

Double-Cross in Phonology: Why Word-Boundary (Often) Acts Like a Consonant

Author(s): Richard D. Janda

Proceedings of the Fifth Annual Meeting of the Berkeley Linguistics Society (1979), pp. 397-412

Please see “How to cite” in the online sidebar for full citation information.

Please contact BLS regarding any further use of this work. BLS retains copyright for both print and screen forms of the publication. BLS may be contacted via <http://linguistics.berkeley.edu/bls/>.

The Annual Proceedings of the Berkeley Linguistics Society is published online via [eLanguage](#), the Linguistic Society of America's digital publishing platform.

Double-Cross in Phonology:
Why Word-Boundary (Often) Acts Like a Consonant
Richard D. Janda
U.C.L.A.

0. Introduction. This paper centers around the proposal that the reason why word-boundary (#, or "double-cross") functions like a consonant in the environments of so many phonological rules of so many diverse languages is that: (1) # is not phonetically present in the speech-signal, and (2) # is typically (= most frequently) followed by a consonant. That is, the "consonantality" of # is not due to any special properties of word-boundaries in themselves, nor--directly--to the alleged automatic cooccurrence of syllable-boundary (\$) with #, but, rather, to the association of word-boundaries with the consonants that usually follow them. Support for this proposal is provided in two ways: firstly, evidence is presented that the (to the best of our knowledge) only two alternative proposals for explaining the quasi-consonantality of # are inadequate in ways that the new one is not; secondly, results are given from a study of twenty languages in which it was found that the percentage of #C (i.e., consonant-initial words) in running text had a mean frequency of 75%, ranging as high as 90% (in Swahili), but never dropping below 63% (in Hungarian). Toward the goal of verifying the new proposal, the paper specifies, finally, a type of potentially disconfirming evidence that can and, of course, should be investigated.

1. Natural Classes in Phonology and the General Problem of Unnatural Classes. The sounds of the languages of the world--of any language and, hence, of human language in general--can be grouped into classes according to their properties, properties of an articulatory, acoustic, or auditory nature. That these natural classes are not merely some artifact of linguistic-phonetic analysis without any reality in the minds--or mouths or ears--of speakers/hearers of language is shown by the fact that such classes act as units in the phonological rules of languages, this unity consisting in the situation that the sounds which make up natural classes act together: undergoing some rule together as input or output, conditioning some rule together (in its environment), or, in a more extended sense, changing into one another or being derived in the environment of one another (cf. Hyman 1975:140).

The success--that is, the consistency and coherency--with which phonological systems (both inventories and rules) in general can be described in terms of natural classes has led phonologists to the actually rather extreme assumption or expectation that practically all of phonology--especially rules--should be describable in terms of such classes, in the sense that the appearance of unnatural classes (in "crazy rules", for example; cf. Bach and Harms 1972) is designated as highly marked, and, consequently, expected to be of extremely rare occurrence. For example, Chomsky and Halle (1968: 335) state: "[a]ny linguist would react with justified skepticism to a grammar that made repeated reference to a class composed of just the four segments [p, r, y, a]. These judgments of 'naturalness' are supported empirically by the observation that it is the

'natural' classes that are relevant to the formulation of phonological processes in the most varied languages, though there is no logical necessity for this to be the case."

In generative phonology, with its adherence to the formal theoretical principle that properties of the object of description should be reflected as directly as possible in the properties of the descriptive notation, so as to maximize predictive/explanatory power, natural classes are described by a set of distinctive features which are established in such a way that they permit a rather mechanical definition of the notion "natural class": a "natural class" is a group of segments (or phonological elements) whose specification as a whole requires fewer features than are required to specify any one member of the group. Thus, whereas a phonetic-alphabetic notation itself shows no difference in naturalness between [p, r, y, a] and, say, [m, n, ñ, ŋ], the feature-notation shows a very clear difference. In setting up a distinctive-feature system, one starts with one's intuitions about natural classes and then defines features--and assigns feature-values for segments--so as (in accordance with the above definition of "natural class") to isolate such classes notationally. The theory having been set up for the clear(er) cases here, one must then let the theory decide for one in those cases where one's intuitions run out--let it tell one which classes are natural (or more so) and which are not (or less so). Then, one can judge any class of segments (or phonological elements in general) in terms of naturalness, and the theory, once having been set up in this way, then also makes strong predictions about which classes will be found in the phonological rules of the languages of the world. When any one such prediction is shown to be consistently not borne out, then one must either revise the feature-system or else make the offending unnatural class go away by some other means (e.g., by questioning the existence, or details of the formulation, of rules referring to it, especially their unity)--this all in line with the general assumption that language/s operate/s in terms of natural classes.

2. The Widespread Unnatural Class {C, #}. Given the traditional generative-phonological framework for defining natural classes, however, a problem immediately arises with the class {C, #}. Since C (consonant) is [-syllabic] (and, for true consonants, [+consonantal]), as well as redundantly [+segment], while # is [+word-boundary] and redundantly [-formative-boundary] and [-segment], as well as unspecified for [syllabic] (and [consonantal]), it is clear that the two share no feature(-value)s and cannot be referred to unitarily, as a natural class. However, {C, #} is extremely common as part of the environment of phonological rules in the most genetically and geographically diverse languages. Here, we present a small but fairly representative sample of such rules drawn from eight languages:

- (1) a. Akan (cf. Schachter and Fromkin 1968:80-81, 262; Hooper 1972:533)

$$\left[\begin{array}{c} \text{V} \\ +\text{high} \end{array} \right] \rightarrow [+nasal] / _ \left[\begin{array}{c} \text{C} \\ +\text{nasal} \end{array} \right] \left\{ \begin{array}{c} \text{C} \\ \# \end{array} \right\}$$

- b. Catalan (cf. Vogt 1970; Anderson 1974:69-70)

$$\begin{bmatrix} +\text{continuant} \\ +\text{voice} \\ +\text{labial} \end{bmatrix} \rightarrow u / _ \left\{ \begin{array}{c} C \\ \# \end{array} \right\}$$

- c. Finnish

- (i) (Part of) Consonant-Gradation (cf. Anderson 1974:85)

$$\begin{bmatrix} -\text{continuant} \\ -\text{voice} \end{bmatrix} \rightarrow \left\{ \begin{array}{l} [+voice \\ \left[\leftarrow \text{continuant} \right] \\ \emptyset \end{array} \right\} / \left[\begin{array}{l} +\text{son} \\ \left[\leftarrow \text{cor} \right] \\ \left[\leftarrow \text{cor} \right] \end{array} \right\} \text{VC} \left\{ \begin{array}{c} C \\ \# \end{array} \right\}$$

- (ii) Three-Consonant-Cluster Simplification (cf. Karttunen 1970:98-99)

$$k \rightarrow \left\{ \begin{array}{l} h / \left[\begin{array}{c} \text{+primary stress} \\ \left[\leftarrow \text{continuant} \right] \end{array} \right\} C \left\{ \begin{array}{c} C \\ \# \end{array} \right\}, \text{ a subpart of } C C \left\{ \begin{array}{c} C \\ \# \end{array} \right\} \\ \emptyset / \left[\begin{array}{c} \text{+primary stress} \\ \left[\leftarrow \text{continuant} \right] \end{array} \right\} C \left\{ \begin{array}{c} C \\ \# \end{array} \right\} \end{array} \right. \Rightarrow 2 \ 3 \text{ (where the sequence } 1 \ 2 \ 3 \text{ violates basic } \begin{matrix} 1 & 2 & 3 \\ \text{sequen-} \\ \text{tial} & & \end{matrix} \text{ constraints)}$$

- d. Hausa (Russ Schuh, personal communication)

$$V \rightarrow [-\text{long}] / _ C \left\{ \begin{array}{c} C \\ \# \end{array} \right\}$$

- e. Korean (cf. Kim 1972:161)

$$(i) \{s, ss, c, ch\} \rightarrow t / _ \left\{ \begin{array}{c} C \\ \# \end{array} \right\}$$

$$(ii) [-\text{continuant}] \rightarrow [-\text{release}] / _ \left\{ \begin{array}{c} C \\ \# \end{array} \right\}$$

- f. Luiseño (cf. Kroeber and Grace 1960:21-23; Langacker 1972:45)

$$\check{c} \rightarrow \check{s} / _ \left\{ \begin{array}{c} C \\ \# \end{array} \right\}$$

- g. Turkish (cf. Lees 1961:36-37)

$$C_1 C_2 \rightarrow C_1 / _ + \left\{ \begin{array}{c} C \\ \# \end{array} \right\}$$

- h. Yawelmani (cf. Kisseberth 1970:294, 296, 297)

$$(i) V \rightarrow [-\text{long}] / _ C \left\{ \begin{array}{c} C \\ \# \end{array} \right\}$$

$$(ii) \emptyset \rightarrow i / C _ C \left\{ \begin{array}{c} C \\ \# \end{array} \right\} \text{ (an identical rule exists in Pales-} \\ \text{tinian Arabic; cf. Brame 1973:21)}$$

In addition to these cases, one might also mention that Javkin (1975) found "15 rules ... which treated [word-]boundaries in a way similar to consonants ..., in that similar processes are triggered by both" in his study of 524 phonological rules in 13 languages then included in the Stanford Phonology Archive (cf. also Vihman 1977). The typological and areal widespreadness of the class $\{C, \#\}$ would thus seem to be in direct contradiction to the putatively unnatural status of that class.

In such a case as this, where the sheer number and the validity of the rules referring to a certain class of phonological elements can scarcely be questioned, we are, in fact, provided with the basis for an intuition that that class is a natural one, and we must then try to revise our linguistic descriptions in some way so that the notation will correctly designate it as such. As pointed out by Kahn (1976:10-11), there are two main approaches to a solution that one may take when faced with the general problem of a phonological rule $A \rightarrow B/X_Y$ (where either or both of X and Y may be null) in which one or more of A, B, X, Y consists of an unnatural class $\{m, q, \dots\}$.

On the one hand, one may try to discover (and independently motivate) some feature [F] such that {m, q, ...} are all [αF], while all other phonological^x elements are [-αF]. Or, on the other hand, one may try to discover (and motivate) some boundary(-marker) φ such that it occurs (only) at {m, q, ...}. In the case of {C, #}, both approaches have been taken by phonologists and can be found represented in the literature.

3. A First Attempt to Make {C, #} a Natural Class: # as a Consonant.

The first proposal that # is a consonant--at least the first one in the generative literature--seems to have been that of Lass (1971), who was soon followed in this by Lightner (1972). The considerations motivating this suggestion are fairly well capsulized by Lightner's statement (p. 333): "We notice that over and over again the disjunction {C, #} appears in many different rules in many different languages. The natural thing to do, therefore, is to assign # the features of a consonant, for, in this way, we can make explicit the fact that # behaves like a consonant." At least in Lightner's formulation, the logic of this argument seems, unfortunately, to be merely that, since rules mentioning {C, #} can be simplified by making # a consonant, then they should be so simplified, tout court. Obviously, however, this is more than a purely notational question, for anything can be simplified; what is really at stake is precisely whether or not the disjunction {C, #} should be simplified. And, indeed, an argument somewhat more along these more motivated lines had already been given by Lass (1971).

