Testing Hypotheses Regarding the Psychological Manifestation of Morpheme Structure Constraints *

JOHN J. OHALA
MANJARI OHALA

INTRODUCTION

Linguists have for many decades recognized that not all possible sequences or types of sounds occur in the morphemes of any single language. These restrictions were called 'phonotactic' or 'distributional' constraints by the classical structuralists and 'morpheme structure' rules or conditions by generative phonologists. Although these constraints may originally have been conceived of as merely descriptive statements one could make about the sound pattern of the language, there is sufficient evidence now that native speakers are also aware of them, that is, that many of the constraints noted by linguists are psychologically real (Esper 1925; Greenberg & Jenkins 1964; Zimmer 1969; Fromkin 1971; Ohala 1972, 1983; Wright 1975).

It is of interest, then, to try to find out how these morpheme structure constraints (henceforth MSCs) are represented in speakers' heads and how they use them. This chapter reports some attempts to do this.

*This research was supported in part by the National Science Foundation and by the Committee on Research, University of California, Berkeley.

PRIOR HYPOTHESES

There have been various proposals about the nature of MSCs, although often they have not been perfectly explicit. Thus in reviewing some of these claims we may inadvertently read into them certain details that were not intended. Also it is not always clear how much psychological reality various authors meant their proposals to have. In any case, our primary purpose is not to do battle with straw men but to attempt to show how it is possible to pit one reasonably explicit and superficially plausible hypothesis regarding MSCs against another and in that way to determine which has more empirical support as a model of how native speakers react to the phonetic shape of words.

Greenberg and Jenkins

Greenberg and Jenkins (1964) (hereafter G&J) offer one of the clearest statements regarding the psychological manifestation of MSCs; we call it the phoneme-substitution procedure. According to them, the degree of closeness of a given 'word' to the native pattern is inversely proportional to the number of zero-, one-, two-, and up to *n*-phoneme substitutions for the *n* original phonemes in the word which resulted in existing words in the language, as given in Table 13.1.

To find the distance from English of /klæb/, /klɛb/, and /zbyk/ using their phoneme substitution procedure, we first note that a zero phoneme substitution would not yield an existing word, as this would happen only if the word we start with were already an existing word, which none of these are. We next try changing single phonemes. If we change the first phoneme we can get /slæb/ from /klæb/ but cannot derive any existing words from /klɛb/ or /zbyk/.[1] Changing the second of the four phonemes also succeeds with /klæb/ but not the other two forms; and so on, for two, three, and finally four simultaneous phoneme substitutions, the latter type, of course, yielding an existing word for any sequence of four sounds. We then total the number of substitutions which give an existing word and subtract that from the number of possible substitutions plus one; this is 17 in the case of sequences of four phonemes. Thus /klæb/, /klɛb/, and /zbyk/ are predicted to be respectively 2, 5, and 15 units away from English on a 16-point scale. The

[1]The Oxford English Dictionary lists *bleb*, defining it as a smaller variety of *blob* and, in fact, this word is widely used today in microbiology. In all likelihood, though, this word is not in the lexicons of most speakers of English and so would not decrease the predicted difference between /klæb/ and /klɛb/. We made similar decisions about other potential phoneme substitutions for the words in Table 13.1 and others used in the experiments described later on, but we do not discuss them in detail here.

