# On the Relationship of Phonology and Phonetics: Comments on B. Lindblom 'On the Origin and Purpose of Discreteness and Invariance in Sound Patterns'[1]

*Sylvain Bromberger and Morris Halle* 1986

## Introduction

In his paper, Lindblom raises some very fundamental questions about the nature of the objects and properties investigated by phoneticians. Though some people may feel uncomfortable about the abstract character of such questions, it is important that they be boldly pursued. They are not only intrinsically interesting, they are also inescapable. Practice requires that we take a stand on them. By failing to do so openly, we do not avoid them, but simply accept some answers uncritically. And the wrong answers may lead us to work on the wrong problems, or to reject certain lines of investigations for the wrong reasons. Thus, though our comments focus on our disagreements with Lindblom's basic proposal, we would like to make clear at the outset that we applaud the spirit of his paper and that we agree with many of the things in it that we do not discuss.

## The Problem

In the first section of his paper, Lindblom reminds us of facts from phonology and from psycholinguistics that, in his words, "force the conclusion that speech is quantal, that it is featurally and segmentally structured." Then, in the second section, he points out that

(I) When articulatory and acoustic events are examined in the light of discrete and invariant linguistic units, the continuous and context-dependent aspects of such events tend to be enhanced. Combining the linguistic-psychological and phonetic-physical perspectives we obtain the classical but paradoxical picture of

> speech that underlies formulations of the invariance and segmentation issues:
>
> For a given language there seems to be no unique set of acoustic properties that will always be present in the production of a given unit (feature, phoneme, syllable ...) and that will reliably be found in all conceivable contexts (INVARIANCE ISSUE).
>
> The speech signal cannot be unambiguously segmented into temporally non-overlapping chunks corresponding to linear sequences of phonemes, syllables and words (SEGMENTATION ISSUE). (p. 504)

The problem, thus, is to come to terms with two apparently incompatible "pictures" of speech utterances: the phonologist's picture of these utterances as built out of a finite, hierarchical set of discrete phonological elements (features, phonemes, syllables, words), and the picture of the working phonetician according to which no finite set of phonetic elements corresponding to phonological ones can be recovered from a careful examination of speech tokens. More important, the problem concerns the relationship between the subject matter of phonology and the subject matter of phonetics.

How is one to think of that relationship?

There are basically two ways in which that relationship has been conceived, and we think that the problem requires us to choose between these two ways. We will call them respectively the *physicalist* and the *mentalist* view.

Before embarking on a discussion of these two points of view, we present a brief general remark. The study of speech sounds requires us to think of concrete utterances produced at a certain time by a certain person, i.e. of *tokens*, as instances of more abstract linguistic entities, i.e. *types*. Thus the acoustic event that would be produced at the end of an utterance would be an instance of the word-type 'types'. Sentences, phrases, words, syllables, etc. also fall into types. Each can fall under more than one type. In particular it can fall under a phonological type as well as e.g. under an acoustical type. To theorize about tokens we need a repertoire of types that tokens can instantiate. The difference between those who subscribe to the physicalist point of view and those who subscribe to the mentalist point of view is at heart a difference about the size and nature of that repertoire of types.



## The Physicalist Point of View

The physicalist point of view looks upon utterances as instances of only articulatory and acoustical types.

Students of language who adopt this point of view consider utterances (tokens) with two sets of questions in mind: (1) What are the configurations (and sequences of configurations) and activities of the speech organs responsible for their production? (2) What are the characteristics of the observed sound waves? And they look upon the utterances that share certain properties according to the first set of questions as instances of a same *articulatory* type, and upon the utterances that share certain properties according to the second set of questions as instances of a same *acoustical* type.

They believe that their task is to find acoustical and articulatory types peculiar to each language, to determine whether any of these types are common to all languages, to find laws about them, and to seek explanations for these laws.

