simultaneously a 'p' and 'q' of the same mapping rule.<sup>4</sup>

There are other examples of complex constructions explained by principle (19). Consider the relative complexity of sentences (21) and (22):

(21)? Maxine did not ask Harvey not to say he would not go.

(22) Maxine did not ask Harvey to say he would not go.

(21) is an example of "triple negation," which has often been recognized as extremely complex, if acceptable at all. Like single embedded sentences, sentences with two negation markers are perfectly comprehensible and acceptable (as in (22)). Principle (14) predicts the relative difficulty of sentences with three negations. The perceptual mapping operation corresponding to the negative marks the second 'not' in (21) as simultaneously the scope of the first negation and the operator on the third negation. The result of this mapping is outlined in (23):

(23)? Maxine did not ask Harvey not to say he would not go.

$$\text{Negator} \left( \begin{array}{c} \text{Negated} \\ \text{Negator} \left( \begin{array}{c} \text{Negated} \\ \text{Negator (Negated)} \end{array} \right) \end{array} \right)$$

Following principles (14) and (19), any sequence containing such a double perceptual function is perceptually complex.

Principle (19) also explains some examples of perceptually complex constructions which are intuitively of the same sort as the preceding examples. Consider the sentences in (24) and (25):

(24)? They were tired of discussing *considering* producing toys.

(25)? They were tired of the discussion of *the consideration* of the production of toys.

In each case the sentences are extremely difficult to understand, if they are acceptable at all. As in double-embedding and triple-negation sentences, the complexity of these sentences is a function of the presence of three superficially identical phrases in which the second phrase is modified by the first

4. Presumably there is an analogous strategy for relating the verbs to each other in sentences like (4), which also underlies the perceptual complexity of embedded sentences.

phrase in the same way that the second phrase modifies the third phrase (in the underlying structure). Consider the relative perceptual ease of these sentences if only two phrases occur:

- (26) They were tired of discussing producing toys.
- (27) They were tired of the discussion of the production of toys.

The sentences in (24) and (25) also become much easier to understand if the internal relations among the three critical phrases are varied as in (28) and (29).

- (28) They were tired of discussing *ceiling* producing toys.
- (29) They were tired of the discussion of the *evolution* of the production of toys.

The explanation is that the middle phrase no longer has a double function since *different* perceptual mapping rules relate the first two and second two phrases. (Note that in (28) the middle phrase ("ceiling") is the underlying structure object of the following phrase ("producing") while the first ("discussing") and second phrase are not directly related. In (29) the middle phrase ("evolution") is the action carried out by the third phrase ("production") but the object of the first phrase ("discussion" . . .).

Finally, (26) and (27) become perceptually simpler if the superficial form of the critical phrases is varied, even while the internal relations are held constant; as in (30) and (31):

- (30) They were tired of discussing the consideration of producing toys.
- (31) They were tired of the discussion of considering the production of toys.

The explanation for the relative ease of (30) and (31) is that the middle phrase does not have a double function with respect to the same perceptual mapping rule. The superficial difference in the middle phrase allows the listener to use a different mapping rule for relating the first two phrases than for relating the second two phrases.<sup>5</sup>

In brief, I have tried to show that if speech perception is viewed (at least in part) as a direct mapping of external sequences onto internal structures, then the tautology in (14) predicts the principle in (19) which in turn, predicts the relative perceptual complexity of doubly embedded sentences, among others. In brief, in sequences of identical constructions, three's a crowd.

5. This suggests that a change in the superficial form of the noun phrases and verb phrase in double-embeddings should reduce their complexity. First, increasing the homogeneity does make the sentence harder, e.g. (a) is harder than (4). Furthermore (b) seems easier than (4), although it is still more complex than (6).

(a) The man the man the man left left left.

(b) The reporter everyone I met trusts says Thieu will resign in the Spring.

### Some Cases of Unacceptability of Uncertain Etiology

It is possible to argue that doubly embedded sentences are unacceptable due to their predictable perceptual difficulty, given our particular theory of sentence perception. If we had available a broad perceptual theory we could take any unacceptable sequence and analyze its complexity in order to determine whether its unacceptability is due to ungrammaticality or to perceptual difficulty. The problem for current research on language is that both our linguistic and perceptual theories are woefully inadequate. Consequently, much of the time, we may be trying to construct grammars in order to rule out constructions which could be discussed more appropriately by a perceptual theory, and vice versa.