Table 13.1
Example of Greenberg and Jenkins' Phoneme-Substitution Method

	Word		
Type of substitution	/klæb/	/klɛb/	/zbyk/
CCVC	—	—	—
$\underline{C}CVC$	*slab*	—	—
$C\underline{C}VC$	*crab*	—	—
$CC\underline{V}C$	*club*	*club*	—
$CCV\underline{C}$	*clam*	*Clem*	—
$\underline{CC}VC$	*stab*	—	
$C\underline{CV}C$	*crib*	*crib*	—
$CC\underline{VC}$	*clip*	*clip*	—
$\underline{C}C\underline{V}C$	*slob*	*slob*	—
$\underline{C}CV\underline{C}$	*slam*	*phlegm*	—
$C\underline{C}V\underline{C}$	*cram*	*Kress*	—
$\underline{CCV}C$	*grub*	*grub*	*creek*
$C\underline{CVC}$	*cream*	*cream*	—
$\underline{CC}V\underline{C}$	*spat*	*bread*	—
$\underline{C}C\underline{VC}$	*flip*	*flip*	—
$\underline{CCVC}$	*gruff*	*gruff*	*gruff*
Total successful substitutions	15	12	2
(17 – total) = "distance" from English	2	5	15

only substitutions G&J allow in this procedure are consonant substitutions for consonants and vowels for vowels. They rule out vowels substituted for consonants and vice versa, and they rule out phoneme additions and deletions.

Early Generative Phonology

Proposals regarding MSCs in early generative phonology (hereafter EGP) (Halle 1959, 1962; Stanley 1967) seem to include the following claims:

1. The function of MSCs is to differentiate between existing or admissible phoneme sequences and inadmissible sequences, that is, a binary classification; or, under some interpretation, a ternary classification between existing words, accidental gaps, and inadmissible forms.
2. MSCs are stated in terms of features, not phonemes.
3. MSCs are separate from the lexical entries themselves and are presumably relatively small in number in comparison to the lexical entries. That is, they capture only rather broad generalizations, such as that in English a nasal consonant may be the second member of an initial cluster only if /s/ is the first, but they would not include such specific constraints as that /ɛ/ may not appear after the initial cluster /kl/ if /b/ follows it.

The Sound Pattern of English

In *The Sound Pattern of English* Chomsky and Halle (1968:417) (hereafter C&H) offer a rather explicit proposal regarding the form and function of MSCs. They claim that the degree of closeness to the native pattern of a given form is inversely proportional to the complexity of the rule—measured by counting the features mentioned in it—which would differentiate that form from all the entries in the regular (existing) lexicon. That is, one finds the simplest rule which would not change any entry in the regular lexicon but which would change an irregular or nonexisting form. One then counts the number of features utilized in this rule (which presumably need not incorporate any features which can be specified by universal marking conventions) and then takes the distance from the native pattern to be the reciprocal of this number. Note that the rule which should change a nonexisting form but no regular, existing form does not, unlike the G&J phoneme-substitution procedure, have to make an existing form out of the nonexisting form to which it applies. Some examples are given in Table 13.2. (a) Gives the rule offered by C&H as being the simplest one which differentiates /blɪk/ from all entries in the English lexicon. Since this form is at best an accidental gap in the English lexicon, the rule must be very specific and thus very complicated if it is to affect /blɪk/ but no regular English morpheme. The rule takes 17 features to state and is thus 1/17 or .059 units away from English on a scale that presumably runs from zero to .5. (b) Gives the much simpler rule which differentiates /bnɪk/ from the rest of the English lexicon, and (c) and (d) give the rules which we offer as differentiating the made-up forms /xlox/ and /xrit/, and /mløf/ and /spøf/, respectively, from the regular forms in English. Since these forms violate more fundamental English MSCs, fewer features are needed to state the rules which apply to them but to no regular English words.

THE CONFLICT BETWEEN THE HYPOTHESES

There are a variety of quite opposite claims made by these models.

Both the G&J and the C&H models predict that speakers are capable of making scalar judgments on the degree to which a given form adheres to the pattern of their native language. The EGP model, however, predicts that speakers could make only ternary distinctions. G&J's experiments proved that subjects could make scalar judgments.[2]

G&J posit that, in making judgments about the well-formedness of spoken

[2]This might account for some results which seemed anomalous in a study by Ohala (1972, 1983). She asked Hindi-speaking subjects to categorize normal and rare Hindi words and made-up words as 'existing', 'possible but nonexisting', and 'not possible' for Hindi. In some cases they unexpectedly assigned what the linguist would consider 'accidental gaps' to the category 'not possible'. This may have happened because the three-way classification was too restrictive; more reasonable responses may have been obtained if they had been allowed to make scalar judgments.