Lass observed that, in the Old English medial-voicing rule for fricatives that gave rise to such alternations as life ~ lives and bath ~ bathe--see (2) below--word-boundary acted like a voiceless sound-segment in blocking the process:

$$(2) \begin{bmatrix} \text{-sonorant} \\ \text{+continuant} \end{bmatrix} \rightarrow \text{[+voice]}/\text{[+voice]}_\text{[+voice]}$$

He similarly pointed out that the generalization (he said "rule", but this is to conflate a phonological with a redundancy-rule) of German phonology that, if an obstruent cluster has one voiceless member, then it has all voiceless members can completely subsume the infamous rule of final devoicing (for which one really need give no examples, here) if # is a voiceless obstruent; cf. (3) and (4) below.

$$(3) \text{[-sonorant]} \rightarrow \text{[-voice]}\% \begin{bmatrix} \text{-sonorant} \\ \text{-voice} \end{bmatrix}$$

$$(4) \text{[-sonorant]} \rightarrow \text{[-voice]}/_\text{\#}$$

But Lass also attempted to provide a phonetic basis for his claim that "at least in terms of power to inhibit voicing, # is really a voiceless obstruent (albeit one with no articulatory features besides voicelessness)" (p. 16): he equated # with silence, which is undeniably voiceless.

This identification turns out to be untenable, but, prior to our stating why, it is instructive to consider some purely formal inadequacies of the Lass/Lightner proposal. It has been pointed out several times before (e.g., by Kahn 1976:12) that, if # is in-

deed a consonant, a voiceless obstruent just like any other such sound, then it must remain entirely unexplained and accidental why there are no phonological rules which mention # in focus-position. For example, no language palatalizes # before high front vowels, as in (5) below:

$$(5) * \# \rightarrow \begin{bmatrix} +\text{high} \\ -\text{back} \end{bmatrix} / _ \begin{bmatrix} +\text{high} \\ -\text{back} \end{bmatrix}$$

In order to prevent rules that otherwise apply quite regularly to (segmental) consonants from overgeneralizing to #, the inputs to such rules would have to be complicated by the addition of [+segment]. Such mention of [+segment] would, in fact, be necessary in almost every phonological or redundancy-rule referring to consonants; for example, in order to prevent a Lass/Lightner-type phonological description from analyzing German as having clusters of up to ten consonants; cf. (6):

$$(6) \text{ du schrumpfst Strümpfe ein 'you (familiar sing.) shrink socks'}$$

/...mpfst##str.../ .

Moreover, even if such complications did not far outweigh the rule-simplifications effected by analyzing # as a consonant, Lass's analysis still would not work for German. It has apparently not heretofore been noticed that, since there does really seem to be a generalization in the language that the presence of one voiceless obstruent in a cluster of obstruents requires the entire cluster to be voiceless (regressively, in the form of a phonological rule similar to (3) above; progressively, in the form of a redundancy rule similar to it, also), then the status of initial word-boundary in German as a voiceless obstruent would lead us to expect that German should have no initial voiced obstruents, which is manifestly and egregiously wrong.

But the main failing of the word-boundary-as-voiceless-obstruent approach is that, as opposed to pause (||), # is "just not there", phonetically; it is absurd to assign segmental features to a non-segment which, though given by morphosyntax in the phonological string, thus can have no direct phonetic effects (assimilatory or dissimilatory) on adjacent segments. This point is well made by Kahn (1976:11): "There is no hope of independently justifying the segmental features imposed on word-boundary in terms of articulatory or acoustic considerations. This is so because word-boundary is simply not articulated and is not present in the acoustic signal. [Trying] to associate word-boundary with silence ... of course ... fails in general: there is no cessation of vocal-cord vibration and sound-output at the points of sentence-internal word-boundary in the phrase [e.g.] John#is#a#bum as normally spoken." In short, while the observation that # acts like a consonant in many ways does contribute to our understanding of linguistic phenomena, the claim that it therefore is a consonant does not, for, then, all of those ways in which # acts differently from consonants become equally inexplicable, once this kind of attempt at reductionism is made. Not sur-

prisingly, the reductionist approach to explaining the quasi-consonantality of # does not seem to have been seriously pursued beyond the early programmatic suggestions. Most phonologists who have addressed the problem since then have espoused Kahn's second possible approach, the common-boundary solution.

4. A Second Attempt to Make {C, #} a Natural Class: # and C as the Locus of Syllable-Boundary (\$) or Margin. The first significant generative phonological proposals for incorporating the syllable into phonological theory were those of Natural Generative Phonology, at a time more or less contemporaneous with Lass's and Lightner's proposals; cf. especially Vennemann (1971, 1972) and Hooper (1972). Recognizing the syllable as a prime phonological unit, of course, allows one to propose that what unites the two disparate members of the disjunction {C, #} in phonological rule-environments is the fact that both C and # in it occur at a syllable-boundary (= \$). This suggestion is, naturally, implicit in much pre-generative and earlier generative work, and has been extensively followed-up by both natural generative phonologists themselves (cf. Vennemann 1974, Hooper 1976, 1977, and references in the last of these) and others of somewhat different theoretical persuasions. For example, Kahn (1976) countenances "ambisyllabicity", the simultaneous membership of a single segment in two syllables, and so speaks in terms of syllable-membership (and syllable-marginality or non-marginality) instead of syllable-boundaries. In either case, however, the relevant claim is that, rather than the usual agglomeration of subscripted C's and V's otherwise customarily needed to characterize open or closed syllables, one need merely refer to syllable-boundary or to relative position of a segment to the end of a syllable; cf. (7):