Sometimes the decision about the basis for sequence acceptability is fairly straightforward. Consider (32) and (33):

- (32) The boat floated on the creek and sank.
- (33)? The boat floated on the creek sank.

Obviously (33) is not an acceptable version of (32) and at first appears to be completely ungrammatical. (Indeed when presented with these sentences out of context many linguists have assured me that (33) is completely ungrammatical.) But suppose (33) is considered in the context of (34) and (35).

- (34) The boat that was floated on the creek sank.
- (35) The boat put in the creek sank.

Suddenly one realizes that (33) is a version of (34) and that its construction is parallel to that of the entirely acceptable (35). The initial perceptual difficulty of (33) arises from the perceptual plausibility and dominance of the first six words as an independent sentence (e.g. "the boat floated on the creek . . ."). The same problem accounts for the initial unacceptability of (36) in which the first nine words comprise a plausible but misleading sentence (e.g. "one editor authors all the newspapers and the magazine"). To understand (36) notice that it has a compound subject and one verb, "agree":

- (36)? One editor authors all the newspapers and the magazine all agree.

Although the perceptual attractiveness of simple declarative sentences is clear on intuitive grounds, much of our research has been devoted to exploring the development and utilization in the adult of the perceptual strategy which maps any 'noun, verb, noun' sequence onto underlying subject-verb-object (Bever, 1970). Other studies have shown that perceptual segmentation occurs whenever there is a plausible underlying structure boundary reflected in the surface sequence (Bever, Lackner and Kirk, 1969). Thus the initial sequences in sentences like (33) and (36) appear to satisfy these perceptual schemata prematurely and make correct perception of the sentences more complex.

Not all cases of unacceptability due to perceptual difficulty are so easy to identify. Consider (37) which at first appears entirely ungrammatical. It would seem that the particle 'over' must be placed before the object noun phrase when that noun phrase is long, as in (38). Although it is clear that (38) is acceptable and (37) is not, it is entirely obscure to me whether (37) should be marked as ungrammatical, or as perceptually incomprehensible but grammatical.

(37)? I thought the request of the astronomer who was trying at the same time to count the constellations on his toes without taking his shoes off or looking at me over.

(38) I thought over the request of the astronomer who was trying at the same time to count the constellations on his toes without taking his shoes off or looking at me.

Consider the intermediate acceptability of (39):

(39)? I thought the request of the astronomer who was trying to count the constellations over.

(40) I thought over the request of the astronomer who was trying to count the constellations.

It is clearly not so acceptable as (40), but is (39) really ungrammatical? It is really grammatical?

There would appear to be a continuum of acceptability governed by the relative complexity of what intervenes between a verb and its particle. The general perceptual source for this restriction is a limit on immediate memory. In general, given two phrases following a verb, the less complex is ordered to come first (cf. Ross, 1967). I have argued elsewhere that placing the less complex phrase first allows less to be held in memory while the final phrase is being analysed (Bever, 1970). That is, in ordering phrases we save the hardest for the last.

However, the acceptability of such sentences is governed by other factors as well. Consider the unacceptability of (41) and (42) and the acceptability of (43) and (44):

(41)? I thought the request of the astronomer looking over the charts over.

(42)? I thought the request of the astronomer looking the charts over over.

(43) I thought over the request of the astronomer looking over the charts.

(44) I thought the request of the astronomer looking at the charts over.

Clearly it is not just complexity of the intervening sequence which determines whether a particle can be placed after the object noun phrase. Other restrictions have to do with apparent identity of adjacent phrases.

Another example of perceptual complexity arises in questions of superficial identity of phrases which appear to play a particularly baffling role in the assignment of pronominalization relations. For example in (45) there is no problem in understanding the way in which the pronouns 'her' and 'him' can cross-refer

(i.e., there is one interpretation in which 'her' refers to 'the girl' and 'him' refers to 'the boy'):

(45) The boy kissed her only after the girl kissed him.

It is easy to see how (45) is composed by combining the two sentences reflected in (46) and (47):

(46) The boy kissed the girl only after the girl kissed him.