Table 13.2

Examples of Chomsky and Halle's Algorithm for Measuring Degree of Distance from the Regular Lexicon[a]

(a) Minimal rule which differentiates /blIk/ from English lexicon

$$[+\text{cons}] \rightarrow [-\text{lateral}] \,/\, [-\text{seg}] \begin{bmatrix} +\text{cons} \\ -\text{voc} \\ +\text{ant} \\ -\text{cor} \\ +\text{voice} \end{bmatrix} \underline{\quad} \begin{bmatrix} +\text{high} \\ -\text{back} \\ -\text{tense} \end{bmatrix} \begin{bmatrix} +\text{cons} \\ -\text{voc} \\ -\text{ant} \\ -\text{cor} \\ -\text{cont} \\ -\text{voice} \end{bmatrix}$$

Number of features required: 17
Distance from English: 1/17 = .059 (min. = 0; max. = .5)

(b) Minimal rule which differentiates /bnIk/ from English lexicon

$$[+\text{cons}] \rightarrow [-\text{nasal}] \,/\, [-\text{seg}] \begin{bmatrix} +\text{cons} \\ -\text{cont} \end{bmatrix} \underline{\quad}$$

Number of features required: 5
Distance from English: 1/5 = .2

(c) Minimal rule which differentiates /xlox/ and /xrit/ from English lexicon

$$\begin{bmatrix} +\text{cons} \\ +\text{cont} \end{bmatrix} \rightarrow [-\text{back}]$$

Number of features required: 3
Distance from English:1/3 = .33

(d) Minimal rule which differentiates /mløf/ and /spøf/ from English lexicon

$$\begin{bmatrix} +\text{syllabic} \\ -\text{back} \end{bmatrix} \rightarrow [-\text{round}]$$

Number of features required: 3
Distance from English: 1/3 = .33

[a]Chomsky & Halle, 1968:417.

forms, speakers refer to the existing words in the lexicon; except for the phoneme-substitution procedure itself, they need not store lexicon independent rules. The EGP model seems to posit that speakers refer to independent rules, the MSCs.

Related to these two opposing details is that mentioned earlier, that is, whether speakers can detect deviations from only the most general patterns of the language or whether they can notice even the specific deviations. The EGP model would allow only the broad generalizations to be reflected in MSCs, whereas the G&J model, since it has the input forms compared to the entire list of existing morphemes, would permit deviations from even the most specific patterns to be detected. It was this difference between the G&J and the EGP model that we sought to test first. (Further conflicting predictions made by these models are addressed later.)

EXPERIMENT 1: G&J VERSUS EGP

To test these two models we constructed pairs of forms for which each model would give different predictions about the relative distance of each member of the pair from the English pattern. Table 13.3 gives five such word pairs. In the case of the first pair, /klɛb/ and /klæb/, we have already explained why the G&J model predicts the latter to be farther from English than the former (see Table 13.1). Presumably the EGP model would predict that they are both equally far from English since the rules that would differentiate them would be too specific (i.e., not general enough) to qualify as MSCs. The four word pairs /θlɛz/, /θlɛd/; /ʃriz/, /ʃrɪd/; /flUt/, /flɪg/; /sθɛʃ/, /sθip/ would pattern like /klɛb/, /klæb/ as far as the two models' predictions are concerned, as indicated in the second and third columns of Table 13.3.

