

## AGAINST MORPHEME STRUCTURE RULES AND OTHER THINGS

THEODORE M. LIGHTNER

1. Examination of the underlying representations of morphemes in a synchronic description of Modern English reveals that the first consonant of a morpheme-initial consonant sequence is restricted to *s* (more detailed discussion, see §3 below). One way to make this fact explicit is to postulate a MORPHEME-STRUCTURE RULE of roughly the following shape:

(MS1)  $C \rightarrow s / + \text{_____} C$

All languages have restrictions on segment sequences that may appear within morphemes, but it is not clear that such restrictions should be stated in terms of rewriting rules of the type (MS1). With respect to the question of how best to capture such restrictions on the shape of morphemes, I would like to examine some data from Semitic (all of the data in what follows are taken from Greenberg, 1950).

In trilateral verb roots one finds, as a first approximation, that homorganic consonants are excluded. Thus Greenberg found, in his examination of 3775 verb roots, that in the I-II position there are no roots with *hm-*, *bf-*, *gk-*, *td-* etc., and in the II-III position no roots with *-ts*, *-bf*, *-kq*, *-lr* etc.

Regardless of what features one uses to characterize point of articulation, it is evidently not easy to write one or more rewriting rules which make explicit the possible distribution of segments within a morpheme. Such a rule (or such rules) would have to assure that each consonant in a trilateral verb root had a different specification for at least one point of articulation feature. One might write

(MS2)  $C \rightarrow [-\alpha F_a] / \text{_____} [\alpha F_a]$  in verb roots, where  $F_a$  represents any arbitrary point of articulation feature.

But if such an analysis is permitted, then a problem of indeterminacy arises because in place of (MS2) one could have

(MS3)  $C \rightarrow [-\alpha F_a] / [\alpha F_a] \text{_____}$  in verb roots

Or one could have (I use the notation described in Bach, 1968):

(MS4)  $C \rightarrow [-\alpha F_a] / [\alpha F_a]$  in verb roots

To write any morpheme-structure rule at all, we are forced to the position of formulating the rule in terms of some such variable as  $F_a = \text{any arbitrary point of articulation feature}$ . And, moreover, we must apparently make a choice as to which of the three consonants is basic; we must decide, in other words, from which one of the three consonants the other two dissimilate.

There is another problem with morpheme-structure rules, one which is difficult to discuss with the Semitic data. I return, therefore, to the English example; the reader will see that the problem occurs also in Semitic. This problem regards the underlying representation of the first consonant in a morpheme-initial consonant sequence.

To what segment should rule (MS-1) apply? If in words like *spot*, *stick*, *scan*, *small*, the first segment is represented in the lexicon as *s*, then rule (MS1) will always apply vacuously and one wonders why the rule should be included in the grammar. On the other hand, one might want to hold the position that segments in underlying representations be minimally marked. Thus, for example, if *p* is the minimally marked obstruent,<sup>1</sup> then *spot*, *stick*, *scan*, *small* would be represented /ppot, ptik, pkan, pmal/, and rule (MS1) would correctly apply to morpheme-initial *p* to derive *s*.

Although we have found, in the last paragraph, a conceivable use for morpheme-structure rules, it is not at all clear how such a use would be applicable to the consonants in Semitic. Moreover, it seems entirely counter-intuitive to have the first segment in *spot*, *stick*, *scan*, *small* be *p* or *p̣* or *t* or, in fact, anything other than *s*.

An alternative to morpheme-structure rules is morpheme-sequence well-formedness conditions. In Semitic we state simply that each consonant must be distinguished from the other two consonants in a verbal root by at least one point-of-articulation feature; in English we state simply that in morpheme-initial consonant sequences the first consonant must be *s*.

The situation in Semitic is more complex than described above in at least two respects. First, although identical consonants (being, of course, homorganic) are

<sup>1</sup> This is Jakobson's position. František Mareš has recently proposed that the glottal stop *ʔ* should be considered the least marked consonant. There are difficulties with both suggestions. Thus the glottal stop, for example, is relatively rare in underlying representations of languages throughout the world. Again, in languages throughout the world, if one of the labial stops is missing it is invariably voiceless *p* and not voiced *b* (Indo-European is exceptional; data and bibliographical references, see Hamp, 1970). If frequency of occurrence is to play any role in the determination of the least marked consonant, it seems that dental *t* should be the choice. Posner, 1961, gives evidence of a different type to support this conclusion; she writes (p. 37) that 'The consonants that are articulated with the tip of the tongue (dentals and alveolars) are *ceteris paribus* easier to pronounce than the others: most difficult are those articulated by movements of the lips. 'Difficulty' here seems to be correlated with the speed at which the muscles controlling the articulatory organs can move: owing to the effect of the facial nerve, the lips move more slowly than the back of the tongue, which in turn moves more slowly than the tip of the tongue. The result is that *ceteris paribus* dental consonants are shorter in duration than the velar, palatal, and labial consonants, and the effort of pronunciation is consequently reduced. A concomitant of the relative ease of pronunciation of the dentals is the fact that they are pronounced with more power - as less effort has to be devoted to articulation - so that they tend to be more audible'.