(47) The boy kissed her only after the girl kissed the boy.

However, within nominalized complement constructions such cross-referring does not occur. For example (48) and (49) cannot be combined to produce (50):

(48) The boy suggested the girl's prediction only after the girl predicted it.

(49) The boy suggested it only after the girl predicted the boy's suggestion.

(50)? The boy suggested it only after the girl predicted it.

That is, (50) cannot mean that the boy suggested that the girl would predict his suggesting her prediction only after the girl predicted that the boy would suggest her prediction of his suggestion. (I am indebted to G. Lakoff for calling examples like (50) to my attention.)

Is the cross-referring interpretation of (50) ungrammatical? At first it would appear that it is, and that there is a structural restriction against cross-referring complements. However, consider (51), which is syntactically parallel to (45), but unacceptable on the same cross-referring interpretation that was perfectly acceptable in (45) (i.e., the first "it" cannot refer to "the boulder" while the second "it" refers to "the rock"):

(51)? The rock bounced on it after the boulder struck it.

Apparently the difficulty lies in the confusion introduced by having the two identical pronouns, "it" cross-referring. Similarly (52) is easy to understand and (53) approaches total gibberish on the cross-referring interpretation (e.g. the interpretation that would mean "the rock bounced on the boulder after the boulder struck it"):

(52) The boy kissed the girl after she kissed him.

(53)? The rock bounced on the boulder after it struck it.

We might try to explain such a restriction as another extension of (14), applied in this case to nominal reference (54):

(54) Superficially identical definite noun phrases in the same discourse corefer.

Suppose we view the perceptual operation of nominal reference as a direct mapping of the surface form of the nominal onto the specific referent (whatever the psychological form of such referents may be). Then it follows that if the

same superficial nominal has two different referents it will be in the position of having a double referential function. Thus (55)–(57) are unacceptable compared with (58)–(60) (notice, of course, that special contexts can be constructed for (55) (56) and (57) that make them more acceptable):

- (55)? Bill and Bill left.
- (56)? It and it fell.
- (57)? Bill hit Bill.
- (58) Bill and Mary left.
- (59) He and she left.
- (60) Bill hit Mary.

If the noun phrase is indefinite then these restrictions do not occur, presumably because such noun phrases do not map onto a unique specific referent; compare (61)–(63).

- (61) Somebody hit somebody.
- (62) A boy hit a boy.
- (63) Boys hit boys.

Even a slight change in apparent form allows (54) not to apply. Thus while (56) and (57) are unacceptable, (64a, b) are acceptable, presumably due to the difference in surface form of 'he' and 'him':

- (64a) He hit him.
- (64b) He and him left.

Principle (54) explains the unacceptability of the cross-referring interpretation of (51) and (53) since it requires that "it" always refer to the same referent. This explains another fact about the interpretation of (51) that I have not mentioned: in (51) not only do the "it"s not cross their reference, they also cannot refer to *two* distinct objects not mentioned within the sentence. Rather, in each case, they must both refer to the same *single* object not mentioned in the sentence. We can now ask whether the unacceptability of (50) is due to the structural requirement that complement constructions cannot cross-refer, or whether it is due to the perceptual principle (54). Principle (54) explains the unacceptability of (51) and the acceptability of (45), both of which are superficially similar to (50). Principle (54) also predicts that (50) is unacceptable. How do we decide whether (50) is unacceptable for structural reasons or perceptual reasons, or both?

Such quandaries can be invented *ad nauseam*. Surely the crucial question for linguistic theory is not whether we can confuse ourselves about the basis for the unacceptability of particular sequences, but whether we can be sure of the cases in which we think we are *not* confused. I have presented some examples of unacceptability due to perceptual complexity which I partially understand and for which I can justify the claim that their unacceptability is not necessarily due

to structural ungrammaticality. There are undoubtedly many other instances of unacceptable sequences which are taken today to be ungrammatical but which may have straightforward perceptual explanations when we come to understand more about perception.

The possibility that many judgments about grammaticality are judgments about perceptual difficulty does not mean that linguistic theory and formal description of intuitions about grammaticality are fruitless enterprises. It is simultaneously the greatest virtue and failing of linguistic theory that sequence acceptability judgments are used as the basic data. The fact that such judgments really are instances of behavior justifies the claim that linguistic theory is a serious psychological theory, not a mere intellectual artifact.