We then obtained native English speakers' subjective judgments on the 'distance' from the English pattern of these forms (and other filler items selected to range over the entire 16-point G&J scale). We did this in essentially the same way as G&J. The individual forms (i.e., not in pairs) were recorded using a high-quality system by a phonetically trained speaker (the first author) who spoke them in two different randomized orders. The tape was played to individual subjects (either over a loudspeaker in a sound-treated room or over high quality earphones), and they were asked to judge the "distance" of each form from English using an 11-point scale and to indicate their responses on an answer sheet. (G&J in their study used both an 11-point scale and the free-magnitude-estimation technique and found both of them to give the same results. We used an 11-point scale since it was simpler to explain to the subjects how to use it.) Instructions to the subjects, given ad lib, went approximately as follows:

> You will hear a list of words. Some will be English words, some have been obtained from languages close to English and some from languages quite unlike English. We want you to listen to each word and judge by how it sounds how far it is from the pattern of English words, using an 11-point scale. If it is an English word check '1' off on the answer sheet; if it is very far away from English mark '11' on the answer sheet. The closer to English you judge the word to be, the smaller should be the number you mark; the farther away the word is from the English, the larger should be the number you mark. Please try to use the entire 11-point scale.

Then a minute or so of the tape was played to familiarize the subjects with the format of the tape and to let them hear the type of words used. We answered any questions they had, and then the tape was started again from the beginning and the subjects gave their judgments on the words. Before the words were played to them in the second randomized order, the subjects

Table 13.3
Stimuli, Predictions, and Results of Experiment 1

Stimuli		Predictions		Results	
A	B	Greenberg & Jenkins	Early generative phonology	Order	Mean difference
klɛb	klæb	A > B	A = B	A > B	0.276
θlɛz	θlɛd	A > B	A = B	A > B	1.557
ʃriz	ʃrɪd	A > B	A = B	B > A	0.107
flUt	flɪg	A > B	A = B	B > A	0.550
sθɛʃ	sθip	A > B	A = B	A > B	1.483
Mean					0.532*

*$p < .001$.

were told that since they now had a better idea of the total range of variation of the word types they could, on this second run, change their ratings of the words if they liked.

The subjects were adult speakers of English and were nonlinguists. We obtained judgments from 21 subjects.

Some subjects failed to use the entire 11-point scale. To make their responses comparable to the others' their ratings were normalized, that is, the portion of the scale they did use was stretched linearly so that the extreme scores equaled 1 and 11.

Although given a chance to change their scores on the second run of the stimuli, our subjects, like those of G&J, gave ratings that were statistically indistinguishable on both runs. (Although this was not our main concern, it turned out that, in agreement with the original results reported by G&J, there was a strong positive correlation for the entire body of stimuli between the ratings obtained by the G&J algorithm and our subjects' ratings.)

In tabulating the results, it was the *difference* between the ratings given to the two words of a pair on any single run (i.e., a given randomized order) that was counted. For example, if one subject gave word A a rating of 5 and word B a rating of 7, then this would be counted the same as another subject who gave the ratings 2 and 4, respectively, to the same words, since in both cases their ratings of the two words differed by +2. As we were just interested in whether or not one member of the pair was judged farther from English than the other, a one-tailed *t*-test was performed to assess the significance of the difference, that is, whether the mean difference in scores for all the pairs of interest was significantly distinct from zero.

The results are given in the rightmost columns of Table 13.3. In two of the five pairs, namely /ʃriz/, /ʃrɪd/ and /flʊt/, /flɪg/, the difference in scores was small and in the opposite direction of that predicted by G&J. In the other three pairs the difference was in the direction they predicted. The average difference for all five pairs was .532, which is highly significant ($p < .001$). It may be safe to conclude, then, that deviations from rather narrow, quite particular, patterns in the language contribute to the perceived distance of a word from the native pattern, as G&J predict.[3]

EXPERIMENT 2: G&J VERSUS C&H

Another point of difference between the models reviewed above is whether subjects' decision on the degree of deviation from the native pattern is based on only *part* of the form or the *whole* form. The C&H model, especially in

[3]Ohala (1972, 1983) found that her subjects not only gave evidence of knowing the MSCs of Hindi but in some cases also the MSCs appropriate to different parts of the Hindi lexicon, for example, the consonant clusters appropriate to English loanwords or words of Urdu origin.