More specifically, they assume that:

(II) (a) There exists a finite (relatively small) number of articulatory types such that each (normal) utterance will turn out to be a token of one of them or a sequence of tokens of some of them.
(b) And similarly for acoustical types.
(c) Utterances which belong to the same articulatory type belong to the same acoustical type.
(d) This follows from a very simple one-to-one relationship between the articulatory and acoustical characteristics or features on the basis of which types are defined, and hence there exists a set of *phonetic* types definable indifferently in either articulatory terms or acoustical terms.
(e) The rules of phonology can be expressed in terms of these characteristics, and therefore phonology can be done using phonetic types and does not require an ontology or ideology (in the sense of Quine) that goes beyond that required to define *phonetic* types.

Since each utterance has an indefinitely large number of articulatory and acoustical properties in terms of which types can be defined, it is obvious that (IIa–e) above will not hold for *every* choice among these

properties. But will it hold for *any*? Is there any set of articulatory or acoustical properties for which (IIa–e) hold? The answer, according to the empirical evidence available so far—and about which Lindblom seems to agree—is negative. This is the heart of the invariance and segmentation issues.

Mitigated forms of the physicalist point of view may replace (d) above with a more complicated proposition. But what characterizes the physicalist position is a consistent refusal to allude to any but articulatory and acoustical properties of tokens and a determination to theorize only about types defined by means of such properties. In other words, from a physicalist point of view, phonological types and features must be identical to phonetic ones; the relationship between them must be that of identity. Hence the segmentation and invariance issues pose a serious problem for phonology: since the traditional units of phonology cannot be defined in exclusively physicalist terms, a physicalist phonology must be created to replace the phonology of the past and it is by no means clear that this can be done.

## The Mentalist Point of View

The mentalist point of view starts from a somewhat different set of presumptions. The mentalist point of view takes seriously the obvious fact that (normally) speech utterances are produced by people intentionally, i.e. each utterance constitutes the carrying out of an intention to say something. The intention is of course very complicated: it is the intention to produce word sequences that have a certain meaning, a certain illocutionary force, a certain syntactic structure, etc, etc. But whatever else it involves, in the cases of interest here it is first and foremost the intention to produce sequences of words, i.e. an intention that can be carried out by moving one's vocal tract in certain ways.

The connection between such an intention and the way it is carried out is a tricky matter. But that is true of most intentions carried out by executing physical acts. Think of the many ways in which the intention of throwing a ball can be executed on any given occasion. Of course, it does not follow that just any type of gesture can count as the carrying out of such an intention. Wriggling one's nose will probably never count as carrying out the intention of throwing a ball. It would nevertheless be absurd to try to define the intention of throwing a ball by specifying in



purely physical terms certain trajectories of the arm. No set of necessary and sufficient conditions for being a case of throwing a ball intentionally can be given in purely macro anatomical and physical terms. (We insert the word "macro" because intentions may turn out to be realized as brain states, a matter about which we know nothing.) In other words, the intention to throw a ball is not the same intention as the intention to move one's arm through a certain specifiable trajectory, nor is there a specifiable trajectory such that only instances of it can count as the fulfillment of the intention of throwing a ball.

From a mentalist point of view, a speaker of e.g. English has a repertoire of phonological intentions. The intention to utter a certain word, or a sentence, or a longer discourse, is a complex intention—structured by rules—and made up of members of that repertoire. The production of an utterance consists in executing such a complex intention in the light of what the speaker knows about the conditions of utterance and of other desires and intentions that he may have at the time, e.g. the desire to be heard above noise, to save time, etc. as well as the preceding and succeeding intentions. To understand what has been uttered, the hearer must reconstruct the phonological intentions of the speaker. He uses the characteristics of perceived sound waves (as analyzed by his hearing apparatus and mind) and his knowledge of the rules, and his beliefs about the circumstances of utterance.

Articulatory phonetics, from the mentalist point of view, is the study of the gestures and activities produced in the vocal tract *by way of carrying out phonological intentions*, and acoustical phonetics is the study of the resulting sound waves. Thus subscribers to this point of view, like most phoneticians, are not interested in just any set of human movements (or movements in the vocal tract) that result in audible sounds, nor are they interested in the acoustical properties of all the sound waves that humans can produce. And they have a relatively clear criterion by which they demarcate their subject matter: the articulatory facts that interest them are those centrally involved in carrying out phonological intentions, and the acoustical facts that interest them are those involved in the recognition of phonological intentions. The relationship between phonological types (and features) and phonetic ones is not that of identity. It is an instance of the relationship that holds between intention types and physical types.