Although there are many properties of our intuitions about acceptability that we do not understand, certain intuitions appear to be stable enough to resist any obvious behavioral explanation. For example, the fact that for every transitive sentence with a direct object there is a corresponding passive sentence in English is not a fact which will someday be shown to be a by-product of our perceptual system. Nor are such facts as the unacceptability of sequences like (2) likely to be explained on trivial behavioral grounds. The purpose of this discussion is not to undercut one's confidence in linguistic theory, but to make one aware of its current empirical uncertainties and limitations.

## THE INFLUENCE OF PERCEPTUAL PROCESSES ON LINGUISTIC RULES

In certain cases the perceptual system appears to influence the structure of linguistic rules themselves rather than merely influencing our acceptability judgments about particular sequences. I shall discuss an example in phonology and an example in syntax of how certain basic perceptual principles influence linguistic structure. The mechanism for this structural influence clearly is to be found in the young child. As he learns the linguistic rules of his language he will tend not to learn rules which produce speech forms that are behaviorally hard to understand or hard to say. Thus, whatever aspects of cognition are utilized directly in speech perception will be reflected in certain properties of linguistic grammars.

### The Influence of Syllabic Structure on Phonological Rules

One of the most basic characteristics of speech perception is the organization of the incoming speech signal into units at different levels. For example, the acoustic continuum can be divided up into syllables, which are themselves composed



of individual speech sounds, 'phonemes'.<sup>6</sup> Accordingly, (65) can be analyzed as the syllable "dak" or as the sequence of three phonemes, 'd', 'a', and 'k':

(65) "dak"

Such a potential duality of description raises the behavioral question, which level of analysis is the "psychologically real" level and which is an artifact of linguistic theory or of the process of linguistic introspection? Most psychologists and many linguists have answered this question in favor of the syllable as the relevant psychological unit of speech production and perception.

The reasons for this are straightforward enough. The syllable can be defined in acoustic terms relatively easily, and it is the smallest unit that can be uttered in isolation (although vowel phonemes can be uttered in isolation, most consonants cannot). Furthermore, it has been shown by various techniques that the acoustic shape of many individual phonemes, both vowels and consonants, is influenced by the surrounding phonemes. That is, there is no acoustically definable invariant that corresponds uniquely to each phoneme. Not only does each phoneme represent a class of sounds acoustically, it is often the case that the acoustic representation of a phoneme in one context overlaps with the acoustic representation of a different phoneme in a different context. If the first step in speech perception were to identify the incoming speech signal, phoneme by phoneme, we would often not know what many of the phonemes were until we had analyzed the local context in which each phoneme occurs. Many researchers have argued that the natural and relevant definition of 'local context' in which to recognize particular phonemes is the syllable. The syllable is relatively invariant, and includes the information about consonant-vowel transitions that determines much of what we perceive.

Most of the studies that underlie the conclusion that the syllable is perceptually primary result from elegant experimental failures that attempted to account for speech perception in terms of sequential isolation of individual phonemes (Liberman et al., 1967). One can demonstrate that it is impossible to identify many individual phonemes out of syllabic context, while whole syllables are identified with relative ease. Similarly, one can demonstrate that there is no unique definition of the phoneme in terms of articulatory movements of the tongue and lips. Just as the perception of individual phonemes is influenced radically by their context, there are effects of equal magnitude on the articulatory movements involved in the production of a phoneme in a particular context.

Of course, the experimental and observational failure to find direct behavioral evidence for the phoneme as a valid perceptual entity does not explicate its obvious existence, its relation to syllables, nor the kind of role that it might

6. In this discussion I leave open the question of how abstract phonemes are for the same sort of reasons mentioned in footnote 2.

play in perception. To explore this kind of question, Harris Savin and I presented subjects with nonsense lists presented at the rate of one syllable/second, of varying length, like those in (66) (Savin and Bever, 1970):

(66) skæm firt mowf kiyg gib nemp *bolf* days saymz

On some trials subjects were told to respond (on a telegraph key) as soon as they heard the first syllable beginning in a /b/ sound; on other trials the subjects were told to respond the first time they heard the particular syllable, which was presented in its entirety ahead of time.