the case of words violating very fundamental patterns, allows only the most serious violation, and then only one instance of it, to influence speakers' reaction to a form. For example, as given in 13.2d, even though there are two violations of English MSCs in /mløf/, namely, a bad initial cluster and a non-English segment /ø/, as opposed to only one violation in /spøf/, namely, the same non-English segment, according to the C&H model, both are equally far from the English pattern since they are both differentiated from existing English words by the same minimal rule, the one given in 13.2d. The G&J model, on the other hand, predicts that all parts of the word affect its judged deviation from the regular lexicon.

We used the same method as in Experiment 1 to test which of these two models better predicts subjects' behavior. The four word pairs used are given in Table 13.4 along with the predictions of how the members of each pair should be regarded by native speakers. In the case of the pair /sfʊb/, /sfɛt/ we assume, perhaps wrongly, that C&H would consider the initial /sf/ cluster as a deviation from the regular English pattern, in spite of the exceptional words *sphere*, *sphinx*, *sphincter*, and *sforzando*. In this experiment there were 11 subjects.

The results are given in the rightmost columns of Table 13.4. In all four word pairs, the difference in ratings is in the direction predicted by G&J and in three of these cases the difference is very large. The mean difference for the four pairs is 1.914, which is highly significant ($p << .001$). This seems to establish fairly conclusively that the whole word and not just part contributes to speakers' evaluation of it.

EXPERIMENT 3: JUDGMENTS ON PAIRED STIMULI

Although both Experiment 1 and Experiment 2 show highly significant results which enable us to evaluate competing hypotheses, the fact that two of the five pairs in Experiment 1 show a difference opposite to that of the group was disturbing. This could have arisen due to any one or a combination of several reasons. First, the G&J model could be inappropriate in very specific cases. Second, the experimental method itself could be deficient in that it permits a great deal of "noise," from, among other things, drift in subjects' frame of reference when making their judgments. To see whether distortions due to this second factor might be eliminated, we conducted a third experiment.

Since we were really interested in the difference, if any, which subjects perceived in the members of each pair with regard to their adherence to the native pattern, in Experiment 3 we presented the words in pairs and asked subjects to judge which member of each pair was closer to the native English pattern. In other respects, the methods used, for example, presentation over

Table 13.4
Stimuli, Predictions, and Results of Experiment 2

Stimuli		Predictions		Results	
A	B	Greenberg & Jenkins	Chomsky & Halle	Order	Difference
mlɔʒ	mlit	A > B	A = B	A > B	0.514
mløf	spøf	A > B	A = B	A > B	3.014
xlox	xrit	A > B	A = B	A > B	2.209
sfʊb	sfɛt	A > B	A = B	A > B	2.018
Mean					1.914*

*$p < .001$.

tape, the background of the subjects, and the words used, were the same as in Experiments 1 and 2. Sixteen subjects participated. The instructions given, ad lib, were approximately as follows:

> You will hear a list of pairs of words. Some will be from languages close to English and some from languages quite unlike English. We want you to listen to each pair and, based on how they sound, judge which member of each pair is closer to English and indicate your judgment on the answer sheet.

This method—obtaining judgments of paired items—should be a more accurate and sensitive way of getting the kind of data needed. Analogously, it is easier to judge which of two individuals is taller if they stand back to back than it is to judge the absolute height of individuals considered in isolation. (This was the method used by Zimmer in his 1969 study on Turkish MSCs.)

The results are presented in Tables 13.5 and 13.6. The trend of the data is similar to that from Experiments 1 and 2 except that this time all pairs show majority judgments in the direction predicted by the G&J model. The differences are again highly significant.