- (7) a. $\underbrace{C_0 V}_{\{(C)V\}} \{ \# \} = (\text{one type of}) \text{ open syllable} = C_0 V \$ \text{ or } \underbrace{CV}_{S_1(S_2)}$
 b. $\underbrace{C_0 VC}_{\{C_1\}} \{ \# \} = (\text{one type of}) \text{ closed syllable} = C_0 VC \$ \text{ or } \underbrace{CVC}_{S_1(S_2)}$

The use of syllables--syllable-boundaries or membership--to eliminate the unnatural class {C, #} from phonological rules presupposes, of course, that the syllable is independently motivated. While the absence of an unambiguous phonetic grounding of the syllable makes its use in linguistic descriptions still, to be honest, somewhat of a matter of faith--because a syllabic treatment can always be translated into a purely segmental one, with fewer ontological entities--nevertheless, lacking necessity-arguments, one's plausibility-arguments for the phonological syllable are quite solid and numerous, and syllabic analyses in phonology in general seem justified on the basis of conclusions like those of Ladefoged (1971:81): "[T]here is no single muscular gesture marking each syllable [, but we should not be] overly simplistic in our view of phonological properties [W]e may still be able to define a physiological unit ... (the syllable) which will account for the timing and coordination of the articulatory movements. There is evidence ... that speakers organize the sequences of complex muscular movements that make up utterances in terms of a hierarchy of units, one of which is the size of the syllable

ble We will therefore assume that a neurophysiological definition is possible, even if one cannot at the moment state it in any way." The collection of papers in Bell and Hooper (1978) provides further phonological support for the syllable.

Given the syllable, then, many phonologists have, to repeat, used it to try to explain the similar environmental behavior of # and certain consonants: both are the locus of syllable-boundaries or margins. In our opinion, however, they have devoted insufficient attention to a relevant class of cases where # and \$ arguably do not coincide--namely, cases of liaison. The claim that, since words are composed of syllables, then the end of a word must be the end of the last syllable in it presupposes that all syllables at all associated with a word are properly contained in it. However, as is explicitly recognized by Hooper (1972:527) and Kahn (1976:30-33), along with numerous others, a word-final consonant preceding a phonetically vowel-initial word is regularly syllabified with that following word (i.e., its first syllable). Significantly, though, even in such cases, where # is not syllable-final and thus does not coincide with \$ on the phonetic surface, word-boundary continues to behave like a consonant, in conditioning certain rules like those in (1) above.

Now, it is possible to analyze such liaison-phenomena (by which we mean occurrences in all languages, and not just in French, of the linking-process by which, say, English hock it! becomes, in connected speech, homophonous with Hockett, with the syllabification more or less as [ha\$ket]) in such a way as to make this surface-phonetic fact unproblematic for the syllabic approach. If liaison is viewed as a reassignment of citation-form or slow-speed syllable-boundaries or membership, and if such resyllabification is seen as due to the operation of a rule or rules ordered after the operation of syllable-sensitive rules (perhaps) including those in (1), then it is irrelevant to such rules' operation that word-boundary will later become non-syllable-final: the damage done, as it were, by the syllable-sensitive rules will then already be a fait accompli, and there are no provisions for reversing it. A global rule could, of course, also be formulated to achieve similar correct results, in such cases.

The decision to adopt or reject proposals of this sort can, it is clear, be made only within the framework of a particular theory of phonology: one that allows rule-ordering or global rules, respectively, for example. While not inclined to dismiss rule-ordering in general as a grammatical mechanism, we feel, however, that the ordering-solution to the liaison-problem effectively abandons any phonetic explanation for at least part of the cases where # behaves consonantly, since that solution incorporates the claim that one phonetically motivated generalization about syllabification is irrelevant and inapplicable underlyingly at the same time that syllabification acts to condition certain phonological rules. Now, this claim is made unhesitatingly in much recent work on the syllable as an abstract phonological unit (cf., e.g., McCarthy forthcoming and references therein), but this does seem to undercut the syllabic explanation for the naturalness of {C, #} in the phonetic-explanatory sense of "natural" (since C is phonetically located at a syllable-boundary, while # is not always coincident with \$).

Lest the importance and evidentiary status of liaison be impugned, let us consider some historical data that underscore the ambiguous word-membership of word-marginal consonants surrounded by vowels (or, more generally, syllabic segments). In the phonological diachrony of English, final -n of the indefinite article and some possessive pronouns was resyllabified before vowel-initial words (i.e., with them); and we know this--that resyllabification made certain phonologically vowel-initial sequences phonetically indistinguishable from originally n-initial words--because, in some words, the resyllabification in question was lexicalized, as in (8a) below. Furthermore, knowledge of this resyllabification must have led speakers to a kind of hypercorrection in other cases, where original initial n- was reinterpreted as belonging to the end of the preceding determiner, as in (8b).