We found that subjects respond about 70 msec faster when they are told the entire syllable than when they are looking for the first syllable beginning in /b/. This might be due to the fact that there are several acoustic versions of /b/, depending on the following context: subjects who were told the exact syllable knew exactly which acoustic variant of /b/ to listen for and so could respond faster. However, we found the same difference in responding to initial /s/-syllables (in sequences like (67)):

(67) kæmf firt mowf kiyg gib nemp *solf* dayt baym

when subjects were told either the entire syllable or just that the target syllable would begin in /s/. This difference is critical, since the acoustic properties of the initial /s/ are only slightly affected by the following context, if at all. Furthermore, there is a recognizable acoustic invariant associated with every /s/ regardless of its context, a burst of high-frequency noise.

Another possible explanation for our results might be that subjects made very slow responses on just a few trials to initial phoneme targets but showed no difference the rest of the time. Therefore we analyzed the subjects' fastest quartile of responses to syllable targets. These, too, were faster than the fastest quartile of responses to phoneme targets, so the difference was not due to only a few slow responses. Furthermore, while subjects' mean reaction time varied from 200 to 450 msec, there was no related difference in the relative delay in reacting to phoneme targets. Thus, consciously to discriminate and respond to the phoneme would appear to take a certain amount of time beyond the time it takes to respond to the syllable. In this sense we can argue that phonemes are perceptually subsidiary to the syllable.

Are phonemes then dispensable as psychological constructs internal to language, since neither the perceptual nor the articulatory facts require them? Not at all. It is quite impossible to do without phonemes in psychological theories of language, but for nonsensory and nonarticulatory reasons. To mention only a few aspects of the behavioral evidence for phonemes, there is the occurrence of alphabetic writing systems, of rhyme and alliteration in non-literate poetry, of segmental phonemic spoonerisms, and of the innumerable well-attested historical changes in language that are described very simply in terms of phonemes and only clumsily and arbitrarily without them.

In addition to evidence from such historical facts (which would hardly be facts but for some psychological facts about the people who make the history), there are numerous regularities in every modern language that can be stated satisfactorily only by referring to phonemic segments. Consider, for example, just part of the rule of plural-formation for modern English nouns (68):

- (68) *If the (singular) noun ends in a voiced sound, add /z/ (boys) and if it ends in an unvoiced sound add /s/ (bits).*

To state such a rule in terms of unsegmented syllables would be a great deal more complex since all the syllables of each kind would have to be listed. Not only would it be inelegant, it would not represent the generalization that underlies the groupings of the three different kinds of syllables. Segmental phonemic features turn out to be just the appropriate concepts for such phenomena (cf. Chomsky and Halle, 1968), neither too specific nor too undifferentiated to describe regularities like this. Again, how could this be if those phonemic features were not part of a psychologically correct description of what people intuitively know about the sound structure of their language?

The conclusion that we draw from such considerations is that phonemes are neither perceptual nor articulatory entities. Rather, they are psychological entities of a nonsensory, nonmotor kind, related by complex rules to stimuli and to articulatory movements; but they are not a unique part of either system of directly observable speech processes. In short, phonemes are behaviorally *abstract*. Just by virtue of standing neutrally between the behavioral systems of sensory input and articulatory output, they can interrelate these perceptual and expressive speech processes.

Since it appears that the initial and primary step in speech perception is to isolate the syllables, we must account for the appearance of phonemes in language learning. One description can be based on work at Haskins' laboratory claiming that adults have a stored "library" of syllable templates. Each time speech stimulus is heard it is matched against the library to determine which syllable it is.<sup>7</sup> Presumably the young child acquires such a library as part of his initial language learning. If this is true, one of the first steps in acquiring the phonemic system which allows him to manipulate and subsequently represent the acoustic regularities in his language will be to acquire the acoustic regularities that characterize syllable structure. For example, in English, the child should acquire restrictions like (69) at an early age:

- (69) Two sibilants do not occur in a row in one syllable.

Such a restriction would enable the child to recognize that (70) has three syllables, that (71) is not English, and so on:

- (70) churchsteep  
(71)\* churchs

7. Notice that it is not necessary to specify if the way in which the library is deployed is analysis-by-synthesis or direct analysis.