Given the mentalist picture, the segmentation and invariance issues do not come as a surprise. Nor do they represent the sort of threat that they represent to the physicalist point of view. On the contrary. From a

position that distinguishes between phonology and phonetics along mentalistic lines, one would not expect tokens that belong to the same phonological type also to belong to the same articulatory type (one might as well expect all deliberate acts of throwing a ball to involve identical arm trajectories) or the same acoustical type.

Whether acoustical types and articulatory types will coincide (i.e. whether every pair of utterances that comes under the same articulatory description also comes under the same acoustical description) is a question of physics, and involves a different relationship than the relationship with phonological types. The answer, for any interesting typology of utterances, is very likely to be negative. But it does not follow from this that no systematic connections exist between them. On the contrary. If the mentalist point of view is correct, then one should expect systematic connections: after all, articulatory types represent ways in which phonological intentions are executed, and acoustical types represent information on the basis of which these intentions can be recognized! But one should also expect these connections to be complex, and it is unlikely that they will be discoverable without appeal to phonological theory.

The mentalist's appeal to intentions and to phonological elements that are metaphysically distinct from phonetic ones strikes some people as suspicious. But this appeal involves nothing that has not always been implicit in the practice of people who study the sounds of language. In fact, physicalists among phoneticians are all closet mentalists. They all agree that not all acoustical outputs of the human vocal tract are within the province of their science: the sounds emitted in snoring or yawning, in blowing out a candle, or in gargling, though often indistinguishable from speech sounds, and though classifiable under some phonetic types (e.g. the sound of blowing out a candle is indistinguishable from what phoneticians would characterize as a voiceless bilabial fricative, and gargling sounds do not differ physically from uvular trills) are rejected by all phoneticians as outside the purview of phonetics. But on what grounds? Obviously, what systematically distinguishes the set of events that phoneticians study from the others is that the former represent intentions of the speaker to produce words in a language and the latter do not. Thus, at least in their practice the most tough-minded physicalists among phoneticians subscribe to something like the mentalist point of view. But then, why not incorporate that point of view in one's research and theorizing?

Note, by the way, that the very formulation of the issue requires one



to classify the units of speech in non-articulatory and non-acoustical terms. Lindblom, for instance, in the passage quoted in (I) above refers to 'discrete and invariant units'. But the units referred to cannot be acoustic types, since that would make his characterization of the invariance issue a self-contradiction. And the phonemes, syllables, and words that he mentions in his characterization of the segmentation issue cannot be acoustically defined phonemes, syllables, and words either, for similar reasons.

Let us then summarize.

The facts mentioned by Lindblom create a problem for the physicalist position. Adherents of that position try to extract from a corpus of utterances phonological features and types that accord with (IIa–e) above, and they have so far failed. Since they do not allow for the possibility that segments of utterances belonging to different phonetic types might nevertheless belong to the same phonological type (and vice versa), and since they do not allow for the possibility that phonological features might have more than one articulatory or acoustical realization, they are caught in a contradiction between the tenets of their position and the empirical evidence.

But these problems do not arise for the mentalist point of view; for that point of view makes a clear distinction between phonological elements and phonetic ones, and allows for the possibility that their joint exemplification in tokens may be governed by very complicated rules. The mentalist must, of course, recognize that little is known about these rules at this point and that much research remains to be done.

## Lindblom's Solution

In his paper, Lindblom cites a number of facts and makes a number of remarks that strike us as incompatible with a physicalist resolution of the invariance and segmentation issues. We have already pointed this out about his very formulation of these issues. Later he mentions evidence that leads him to "... the conclusion ... that solutions of the invariance and segmentation problems cannot a priori be expected to reside exclusively in the signal since experimental observations clearly show that listeners rely partly on the explicit stimulus contents, and partly on active topdown mechanisms capable of supplying missing, implicit information." And later again he draws the following implication from observations about speech production "... speech motor control is *not*

organized to generate strictly invariant and clearly segmented acoustic correlates of speech units."