excluded in the I-II position (i.e., there are no roots of the type *hh-*, *pp-*, *qq-*, *kk-*, *rr-*, *bb-* etc.), identical consonants are frequent in the II-III position (Greenberg lists 10 roots with *-qq*, 11 with *-ss*, 22 with *-bb* etc.). Needless to say, the formulation of a rewriting rule which changes feature specifications in such a way as to exclude all homorganic consonants EXCEPT IDENTICAL CONSONANTS will be an even more difficult task than the oversimplified task outlined a few paragraphs above.

Second, in addition to the restriction on the I-II and the II-III positions, there is an identical restriction on the I-III position; neither homorganic nor identical consonants may appear here. Thus Greenberg found no morphemes with *d-d*, *z-z*, *m-b*, *z-s*, *s-z*, *s-s* etc. Again, it is easy to state the well-formedness condition.

2. I would like to turn now to a different type of example. Consider English sentences of the type  $S_1$  and  $S_2$  too. We have

- (1) John bought a banana and Mary bought one too.
- (2) John bought a banana and he bought an odalist too.
- (3) John bought a banana and he ate it too.
- (4) \*John bought a banana and he bought a banana too.
- (5) \*John bought a banana and he played the violin too.
- (6) \*John bought a banana and Mary bought an apple too.
- (7) \*John bought a banana and Mary drinks water too.

Examples like these show that in conjoined sentences of the type  $NP + V + NP + \text{and} + NP + V + NP + \text{too}$ , one (and only one) of the elements in the sentence before *and* must be different from the corresponding element in the sentence after *and*. Thus in (1) the subject NP's differ; but in the ungrammatical (4) no elements differ and in the ungrammatical (5) two elements – the verb and the object noun phrase – differ.

It is evident that here – just as in the case of Semitic consonants – it will be extremely difficult to give a rewriting rule governing the conjunction. But the statement of a condition is straightforward.

There are a few difficulties. Thus in (1) *banana* and *one* do not refer to the same banana (= coreferentiality), as can be seen by comparison with

- (8) John bought a banana and Mary bought it too.

Thus identity of reference is not involved. Nor is morphemic identity involved because of well-formed sentences like

- (9) John ate a banana and he consumed an apple too.
- (10) John arrived with scotch and Mary came with scotch too.

In (9) *consumed* can only be interpreted as meaning *ate*, and in (10) *came* can be interpreted only as meaning *arrived*. Thus the condition seems to be that semantic identity (but not identity of reference) of precisely two elements is required.

I find sentences like (11) *John's going to the Riviera, and Harry's going to Bermuda, too* well-formed, especially with certain types of special intonation envelopes. But the meaning here is (11)' *John's going to take a pleasure trip and Harry's also going to take a pleasure trip*.

Less easy to explain are sentences like

- (12) John loves Mary and Mary loves John too.  
 (13) Bill hit Harry and Harry hit Bill too.

But in these sentences (which can be roughly paraphrased by (12)' *John and Mary love each other*, and (13)' *Bill and Harry hit each other*), there must be identity of reference between the two occurrences of *John, Mary, Bill, Harry*; perhaps (12), for example, is to be interpreted along the lines of

- (12)'' There was loving going on between John<sub>1</sub> and Mary<sub>1</sub>, and between Mary<sub>1</sub> and John<sub>1</sub> too.

Sentences like (14) *John hit Harry and Bill hit Sam too*, are presumably interpreted as

- (14)' There was hitting going on between John and Harry, and between Bill and Sam too.

As can be seen from this extremely brief look at sentences of the form  $S_1$  and  $S_2$  too., the situation is not crystal clear. Nevertheless, the general outlines seem clear; however these sentences are best analyzed, it will be a condition – not a rewriting – rule – that plays the key role.