If acquisition of the acoustic regularities of syllables is one of the first steps in the child's acquisition of the phonemic system in his native language, we might expect that the syllabic properties of the language would influence the regularities that relate individual phonemic segments. For example, suppose the plural rule in (68) were applied to singular words that end in sibilants, like those in (72). The results would be incorrect forms like those in (73), which violate restriction (69):

- (72) church, ax, bush, fudge  
(73)\* churchs, axs, bushs, fudgz

Rule (68) must be modified so that it will not form syllables that violate the restriction against two sibilants in a row. The plural rule as stated in (74a, b) captures the regularities offered by the use of segmental phonemes with internal features and a system of ordered rules:

- (74a) *If the singular noun ends in a sibilant add the neutral vowel, 'uh'.*  
(74b) *Add /s/ to all forms, assimilating the voicing to the preceding sound (giving 'z' following voiced consonants and vowels, including the vowel introduced by (a)).*

But it does not reveal within the linguistic description any hint that (74a) is, in fact, an accommodation to the syllabic structure of English that would otherwise be violated by (74b).

Of course, one could argue that syllabic restrictions on the phonemic rules can be stated as part of the universal form of phonological grammar. On this view all phonological rules and systems of rules are evaluated against a paradigmatic statement of the syllable structure of that language. Those rule sequences that produce impossible syllable sequences are modified appropriately. In this way one could make it appear that syllabic constraints on phonological rules operate *within* linguistic structure, rather than being the result of the interaction of systems of speech perception and production with the system of linguistic structure. Such a move would formally save the descriptive linguist the embarrassment of having to turn to nonstructural aspects of speech behavior. But the saving would be spurious. Including syllabic constraints within phonological theory is merely a way of stating that there is a descriptive problem of accounting for certain kinds of constraints on segmental sequences. The solution of the problem lies beyond the traditional domain of linguistic structure itself.

#### The Influence of Sentence Units on Syntactic Rules

At a higher level of analysis one must decide whether the sentence is the basic level of perceptual processing or whether we listen to and organize speech initially as a series of isolated words. At first it would appear obvious that sentences are composed of words and phrases perceptually—that is, as we hear the individual words and phrases we organize them into sentential units. Indeed

some of the experiments reviewed above demonstrate the effects involved in the act of organizing a sequence of phrases into a sentence. But the question remains whether we hear sentences *before* the constituent words. For example, a classic experiment is baffling on this view of perception. If sentences are played in noise at certain levels, probable sentences (75) are perceivable but improbable sentences are not (76):

- (75) The mother patted the dog.  
 (76) The dog patted the mother.

How can this be? Even more baffling is the fact that if the noise level behind a sentence like (75) is set so the words just cannot be perceived in isolation, they can be perceived when arranged in a sentence order. Somehow the bits and pieces of words that we hear can contribute mutual information when they are placed in a sentence. The striking aspect of this phenomenon is that as we hear sentences presented in noise we do not perceive that we hear bits and pieces. Rather, as the sentence organization brings the series of words into consciousness the perception of the individual sounds is simultaneous with it.

Of course, the technique of presenting sentences in noise or under other forms of distortion leaves a great deal to be desired and may introduce special properties of its own. To look at this in a different way, Savin, Hurtig and I are examining subjects' reaction times to target sentences beginning with the word "boys" in sequences like (77) (spoken at 1 sentence/second):

- (77) Monks ring chimes, cows give milk, shoes help feet, plants have seeds,  
*boys like girls, . . .*

Before each trial, subjects are told the entire target sentence or told to respond to the first sentence beginning in "boys." Our results so far show that subjects respond consistently faster when they know the entire sentence target than the initial word target (the word 'boys' is always the initial word in the target). This difference is consistent even comparing the fastest quartile of responses to sentence and word targets for each subject (all these responses are *completed* before the end of the word "boys"). If these results hold up, and are not due to some artifact, they will demonstrate that conscious perception of a whole sentence is more direct than perception of the first word in the sentence (Bever, Savin and Hurtig, in preparation).