In spite of this, he seeks a physicalist approach to the resolution of the invariance and segmentation issues. Thus he writes about the approach that he advocates: "... we shall assume that the invariance and segmentation issue are real problems only so long as we insist on keeping the phonetic and phonological point of view strictly separate. Under this assumption then, the conflict created by juxtaposing the linguistic and physical perspective could in principle be avoided if we were able to seek a new alliance rather than continue to promote the traditional two cultures represented by phonetics and phonology."

The alliance Lindblom seeks is essentially a physicalist alliance. It is not an alliance that strives for a theory showing how phonological units (intentions, as we think of them) receive different phonetic realizations (articulatory and acoustical) under different circumstances, or how the rules of phonology are reflected through these realizations, or how acoustical information can serve to reconstruct the phonological intentions responsible for their production.

It is an alliance which treats the units of phonology as illusory by-products of certain general desiderata that govern the process of vocal communication such as the need to keep different messages distinct, economy of effort, etc.. Lindblom likens phonemes and phonological features to the pillars and arches that can be seen in the nests of termites, and he points out that these structures are built by the insects not by following blueprints and plans in which pillars and arches figure explicitly (i.e. that termites lack the intentions to build arches and pillars). Rather the insects follow the procedure of depositing matter at points where there is a maximum concentration of a chemical (pheromone) that is contained in the deposits. "In the beginning the deposits are randomly distributed ... Somewhat later local peaks have begun to appear, serving as stimuli for further deposits that gradually grow into pillars and walls by iteration of the same basic stimulus-response process."

Lindblom comments that "in spite of the seemingly purposeful and intricate design of these nests termites cannot be assumed to have a 'mental blue-print' or 'target' for the finished product." He observes further that "although there is obviously more to how people build Gothic arches in cathedrals than how termites build arches in their nests, the termite story ... teach(es) a lesson relevant to theories of linguistics." The lesson, according to Lindblom, is that phonemes are not real components of utterances, nor part of the speaker's intention (in our terminology) but that like the arches of the termites, discrete phonemes arise "in a self-organizing way from an interaction between vocabulary growth and phonetic constraints."



Thus, according to Lindblom, the units of phonology are not only illusory aspects superimposed on reality by theory, but, like the arches of the termites (arches are arches, no matter how they might have come about) they are fully specifiable in physicalist terms. The method that generates them requires that they turn out to be describable in articulatory and acoustic terms, and the "alliance" that Lindblom advocates is one that identifies phonological types as phonetic ones. Though Lindblom does not say so explicitly, that is a clear implication of his approach.

Let us turn to that approach.

Lindblom postulates that the primitive vocalizations of an infant are made up of units which are "holistic patterns" that would resemble CV syllables if presented on a spectrogram. A lexicon is then acquired by the child through the application of the following algorithm:

(III)
1. Determine how many meanings you have to represent. Let that number be $k$.
2. Select next a syllable at random from a given inventory of $n$ possibilities. Let us call it the first syllable.
3. Select a next syllable so as to optimize certain performance constraints criteria.
4. Apply the previous step recursively until $k$ syllables have been obtained. (Let us call a set of $k$ syllables selected that way a $k$-set.)
5. Apply the steps above recursively until every syllable has served as a 'first' syllable at least once.
6. Pick the $k$ syllables that occur in the largest number of $k$-sets. (Let us call a set of $k$ syllables selected that way a final $k$-set.)
7. Assign a different meaning to each member of the final $k$-set so obtained.

Lindblom makes a number of claims about his algorithm that presumably have a bearing on the invariance and segmentation problem.

His first claim is that the algorithm is a plausible simplified model of how language is acquired by children and developed by language communities, i.e. a theory "that reflects the process of speech development and its transition into adult behavior ..." This has a bearing on the issues

because Lindblom thinks that "... every child solves the segmentation problem ...".

But note that the algorithm does not require information about the language community in which the child is doing his learning. Thus two children growing up in two different language communities that have encoded the same number of meanings (whatever that means) should, according to this theory, acquire the same language; and in general languages that do not differ about the number of meanings they encode will have exactly the same vocabulary (though perhaps with some permutations in the pairing of meanings and sounds.) Furthermore, changes in the number of items in the lexicon should have a drastic effect on the phonology of the language. Both of these implications fly in the face of the known facts of language acquisition.