- (8) a. an#ef(e)t(a) 'a newt'
min(e)#uncle 'my nuncle' (cf. King Lear)
min(e)#Ed(ward) 'my Ned(ward)'
 b. a#nap(e)ron 'an apron'
a#næd(e)r(e) 'an adder'
a#na(f)ogar 'an auger'

Similar exchanges have also been observed--to remain with Germanic languages--between verbs and following pronouns in one direction in Old Norse (cf. (9a), from Gordon 1957:294) and in the other direction in Old High German (cf. (9b), from Braune/Mitzka 1967:259), as well as in English (cf. (9c)).

- (9) a. skulud#er 'you (plural) shall'
 b. gilaubis#thu 'thou believest'
 c. (the equivalent of) believes#thou

Given that such liaison-phenomena cannot be handled without a certain unnaturalness (non-phoneticity) in the syllable-structure approach to eliminating the unnatural class {C, #}, one may well ask whether any alternative solution is available. That is, because ...C#V... is quite regularly syllabified as ...\$C(#)\$V..., not ...C(#)\$V..., while # continues to act like a consonant in such cases, it is clear that, barring the ordering of syllable-sensitive rules before liaison-resyllabification, syllabic considerations cannot be the primary (much less the sole) reason for the quasi-consonantality of #. And, actually, of course, even when existing solutions exhibit no obvious deficiencies, one should still, in principle, ask whether other, alternative solutions exist. In the remainder of this paper, we will propose such an alternative solution and attempt to motivate it.

5. The Frequency of #C--the Typical Association of # with Following C--as the Reason for the Consonantal Behavior of #. Let us now work our way toward accounting for the commonness of the environment(-part) {C, #} from another angle. The environment /_# abbreviates, by convention, the set of environments #C..., #V..., and #||.... Now, first off, it is clear that word-boundaries before pause/s will be relative-

ly infrequent as compared to those before vowels or consonants. However, pre-pausal position may have a status proportionally far greater than its frequency, and we will return to the role of pause toward the end of this paper. This then leaves #V and #C--initial vowels and consonants--to consider. Here, we can recall the well-known fact that CV is, in general, the canonical syllable-shape, and derive from it inferences about word-initials. According to Bell (1977), all languages have some C₁V syllables, and 20% to 40% of all languages require C₁V--but these facts/figures are drawn from sources treating the syllabification of words in isolation, where the first consonant of the first syllable is clearly the first consonant in the word. One might then surmise that the initial syllable of a word is prototypically likely to begin with a consonant, and that words in general are thus likely to begin with consonants (i.e., at least one) in universal phonology/grammar.

But if # is regularly associated with an immediately following C, then this goes a long way toward explaining the consonantal behavior of #, as well as toward guaranteeing that # and C in the disjunction {C, #} should occur with the same syllabification, regularly. Then, if #V (= word-initial vowels) are of sufficiently low frequency in a language, we might assume that speakers/hearers consequently simply disregard the liaison-phenomena that counterexemplify the generalization that # coincides with a following C and, therefore, with \$. Hence, syllable structure would be at work here, too, but only indirectly, derivatively. We would suggest, then, that the primary reason why # in general acts like a consonant is that: (1) # is not phonetically present in the physical speech-signal, and (2) it is typically (= most frequently) followed by a consonant. Hence, the extension of (presumably) phonetically motivated processes occurring in the environment /X__Y(#)C to the environment /X__Y#V represents an (over)generalizing reanalysis and, in a sense, a phonologization; as a result, we have the equivalence given in (10):

$$(10) /X_Y\left\{\begin{array}{l} (\#)C \\ \#V \end{array}\right\} = /X_Y\left\{\begin{array}{l} C \\ \# \end{array}\right\}$$

Thus, {C, #} itself is not an entirely natural class, but it is the result of an extremely natural reanalysis whereby the # in #C comes to interpreted as more criterial than the C, due to the extreme infrequency of #V. This is, in a sense, a (more complete) morphologization (morphosyntacticization) of #C; for, after all, #, though not phonetically present, is nevertheless known from the morphology and syntax, and it does have the potential to affect other segments as a conditioner of phonological rules.

We have begun to speak of #V as actually being very infrequent, and it is now time to present some evidence that this is indeed the case.

6. Evidence for the Infrequency of #V (vis-à-vis #C). First of all, it can be mentioned that, in the Basic German corpus of 600,000 running words of colloquial spoken-German texts (cf. Pfeffer 1964), the 250 most common words--which account for over 400,000 words of the total--are 73% consonant-initial. (Of course, if one takes into account

the phonetic fact that all (otherwise) vowel-initial words in German are articulatorily preceded by a glottal stop, then German actually is 100% consonant-initial in its words.)

Secondly, however, we can report, here, the results of a study of twenty languages in which the frequency of #C (again: effectively, of consonant-initial words) was the object of investigation. This study, whose findings are summarized in (11) (= Table 1) on the next page, was performed by the present author expressly to test the hypothesis just proposed in this paper, and involved randomly-chosen written texts from each of the languages in question. In each text, the first 250 junctures between orthographic words--excluding those across punctuation, which was taken to be most probably equivalent to some kind of pause--were translated into their phonological values (systematic phonemic ones, but actually only rarely differing from autonomous phonemic ones) and then listed and analyzed according to type. The total number of words of text involved in each case was thus always slightly over 500, and it should also be mentioned that the fit between divisions into orthographic and phonological words was always extremely close. The focus of the study was on word-initial segment-type (#C vs. #V), but also initial segment-voicing and place of articulation (for consonants only), as well as on abutting segment-types (juxtaposed C and V in either order, as opposed to successive C's or V's); covered were six language-families (albeit three within the larger, Indo-European group, and two within Niger-Congo/Kordofanian), and up to three diachronic stages per language. The Germanic and Romance bias of the results must, of course, be freely admitted.