What then of the word? Is it a mere artifact of linguistic analysis? Obviously not. The kinds of evidence for the behavioral reality of the word are too numerous and obvious to list. But its reality may be analogous to that of the phoneme—words serve as the units with abstract properties that are intermediate between sentence production and perception. The word serves as the reflection of a point of intersection between acoustic, semantic, and syntactic structure.

Accordingly, many syntactic transformational rules are stated on constituents, ranging from clauses to individual words. For example, consider the

sentences in (78), which exemplify the restrictions on the optional deletion of the complementizer markers "that" and "the fact":

- (78a) John mentioned the fact that Sam is a fool.  
 (78b) John mentioned the fact Sam is a fool.  
 (78c) John mentioned that Sam is a fool.  
 (78d) John mentioned Sam is a fool.

It would appear that these words can be freely deleted, as represented by the syntactic rule (79):

- (79) *Delete complementizers that, the fact freely.*

As the young child learns rules like these we can expect that he will bring to bear his perceptual strategies based on the sentence as the main unit of perception. We have found that at first (age 1½–2 years) young children repeat the first plausible sentence they hear; for example they repeat "the elephant jumped" in response to (80):

- (80) The elephant that jumped kissed the cow.

That is, the young child takes the first noun-verb sequence as the main clause of a complex sentence, even in cases like (80) where that is inappropriate. Certain facts of adult English syntactic structure appear to accommodate this perceptual strategy: for example, R. Kirk (personal communication) has observed that a subordinate clause verb which precedes its main verb is generally marked as subordinate by the end of its verb phrase. The most obvious device is the subordinate clause conjunction, as in (81a, b):

- (81a) *Although* the research was secret the liberated files revealed that it actually concerned the metabolization of sauce Bearnaise.  
 (81b) *Because* the demands were non-negotiable nobody wanted any.

In each case the first verb is marked by the subordinate conjunction as subordinate.

Subordinate conjunctions are specific lexical items which act as specific markers of those cases in which a subordinate clause occurs before its main clause. There are also certain syntactic rule systems that appear to have formed in response to the principle that initial subordinate clauses are marked in English: restrictions on the 'syntactically' allowed deletion of words which mark functional relations among clauses. The sentences in (82)–(84) exemplify a heterogeneous set of grammatical restrictions on the stylistic deletion of 'that' or 'the fact' in initial position.

- (82) \*Sam was a fool was mentioned by John.  
 (83) That Sam was a fool was feared by John.  
 (84) The fact Sam was a fool was feared by John.

Rule (79) is incorrect as shown by the incorrect sentence (82) in which both complementizers are deleted. However, one of the complementizers can be deleted in initial position, *so long as the other remains*. Thus rule (79) must be reformulated to (85):

- (85) *Delete the complementizers, except in sentence initial position, in which case at least one complementizer must remain.*

Notice that (85) also applies to allow deletion of the complementizers so long as the complement clause is marked by some lexical item (e.g., "the discovery . . ." in (87)).<sup>8</sup>

- (86) The discovery that Sam was a fool frightened John.

- (87) The discovery Sam was a fool frightened John.

Thus as the child acquires the rules of syntax he applies global restrictions about sentences to the properties of sentences. One might argue that the fact that such restrictions are reflected in particular syntactic rules is sufficient evidence that the restrictions be included as a proper part of linguistic theory. For example, we might propose a universally available syntactic restriction that initial subordinate clauses always be uniquely marked as such. In a language which has such a restriction any set of syntactic rules that would produce constructions violating that constraint are modified appropriately. Of course, this would merely represent within the linguistic theory a structural property while leaving opaque the behavioral explanation for the existence of that property.

## CONCLUSION

I have argued that there are two forms of interaction between speech behavior and linguistic description. First, certain intuitions about the unacceptability of possible sequences are due not to structural facts of the language, but to facts that make them hard to process behaviorally. Linguistic investigations sometimes may attempt to explain on structural grounds the unacceptability of sentences that have a more direct explanation on perceptual grounds. Second, some structures in adult language appear to be accommodations to behavioral strategies in the child's use of language. Consequently, certain structural properties of the grammar of a language can be attributed, not to the form of the child's (innate) linguistic structure, but to the interaction between the process of language

8. A similarly motivated restriction exists on the deletion of relative clause-introducing relative pronouns on initial nouns. Thus (b) is not a grammatical version of (a):  
(a) The boy who is a fool frightened Sam.  
(b) \*The boy is a fool frightened Sam.