Lindblom's second claim is that the algorithm does not presuppose features and segments, i.e. that a learner could apply it to a corpus of "gestalt trajectories through the phonetic (articulatory/perceptual) space" without analyzing the members of that corpus (presumably tokens) into segments and features. Lindblom writes "It should be clear ... that there are neither features nor segments in the generation of these phonetic signals and that it is our preceding linguistic analysis that imputes discrete 'segments' and 'features' to them. Just as 'arches' and 'pillars' are implicit in the behavior of termites, the 'segments' and 'features' represent phonological structures implicitly and nondiscretely present in the process of selecting the phonetic system." (We will come to that linguistic analysis in a moment.) The suggestion is that features and segments are undetectable in individual tokens but emerge only in the comparison of different tokens and that approaches to the invariance and segmentation issues should take that into account.

This claim strikes us as mistaken. Lindblom's algorithm—in particular the application of the performance criteria in the third step—crucially involves a comparison of the beginnings and endings of syllables. Among the parameters that Lindblom invokes is *perceptual salience*. This he describes as "extent of trajectory in auditorily motivated dimensions, e.g. a palatal closure followed by an [i] comes out as less salient than a palatal closure changing into [u]." To determine the "salience" of a given syllable it is thus necessary to compare its beginning with its ending. But this implies that there is some way of separating the beginning of a syllable from its ending. Thus the segmentation of the syllable is not an *outcome* of the algorithm, but a *condition* for its applicability! The case is quite unlike the case of arches and columns



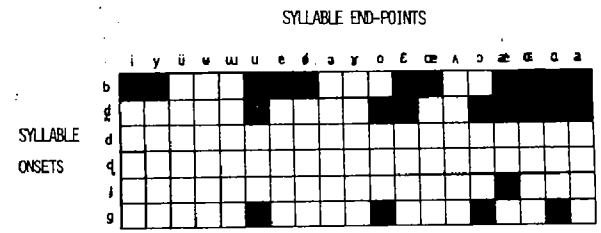
built by termites. Arches and columns can be the outcome of a process that can be described without presupposing anything about the shape of the outcome. But the algorithm in III cannot be thought to have any application without presupposing certain things about its input, and in particular that its input is segmented into phoneme-like units.

The third—and from our point of view the most important—claim that Lindblom makes for the algorithm concerns the final $k$-sets it yields. These $k$-sets are held to be themselves amenable to a linguistic analysis that produces quasi-features and quasi-segments for which the invariance and segmentation issue do not arise.

A brief review of that analysis may be helpful. The algorithm that we have described above is first applied to an inventory of discrete syllable-like elements describable in phonetic terms. (Lindblom must assume here—contrary to fact, and contrary to his own statement quoted above—that the segmentation issue does not arise for syllables, but only for units shorter than syllables.) The algorithm assumes that every syllable in the original inventory can be characterized physicalistically in terms of its two termini, its beginning and its ending. In the paper, Lindblom recognizes seven beginning and nineteen ending types, which yield a total repertory of 133 primitive syllables. He represents these termini not with integers or some other arbitrary symbols, but rather with letters drawn from the International Phonetic Alphabet. Thus the seven beginnings are represented with the voiced stop symbols given along the ordinate of the chart (IV) below which we reproduce from Lindblom's paper (top part of Fig. 23.3), and the nineteen endings with the vowels given along the abscissa of the chart.

The algorithm (III) is then applied to select the needed number of distinct syllables from the repertory of 133. It turns out that for $k = 24$ (which one might suppose to represent a fairly early stage in the child's course of language acquisition) the syllables are those indicated with black squares in the chart.

(IV)



The "linguistic" analysis is now applied to that chart. It proceeds from the fact that the syllables selected form rows and columns rather than diagonal lines (or arrangements that can be turned into diagonals by reordering the points along the axes.) "Rows and columns with multiple entries contain syllables that keep one segment constant while varying the other. They identify *minimal pairs*. Since by definition all syllables have distinct meanings, we must conclude that according to standard procedure the minimal pairs contain distinct phonemes. The existence of [bu du gu] thus appears to suggest that in these derivations [b, d, g] come out as separate *phonemic segments* ..." And the analysis proceeds in this way to create quasi-segments. Nothing in the analysis requires departure from the physicalist approach. The types generated by this method are physicalist types.