What is important here is that #C ranges from 63% to 90% in frequency, with an overall mean of 75%. And this high percentage obtains even in languages (like Spanish and Hungarian) where, say, both the definite and indefinite articles--obviously extremely frequent words--are vowel-initial. It will be noticed that Catalan and Finnish, on the list, are also given in (1) as having rules which treat # like a consonantal environment; furthermore, two languages that figure in (1) but were not included in the text-count study summarized in (11)--namely, Luiseño and Yawelmani--are reported by Kroeber and Grace (1960:9-10 vs. 249-251) and Newman (1944:26-27), respectively, to be entirely consonant-initial. In fact, the former state (p. 251) of the situation "never ... a vowel initially[;] if there is no other consonant, a glottal stop ... precedes the vowel": "This is a description made in almost all new [accounts] of western American [Indian] languages, until we wonder whether there are any exceptions." (Akan, Hausa, Korean, and Turkish--the remaining languages in (1)--will have to be studied and added to (11) at a later time.)

In regard to abutting segment-types, quite pronounced language-(family-)particular "preferences" can be discerned: Germanic easily allows or even favors C#C, while avoiding hiatus (V#V); Italian and the two African languages allow some hiatus, but shun C#C to an extreme degree, etc. The consistency of these favored juxtapositions within language-families (and their consistent differences across such families) seems anything but accidental, appearing to be, rath-

(11) Table 1

 TYPES OF ENVIRONMENT FOR # IN 20 LANGUAGE(-STATE)S
 (in per cent)

(LANGUAGE)	X#C	V#C C#V	C#C	V#V	X# [C] [-voice]	X# [C] +cor	X# [C] +ant -cor	X# [C] +bk
Hungarian	63	53	36	11	30	28	24	10
Finnish	76	47	39	14	44	40	22	14
(URALIC Mean)	(70)	(50)	(38)	(12)	(37)	(34)	(23)	(12)
Latin	72	43	46	11	41	32	23	16
Catalan	74	64	29	7	37	41	22	11
French	75	64	23	13	28	45	21	10
Italian	72	64	8	28	30	43	20	9
Rumanian	82	70	14	16	41	53	22	6
Spanish	69	56	25	19	38	48	12	9
(ROMANCE Mean)	(74)	(60)	(24)	(16)	(36)	(44)	(20)	(10)
Dutch	74	47	52	1	30	34	23	18
Old English	83	45	46	8	48	44	14	25
Middle English	72	52	44	4	40	36	23	13
Modern English	73	40	56	4	30	36	22	15
Middle High German	76	38	60	2	32	50	14	11
New High German	74	38	55	7	25	44	17	12
Gothic	79	41	52	7	38	41	19	18
Old Norse	71	44	47	9	45	30	18	23
(GERMANIC Mean)	(76)	(43)	(52)	(5)	(36)	(39)	(19)	(17)
Czech	84	62	30	8	50	54	27	2
Polish	82	62	28	10	34	48	29	4
(SLAVIC Mean)	(83)	(62)	(29)	(9)	(42)	(51)	(28)	(3)
Swahili (BANTU)	90	90	0	10	31	43	21	26
Yoruba (KWA)	67	67	0	33	27	40	17	11
(NIGER-CONGO/ -KORDOFANIAN Mean)	(78)	(78)	(0)	(22)	(29)	(42)	(19)	(18)
((OVERALL Mean))	((75))	((54))	((33))	((11))	((36))	((42))	((20))	((13))

er, a systematic part of the grammar of the language/s in question. Studies done according to the present method thus promise to contribute toward a way of characterizing exactly what it is that makes up the auditory impression of a (foreign) language, in addition to its phonetic and phonological inventory and its rules.

The (relatively) extremely small percentage of initial voiceless consonants (and, hence, of initial voiceless segments in general)--not over 50% in any language, and averaging only 36%--makes Lass's claim that final devoicing is (in general) induced phonetically by (voiceless) silence even more dubious. Thus, for example, in Germanic, even though Old English has the highest percentage of initial voiceless consonants, it had no final-devoicing rule (only an allophonic regularity which kept fricatives voiceless word-marginally), while German, with only 32% (MHG) or as low as 25% (NHG) of such consonants, does have a general final-devoicing rule, for all obstruents. (Although the picture changes somewhat for German if all #V-words are analyzed as phonologically #?V- (i.e., #C-) words, the relative disparity of it vis-à-vis OE does not, since there is evidence that OE vowel-initial words may have been preceded by a glottal stop, too--in alliterative OE verse, all vowels alliterate with one another, a fact which is explained if #V = #?V.)

Finally, it is salient that, of initial consonants, coronal ones always greatly outnumber non-coronal anterior ones, while the latter--with two exceptions (OE and Old Norse)--in turn outnumber back consonants. Thus, if # is (wrongly) to be given the features of a consonant, one could make a good case for its being [+coronal] (basically, dental or alveolar). Several people have mentioned to us a vague recollection that this has, in fact, been suggested in the literature, but we have as yet found no evidence to corroborate this. But the most salient result remains, of course, the one that #C is of a frequency of always over near two-thirds, averaging three-quarters and reaching even nine-tenths, since it then follows inexorably that #V is, relatively speaking, very infrequent, just as hypothesized.