DISCUSSION

We believe that the results of these experiments suggest (although perhaps do not conclusively prove) that the speakers make reference to the words in the lexicon, not to lexicon-independent rules, when making judgments of the kind we required of them. If they used rules, these results suggest that they would have to incorporate implausibly fine detail on the sound pattern

Table 13.5
Stimuli and Results of Experiment 3: Greenberg & Jenkins versus Early Generative Phonology (see Table 13.3 for predictions of each model)

Stimuli		Number of subjects	
A	B	A closer to English	B close to English
klɛb	klæb	4	12
θlɛz	θlɛd	1	15
ʃriz	ʃrɪd	5	11
flʊt	flɪg	7	9
sθɛʃ	sθip	7	9
Total		24	56*

*$p < .001$.

Table 13.6
Stimuli and Results of Experiment 3: Greenberg and Jenkins versus Chomsky and Halle (see Table 13.4 for predictions of the two models)

Stimuli		No. subjects	
A	B	A closer to English	B closer to English
mlɔʒ	mlit	4	11[a]
mløf	spøf	1	15
xlox	xrit	5	11
sfʊb	sfɛt	1	15
Total		11	52*

[a]One of the 16 subjects gave no response.
*$p < .001$.

of the language. In general, then, our results give support to the G&J model.[4]

There are still many other details to test, of course—among others, whether the speaker operates with phonemes or features in comparing the target form with existing words. Relevant to this question is the research of Derwing and Nearey (this volume), who show that when subjects judge the phonetic similarity of pairs of forms, there is some evidence that a featural analysis is performed.

[4]Halle (personal communication, 1975) acknowledges that the proposals in *The Sound Pattern of English* on the quantification of morpheme well-formedness were tentative and incorrect to the extent that they did not predict /mløf/ to be further from English than /spøf/.

Another issue is whether, as G&J somewhat arbitrarily maintain, deletions and additions play no part in the transformation made in the comparison precedure. Implicitly, such operations are allowed by the algorithm of Vitz and Winkler (1973), which was offered as a model of how subjects compare the phonetic similarity of forms and which received some empirical support from their experiments as well as those of Derwing and Nearey. The resolution of this issue has potentially important implications for the interpretation of the results of an interesting experiment conducted by Pertz and Bever (1975). They investigated the awareness which monolingual speakers of English have of universal consonant-cluster patterns. They found that their subjects judged a word such as [ntel] to be more common in the world's languages than [nkel] (which, of course, is the case). They concluded that since neither of these clusters occurs in initial position in English and that since both words would be rated equally far from English using the G&J measure, including the prohibition on addition or deletion of phonemes, their subjects could not have used their knowledge of English in making their judgments. (Pertz and Bever concluded that the source of knowledge which subjects invoked in making these judgements must have been formed during language acquisition.) However, this conclusion, like ours, is only as good as the assumptions it is based on. If subjects can use something like the G&J measure but with addition or deletion of phonemes as part of the transformations permitted, then these initial clusters could become final or medial clusters. In that case the homorganic nasal + stop clusters would certainly be judged preferable or more common than the nonhomorganic clusters, since the former are also more common in English. The word *lintel* (top part of a doorway) might form the nearest match in that case, whereas there seems to be no comparably close match exhibiting the sequence [-nkel-]. Thus we do not necessarily have to assume that Pertz and Bever's subjects gave the responses they did from knowledge not based on the sound patterns in their mental lexicons of English.

Other factors deserve study, too. Does the number of existing word(s) that the input forms are transformed into or their frequency of occurrence influence speakers' judgments? Would sound sequences that are frequently or easily pronounced by English speakers, but are not necessarily lexical, be regarded by speakers as close to English? Would existing, though rare, clusters in common loan words still be judged far from English? As a preliminary answer to these last two questions, the results from G&J-type experiments done by our students (at Berkeley and San Jose) as class assignments have quite consistently given the result that forms like [θlæt] and [sræt] are rated by linguistically naive English speakers as being as close as or closer to the English pattern than a form like [sfæt], even though [sf-] clusters appear in existing English words—loans from Greek and Italian—as noted above.