This, too, is presumably caused by the fact that in (b) the first sequence appears incorrectly as a main clause, cf. Bever and Langendoen (1971) for a discussion of this phenomenon in the history of English.

learning and behavioral strategies used by young children to process actual sentences.

Such facts must modify the dictum that a "correct" grammatical description of a language is at the "center" of all language behavior. A more accurate statement would be that language in its entirety is a function of an interaction between different systems of speech behavior. In the adult, different systems contribute to intuitions about particular utterances, and in the child different systems interact to determine the final form of the language in the adult. In this paper I have concentrated on the effects of the perceptual system on the structural system of language. Of course, both of these systems interact with systems of language learning, speech production, and so on. A fuller understanding of the complexities of language behavior will require study of the interactions among all the psychological systems that are recruited by our instinct to communicate.

## REFERENCES

- Bever, T. G. 1968. "A Survey of Some Recent Work in Psycholinguistics." *Specification and Utilization of a Transformational Grammar*, In W. J. Plath ed. Scientific Report No. 3, Air Force Cambridge Research Laboratories, Cambridge, Mass.
- . 1970. "The Cognitive Basis for Linguistic Structures." In *Cognition and the Development of Language*, J. R. Hayes ed. Wiley.
- Bever, T. G., J. R. Lackner, and R. Kirk. 1969. "The Underlying Structures of Sentences Are the Primary Units of Immediate Speech Processing." *Perception and Psychophysics* 5: 225-234.
- Bever, T. G., and D. T. Langendoen. 1971. "A Dynamic Model of the Evolution of Language." Reprinted in this volume.
- Bever, T. G., H. Savin, and R. Hurtig. The Non-Perceptual Reality of the Word. In preparation.
- Blumenthal, A. L. 1967. "Prompted Recall of Sentences." *Journal of Verbal Learning and Verbal Behavior* 6: 674-676.
- Chomsky, N. and M. Halle. 1968. *The Sound Pattern of English*. Harper and Row.
- Chomsky, N. and G. A. Miller. 1963. "Introduction to the Formal Analysis of Natural Languages." In R. D. Luce, R. R. Bush and E. Galanter eds. *Handbook of Mathematical Psychology*, Vol. 2, Wiley.
- Fodor, J. A., T. G. Bever, and M. Garrett. 1968. "The Development of Psychological Models for Speech Recognition." Report on Contract No. AF 19 (628)-5705, M.I.T.
- Fodor, J. A. and M. Garrett. 1967. "Some Syntactic Determinants of Sentential Complexity." *Perception and Psychophysics* 2: 289-296.

- Halle, M. and K. N. Stevens. 1964. "Speech Recognition: A Model and a Program for Research." In *The Structure of Language*, J. A. Fodor and J. J. Katz, eds. Prentice-Hall.
- Herzberger, H. 1966. "Perceptual Complexity in Language." Unpublished paper.
- Lakoff, G. 1971. "On Generative Semantics." In *Semantics*, D. Steinberg and L. Jakobovits, eds. Cambridge University Press.
- Liberman, A. M., F. S. Cooper, D. P. Shankweiler, and M. Studdert-Kennedy. 1967. "Perception of the Speech Code." *Psychological Review* 74: 431-461.
- Miller, G. A. and J. A. Selfridge. 1950. "Verbal Context and the Recall of Meaningful Material." *American Journal of Psychology* 63: 176-185.
- Perlmutter, D. 1971. *Deep and Surface Structure Constraints in Syntax*. Holt, Rinehart, and Winston.
- Phillips, J. R., and G. A. Miller. 1966. "An Experimental Method to Investigate Sentence Comprehension." Unpublished paper.
- Postal, P. 1972. "The Best Theory." In *Goals of Linguistic Theory*, S. Peters ed. Prentice-Hall.
- Ross, J. R. 1967. *Constraints on Variables in Syntax*. Indiana University Linguistics Club.
- Savin, H., and T. G. Bever. 1970. "The Non-Perceptual Reality of the Phoneme." *Journal of Verbal Learning and Verbal Behavior* 9: 295-302.