7. Conclusion/s. Such, then, are the results of our study, which the present author intends to keep expanding to more languages and language-families. The size of the sample of text in each language has, admittedly, thus far been rather small (500+ words, 250 junctures), but the agreement across twenty languages is striking, and the almost exact match of our figures for #C in NHG with those determinable from Pfeffer (1964) (viz., 74% vs. 73%, respectively) provides further heartening confirmation. The data collected here give initial plausibility, at least, to the hypothesis that # (often) acts like a consonant in phonological rule-environments essentially because word-boundaries are usually followed by consonants.

Let no one mistake or misconstrue the claims made here: we are most emphatically not claiming that the language-learning child, for example, has access to the overall mean that we have arrived at cross-linguistically, or, in fact, even necessarily to the high percentage of #C in any language but his or her own. Our claim is simply that, on the basis of the fact that consonant-initial words predominate so overwhelmingly in every language studied here, without exception, it seems reasonable to expect a similar situation in other languages, es-

pecially ones in which there are phonological rules that treat # like a consonant, and, thus, that it seems likely that #C will be the prototypical initial in the speech that a child is likely to hear. As a sort of statistical hypothesis, the proposal made here is obviously not exactly subject to immediate refutation on the basis of one, or even a handful, of counterexample-languages. However, serious doubt would indeed be cast on it if a large number of languages should turn out to have huge percentages of vowel-initial words--and, more to the point, our claim that the high frequency of consonant-initial words is responsible for the quasi-consonantality of # would appear to be definitively falsifiable by any well-documented cases of a language that has mainly vowel-initial words (with no recent history of initial-consonant loss) but still contains a rule or rules referring to the disjunction {C, #}. It is thus quite clear how the present hypothesis can be tested synchronically.

But, lastly, we should also mention that diachronic evidence can additionally be brought to bear on this issue. Since # at pause obviously does not occur before a consonant, the hypothesis forwarded here predicts that prepausal # should not begin to behave consonantally until after the phonologization of rules referring to it (later); thus, historically, there should be a stage during which certain (phonetically motivated) processes occur at #C, but not before #|| (or before #V). The syllable-structure hypothesis, on the other hand, predicts that prepausal #--which undeniably coincides with \$, the end of a syllable--will act consonantally at precisely the same time as #C does (or begins to do so). In this regard, we explicitly want to separate word-final devoicing from other word-final processes. Ohala (personal communication) suggests that the only real phonetic motivation for such devoicing is that which is found prepausally, and that anticipatory (to some later phonological element) devoicing in other positions is the result of an extension of this beyond its phonetic basis; i.e., a phonologization. But, then, German final-devoicing, for example (which actually can be convincingly shown to be syllable-final; cf., among various relevant articles, Vennemann 1972:12-13, Hooper 1972:531-532, 539; and Janda in preparation), does not at all conflict with the major hypothesis of this paper--that the consonantal behavior of # begins preconsonantly--since devoicing is not a consonantly-conditioned process. There is, it should be mentioned, still another competing hypothesis which suggests that all #-sensitive processes generalize from prepausal #, but, since that proposal is at least partly orthogonal to the question of why word-boundary seems to have some consonantality to it, we have nothing further to say about it here, beyond merely mentioning it. In the event, it is clear where to look for evidence that will bear on, and help one decide among, this and the other hypotheses. Detailed instrumental-phonetic studies are called-for, in this respect, since it may well be the case that the perceptual salience of phenomena at prepausal (i.e., utterance-final) position perhaps distracts the ear from similar phenomena in other positions (by attracting so much attention to itself).

In sum, then: there is a growing body of evidence to support the present paper's proposal that word-boundary often acts like a consonant for the reason that it is usually followed by one, and addi-

tional relevant data will be easy to find, presumably in great abundance. But, until (such) differentiating evidence is found, our suggestion remains every bit as valid as the syllabic approach (both C and # as the locus of syllable-boundaries or margins), while both of these are, of course, clearly superior to the #-as-consonant approach, as we have seen previously. And, thus, it is certainly true that the plausibility of the new hypothesis invalidates any attempt to use a "how else?" argument (in the technical sense in which Rudolf Botha uses this term) to justify the old, other one as solely correct.*

Footnotes

* A number of people have made helpful comments, both positive and negative, from which this paper has greatly benefited, and I would like to thank them, here. The paper itself grew out of a term-paper written in March, 1978 for a proseminar on suprasegmentals taught by Steve Anderson, who deserves special thanks. It was later presented at a U.C.L.A. Linguistics Department colloquium in November, 1978; for comments made both then and later, I would especially like to acknowledge Raimo Anttila, Bill Bright, Ian Maddieson, Russ Schuh, Bob Stockwell, and Andreas Wittenstein, but also Ava Berinstein, Brent de Chene, Sandy Disner, Andrew Fox, Vicki Fromkin, Júlia Horváth, Peter Ladefoged, and George Papçun. Finally, I am grateful for beneficial discussions at BLS regarding this paper with Pierre Divenyi, Jeri Jaeger, Hector Javkin, Mark Mandel, Johanna Nichols, and John Ohala, as well as several people whose names I neglected to note, at the time. To the extent that I have taken any of their advice, one way or another, they really cannot be totally absolved from blame (or credit) for what appears here, but, on the other hand, there is no getting around the fact that the final responsibility is mine alone.