3. The decision to introduce conditions to describe linguistic phenomena – a decision which seems inescapable not only in view of the Semitic triliteral roots, but also in view of a large number of syntactic constraints and in view of the general nature of the lexicon – raises a number of interesting questions, two of which I discuss here:

Let us look more carefully at the morpheme-initial consonant sequences in English. Except before the liquids *r* and *l*, only *s* is permitted: we have *trick* but no \**tmick*, *glade* but no \**gbade*, *shriek* but no \**shniek*, and so on. This holds true for the non-Greek part of the English lexicon (with, needless to say, a few well-known exceptions such as, for example, *zweiback*, from Ger *zwei* 'two' + *backen* 'to bake', pronounced [tsv-], [tsw-], [zv-], [zw-], [sw-], depending on the dialect). But in the Greek part of the lexicon this does not hold true: *amnesia* has morpheme-initial *mn-*, *autochthonous* has *kθ-*, *apnea* has *pn-*, *archaeopteryx* has *pt-*, *agnostic*, *diagnostic* have *gn-*,<sup>2</sup> *apsychia* has *ps-*, and so on. The interesting thing about words like these is that they all have

<sup>2</sup> In Germanic, *kn-* and *gn-* are permitted: thus historically related to Greek *gnō-* is English *know*. One can argue that *know* must have underlying *kn-* in a synchronic description on the basis of *acknowledge* with phonetic [-kn-], derived from prefixal *ad-* in the following manner: *ad+kn* → *ak+kn* → *ækkn*.

related forms in which the morpheme in question occurs in word-initial position; and phonetically, these words all begin with a single consonant (barring spelling pronunciations, about which I have nothing to say); thus, *mnemonic* with [n-], *chthonian* with [θ-], *pneumonia* with [n-], *pterodactyl* with [t-], *gnostic* with [n-], and *psychology* with [s-].

In English there seems to be a particularly strong constraint against having a word begin with phonetic CC- (where C is not a liquid) unless the first C is *s*; this constraint is manifested first of all in the morpheme-structure constraint and second of all in the phonological rule which drops any consonant except *s* in word-initial pre-consonantal position (*mnemonic* → *nemonic* etc.). The situation is peculiar because the morpheme-structure constraint does not apply to the Greek part of the lexicon, whereas the phonological rule applies only to the Greek part of the lexicon.

Another interesting question raised by the necessity to include constraints in grammars regards the two distinct resulting types of rules: the rules of constraint and the usual rewriting rules.

Let us consider the case for rewriting rules. If one finds a language at one stage with phonetic velars before front vowels and at a later stage with palatals in place of the velars before front vowels, it seems natural to say that the change which has occurred is expressed precisely by the addition of the rewriting rule  $K \rightarrow \check{C} / \text{---} \check{V}$  to the grammar. Similarly, if VP's, say, appear in surface form either as V or as V+NP, it seems natural to say that the relations are expressed precisely by the rewriting rule  $VP \rightarrow V (+ NP)$ .

Presumably it is the naturalness of the rewriting rules in cases like these (and there are many such cases) which led to the introduction of rewriting rules in the first place. One then assumed that *all* rules were rewriting rules: there was only one type of rule. Now we have shown that there must be rules of constraint. Since it would be preferable to have only a single type of rule, we can ask whether rewriting rules are not altogether a mistake. Perhaps we can make do with only a single kind of rule, rules of constraint.

Now it is easy to see that rewriting rules can be reformulated in terms of rules of constraint, whereas the reverse, as we have shown, is not always true (or at least is not true in any LINGUISTICALLY relevant sense). The problem that arises is whether there is any empirical difference between a rewriting rule and a rule of constraint in those cases where it is not clear that a rule of constraint is required (in cases like velars shifting to palatals before front vowels).

This is an extremely difficult question, for which I have no answer. But we might approach the problem by looking at rules which deal with more than a single segment. Consider, for example, a rule which drops all word-final consonants:

- (15)  $C_1 \rightarrow \emptyset / \text{---} \#$ , where  $C_1$  represents one or more consonants,  $\emptyset$  represents zero, and  $\#$  represents word-boundary.

By virtue of this rule, forms in *-at* are realized as [-a], forms in *-ozd* as [-o], forms in *-ikθs* as [-i], and so on. The question is: how does the rule apply when an underlying representation ends in more than a single consonant? Take the case in *-ikθs*: are all three final consonants dropped in one step? Or are three steps involved? And if three steps are involved, which consonant drops first and which last? If three steps are involved, the natural steps, it seems to me, are: *-ikθs* → *-ikθ* → *-ik* → *-i*. It is the word boundary following the consonant which is significant. A derivation like *-ikθs* → *-iθs* → *-is* → *-i* misses the fact that the word boundary is significant.

But we have not yet answered the question as to whether the derivation should proceed in one or in three steps. Perhaps this question can be related to the question of whether rewriting rules are to be permitted. Thus one might argue that if all consonants are to be dropped in a single step, the correct way of writing the rule is a condition placed on two successive steps in derivations. But if one wants to argue for the stepwise process – presumably the sequence of events that reflects a historically natural process – the correct formulation of the rule is in terms of a rewriting rule. One could then give a precise meaning to the subscript *one* in the symbol  $C_1$ .