Although [sr-] is not a normal cluster of English, according to most analyses, it does occur phonetically in the most common pronunciation of polysyllabic words such as *ceramic*, [sræmIk]. [θl-] is a well-known substitute for [sl-] in speech that attempts to imitate a lisp. (On the problem of accidental gaps in consonant clusters, see Fischer-Jørgensen, 1952.)

On the other hand, a pilot study by Larson (1982) showed that the predicted "Englishness" for most clusters by the G&J metric correlates (inversely) with their frequency of occurrence in the English lexicon (this is not the same thing as frequency of occurrence in running text or running speech). However, in the cases she examined where the two would give different predictions (for example, /glip/ and /trip/ would be equally close to the English pattern by the G&J metric—according to our analysis of Larson's stimuli—but /trip/ has a cluster that is more frequent in the lexicon), it turns out that frequency of occurrence is a better predictor of native speakers' ratings.

It would also be worth investigating the role that spelling might play in influencing speakers' judgments of the well-formedness of forms, that is, the extent to which their reactions are based on "graphotactic" rather than phonotactic patterns. In a preliminary study which addressed this question, Perucca-Ramirez (1982) found that six-year-old American children, though literate in a relatively small fraction of their aural vocabulary (and thus presumably less influenced by spelling), nevertheless treated made-up words virtually identically to adults in a G&J-type experiment.

CONCLUSION

We believe that the methods we report here, first applied to the study of phonological problems by G&J and by Zimmer, are capable of giving answers to these and additional questions regarding the psychological manifestation of MSCs. All that is really necessary to make this or any experimental technique useful is the formulation of hypotheses in a fashion that is explicit enough to render them testable.

REFERENCES

Chomsky, N., M. Halle. 1968. *The sound pattern of English*. New York: Harper & Row.

Esper, E. A. 1925. *A technique for the experimental investigation of associative interference in artificial linguistic material*. Language Monographs, no. 1.

Fischer-Jørgensen, E. 1952. On the definition of phoneme categories on a distributional basis. *Acta Linguistica Hafniensia* 8:8–39.

Fromkin, V. A. 1971. The non-anomalous nature of anomalous utterances. *Language* 47:27–52.

Greenberg, J. H., & J. J. Jenkins. 1964. Studies in the psychological correlates of the sound system of American English. *Word* 20:157–177.

Halle, M. 1959. *The sound pattern of Russian*. The Hague: Mouton.

Halle, M. 1962. Phonology in generative grammar. *Word* 18:54–72.

Larson, M. 1982. Referring to one's lexicon for judgments on morpheme structure constraints: Evidence from the issue of frequency. San Jose, CA. Ms.

Ohala, M. 1972. Topics in Hindi-Urdu phonology. Ph.D. diss. University of California, Los Angeles.

Ohala, M. 1983. *Aspects of Hindi phonology*. Delhi: Motilal Banarsidass.

Perucca-Ramirez, M. 1982. Experiment on morpheme structure conditions: Adult and child native-speaker intuition. San Jose, CA. Ms.

Pertz, D. L., & T. G. Bever. 1975. Sensitivity to phonological universals in children and adolescents. *Language* 51:149–162.

Stanley, R. 1967. Redundancy rules in phonology. *Language* 43:393–436.

Vitz, P. C., & B. S. Winkler. 1973. Predicting the judged 'similarity of sound' of English words. *Journal of Verbal Learning & Verbal Behavior* 12:373–388.

Wright, J. 1975. Nasal-stop assimilation: Testing the psychological reality of an English MSC. In C. A. Ferguson, L. M. Hyman, & J. J. Ohala, eds., *Nasalfest. Papers from a symposium on nasals and nasalization*, 389–397. Language Universals Project, Stanford University.

Zimmer, K. 1969. Psychological correlates of some Turkish morpheme structure conditions. *Language* 46:309–321.