References

- Anderson, S. R. (1974). The Organization of Phonology. New York: Academic Press.
- _____ and P. Kiparsky (eds.) (1973). A Festschrift for Morris Halle. New York: Holt, Rinehart and Winston.
- Bach, E. and R. T. Harms (1972). "How Do Languages Get Crazy Rules?", in Stockwell and Macaulay (eds.), 1-21.
- Bell, A. (1977). "Segment Organization Phenomena and Their Explanations". Paper presented at the Symposium on Segment Organization and the Syllable; Boulder, Colorado; October 23-24, 1977.
- _____ and J. B. Hooper (eds.) (1978). Syllables and Segments. Amsterdam: North-Holland. (North-Holland Linguistic Series 40.)
- Brame, M. K. (1973). "On Stress Assignment in Two Arabic Dialects", in Anderson and Kiparsky (eds.), 14-25.
- _____ (ed.) (1972). Contributions to Generative Phonology. Austin: University of Texas Press.
- Braune, W./[revised by] W. Mitzka (1967). Althochdeutsche Grammatik (12. Auflage). Tübingen: Max Niemeyer. (Sammlung kurzer Grammatiken germanischer Dialekte. A. Hauptreihe Nr. 5.)

- Bruck, A., R. A. Fox, and M. W. La Galy (eds.) (1974). Papers from the Parasession on Natural Phonology. April 18, 1974. Chicago: Chicago Linguistic Society.
- Chomsky, N. and M. Halle (1968). The Sound Pattern of English. New York: Harper and Row. (Studies in Language [Series].)
- Gordon, E. V. (1957). An Introduction to Old Norse. (Second edition, revised by A. R. Taylor.) Oxford: Clarendon Press. (First edition 1927.)
- Hooper, J. B. (1972). "The Syllable in Phonological Theory". Language 48.3:525-540.
- _____ (1976). An Introduction to Natural Generative Phonology. New York: Academic Press.
- _____ (1977). "Syllabic and Non-Syllabic Issues in Phonology". Paper presented at the Symposium on Segment Organization and the Syllable; Boulder, Colorado; October 23-24, 1977.
- Hyman, L. M. (1975). Phonology: Theory and Analysis. New York: Holt, Rinehart and Winston.
- Janda, R. D. (in preparation). "A Final Proof That German Final-Devoicing Really Is Syllable-Final".
- Javkin, H. R. (1975). "Phonological Rules Influenced by Boundaries". Paper presented at the 90th Meeting of the Acoustical Society of America; San Francisco, California; November 3-7, 1975. (Abstract published in Journal of the ASA 58, Supplement 1 (1975):119.)
- Kahn, D. (1976). Syllable-Based Generalizations in English Phonology. M.I.T. Ph.D. dissertation distributed by the Indiana University Linguistics Club.
- Karttunen, F. (1970). Problems in Finnish Phonology. Indiana University Ph.D. dissertation.
- Kim, C.-W. (1972). "Two Phonological Notes: A-sharp and B-flat", in Brame (ed.), 155-170.
- Kisseberth, C. W. (1970). "On the Functional Unity of Phonological Rules". Linguistic Inquiry 1.3:291-306.
- Kroeber, A. L. and G. W. Grace (1960). The Sparkman Grammar of Luiseno. Berkeley: University of California Press. (University of California Publications in Linguistics, Volume 16.)
- Ladefoged, P. (1971). Preliminaries to Linguistic Phonetics. Chicago: University of Chicago Press.
- Langacker, R. W. (1972). Instructor's Key to Fundamentals of Linguistic Analysis. New York: Harcourt Brace Jovanovich.
- Lass, R. (1971). "Boundaries as Obstruents: Old English Voicing Assimilation and Universal Strength Hierarchies". Journal of Linguistics 7.1:15-30.
- Lees, R. B. (1961). The Phonology of Modern Standard Turkish. Bloomington: Indiana University Press, and The Hague: Mouton. (Indiana University Publications, Uralic and Altaic Series, Volume 6.)
- Lightner, T. M. (1972). Problems in the Theory of Phonology. Volume 1. Russian Phonology and Turkish Phonology. Edmonton: Linguistic Research. (Current Research into Language and Linguistics 6.)
- McCarthy, J. J. (forthcoming). "On Stress and Syllabification". To appear in Linguistic Inquiry.
- Newman, S. (1944). Yokuts Language of California. New York: Viking Fund Publications in Anthropology; Number 2.

- Pfeffer, J. A. (1964). Grunddeutsch: Basic (Spoken) German Word List: Grundstufe. Englewood Cliffs: Prentice-Hall. (Number One of the Publications of the Institute for Basic German, University of Pittsburgh.)
- Schachter, P. and V. A. Fromkin (1968). A Phonology of Akan: Akwapem, Asante and Fante. Working Papers in Phonetics 9, U.C.L.A.
- Stockwell, R. P. and R. K. S. Macaulay (eds.) (1972). Linguistic Change and Generative Theory. Essays from the UCLA Conference on Historical Linguistics in the Perspective of Transformational Theory. February, 1969. Bloomington: Indiana University Press. (Indiana University Studies in the History and Theory of Linguistics.)
- Vennemann genannt Nierfeld, T. (1971). "Natural Generative Phonology". Paper presented at the LSA Annual Meeting; St. Louis, Missouri; December, 1971.
- _____. (1972). "On the Theory of Syllabic Phonology". Linguistische Berichte 18:1-18.
- _____. (1974). "Words and Syllables in Natural Generative Phonology", in Bruck et al. (eds.), 346-374.
- Vihman, M. M. (ed.) (1977). A Reference Manual and User's Guide for the Stanford Phonological Archive. Part I. Stanford: Phonology Archiving Project, Department of Linguistics, Stanford University.
- Vogt, E. (1970). "Topics in Catalan Phonology". Harvard University B.A. essay.