Similar examples are found in a rule which palatalizes all consonants in a cluster preceding a front vowel (*stri* → *s't'r'i*, and so on), in a rule which devoices all word-final obstruents (*služb* → *služp*, and so on), although in derivations like these a different interpretation is available if one is searching for a rule to reflect natural processes. One might want to say, for example, that it is only the consonant immediately preceding the front vowel which is palatalized (*stri* → *str'i*) and that it is only a word-final obstruent which is devoiced (*služb* → *služp*). The problem of sequential ordering still arises, however, because one now needs a rule to assimilate palatalization in consonant sequences (*str'* → *st'r'* → *s't'r'*) and to assimilate voicing in obstruent sequences (*vzt* → *vst* → *fst*).

The point at issue here is to find some empirical difference between a rule of condition and a rewriting rule.

4. The Semitic data pose two further problems, one dealing with the exceptions to the general constraint on what types of consonants may appear in the three different positions, the other dealing with the choice of the universal set of phonological features.

4.1 There are exceptions in all three positions: in I-II, Greenberg found one morpheme with *td'*, five with *m-* etc.; in II-III three morphemes *-td*, one with *-bm* etc.; in I-III one morpheme with *k-g*, five with *b-m* etc. The problem with these data is that there is a gradation in the number of exceptions and the gradation seems to coalesce with the probability of occurrence of a vowel between two consonants. Thus there are more exceptions in the I-III position than in the II-III position than in the I-II position; in the I-III position it is certain that a vowel will intervene

between the first and third consonant, in the II-III position it is more likely that a vowel will intervene between the consonants than in the I-II position (I am indebted to Robert Freund for this observation).

If the reason underlying this dissimilative restriction on the distribution of consonants in verbal trilateral roots has to do with maximal differentiation of adjacent segments (see Jakobson, 1940), then it seems reasonable to suppose that the PHONETIC representations (= consonant sequences with or without intervening vowels) are affecting the distribution of exceptions in the shape of the underlying representations.

There is also another gradation in the type of exceptions. Take the consonants *r*, *l*, and *n*. These consonants do not cooccur in verbal trilateral roots, but they occur freely with all other consonants. According to this classification, therefore, *r*, *l*, and *n* are homorganic consonants. But within the exceptions there are less in which liquids cooccur than in which a liquid and *n* cooccur. The reason is clear: *n* is less homorganic to *r* and *l* than *r* and *l* are to themselves or to each other.

Thus it looks as if one can say that the number of exceptions is directly proportional to the number of segments intervening between the two segments in question and to the number of feature differences between the two segments in question. It is not clear how such a statement (if this is indeed the correct statement) is to be incorporated in a grammar.

4.2 The problem regarding features has to do with the back consonants, which Greenberg divides into four series: laryngeals (*ʔ* and *h*), pharyngeals (*ʕ* and *ħ*), post-velars (*x* and *ɣ*), and velars (*k*, *g*, and *q*). These back consonants do not cooccur freely within a verbal trilateral root, but they do occur freely with all other consonants, with liquids (*r*, *l*, *n*), with front consonants (*q̄*, *ʃ*, *s̄*, *z*, *ʂ*, *t*, etc.), and with labials (*p*, *b*, *m*). The back consonants therefore form a homorganic class. But within this class there are three groups: laryngeal-pharyngeals, post-velars, and velars (going from back to front). Post-velars may not cooccur with laryngeal-pharyngeals or with velars. But the velars (*k*, *g*, *q*) occur freely both with pharyngeals (*ʕ* and *ħ*) and with laryngeals (*ʔ* and *h*).

*University of Texas at Austin*

#### REFERENCES

- Bach, Emmon  
1968 'Two proposals concerning the simplicity metric in phonology', *Glossa* 2, 128-149.
- Greenberg, Joseph H.  
1950 'The patterning of root morphemes in Semitic', *Word* 6, 162-181.
- Hamp, Eric P.  
1970 'Maya-Chipaya and the typology of labials', *Papers from the Sixth Regional Meeting of the Chicago Linguistic Society*, 20-22.
- Jakobson, Roman  
1940 *Kindersprache, Aphasie und allgemeine Lautgesetze* (English translation: *Child Language, Aphasia, and Phonological Universals* (= *Janua Linguarum* No. 72) (Mouton and Co.).

Mareš, František

Forthcoming 'Die Grundbegriffe der phonetischen Beschreibungen', *The Slavic Word* (= *Proceedings of the International Colloquium held at UCLA, September, 1971*) (to be published by Mouton and Co.).

Posner, Rebecca R.

1961 *Consonantal Dissimilation in the Romance Languages* (Oxford).