We do not believe that Lindblom's third claim can be sustained either. Remember that the solution of the problem requires one to find a set of phonological features and segments (at the type level) and a set of phonetic features and segments (at the type level again) such that any two utterances that can be analyzed as containing tokens of the same member of the first set can be analyzed as containing tokens of the same members of the second set. Without going into the details of Lindblom's analysis, we need only notice that the quasi-phonological units generated by that analysis are physicalistic units. As we have seen in chart (IV) above, Lindblom has identified these units with elements from the International Phonetic Alphabet. But that defeats the claim, for if anything is clear, it is that the issues of invariance and segmentation arise most sharply for the elements of that alphabet.

Put very succinctly, our objection runs like this: if Lindblom's method could generate elements that "solve the problem" then the problem would not have arisen. It has arisen. Therefore the method cannot generate elements that solve the problem.

Of course Lindblom could reply that he could have started with a different "inventory" and thereby avoided our objection. But that leaves the matter up in the air until a proper inventory is discovered. We do not believe that such an inventory exists.

## Conclusion

In this paper, Lindblom invites us to develop a new perspective on the invariance and segmentation issues and to add to the set of facts deemed



relevant to their solution. We think that this is an invitation that must be taken very seriously. We also agree with him that pragmatic facts will turn out to be important. However we believe that they will not suffice. We believe that serious work on the theory of intentions and the execution of intentions will also be necessary. Little is known about intentions. We will have to grope in the dark for a while. But that cannot be helped.

## *Note*

1. In Joseph S. Perkell and Dennis H. Klatt, eds. *Invariance and Variability in Speech Processes* Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers, 1986, pp. 493–510.

# Why Phonology is Different

*Sylvain Bromberger and Morris Halle* 1989

## 1. Introduction

Until well into the 1970s research in generative syntax was based on the presumption that the syntactic component of a grammar should consist of ordered transformations that derive surface structures (and later logical forms) from deep structures (themselves generated by phrase structure rules) through intermediate structures. Though there were discussions on whether the ordering of transformations peculiar to each language was entirely the consequence of universal principles or was in part language-specific, and though questions were raised about the significance of intermediate structures, there was general consensus about the basic premises. Even a cursory perusal of the journals of the period or of such influential textbooks as those by Akmajian and Heny (1975), Baker (1978), or Perlmutter and Soames (1979) shows that problems concerning the right ordering of transformations and their applicability to intermediate structures were at the center of research and teaching. This earlier state contrasts dramatically with the situation in syntactic theory in more recent times, especially since the publication in 1981 of Chomsky's *Lectures on Government and Binding*. Questions about the ordering of transformations and about intermediate representations have all but disappeared from syntax—at least in the versions of the theory that have accepted the Government-Binding framework and its later developments.

This course of events obviously raises the question whether phonology should not undergo a similar development—that is, whether phonological theory should not be restructured in such a way as to exclude rule ordering and representations that are neither underlying representations nor surface forms.[1] Indeed, there have been a number of attempts in recent years to reformulate phonology without recourse to extrinsic rule order, strict cyclicity, and so on; see, for example, Kaye and Lowenstamm (1986) and Majdi and Michaels (1987).

Reprinted with permission of MIT Press from *Linguistic Inquiry*, 20, 51–70, 1989.



We shall argue here that derivations based on ordered rules (that is, external ordering)[2] and incorporating intermediate structures are essential to phonology—in other words, that they represent an uneliminable aspect of linguistic knowledge. Some—though not all—of our arguments will turn out to be updated versions of original arguments advanced in support of extrinsic ordering in phonology (see, for example, Halle (1962)), for many of these appear to be no less sound now than they were a quarter of a century ago when they were first advanced. The crux of our position is that facts pertaining to the two domains—phonology, on the one hand, and syntax and semantics, on the other—are of a very different nature and that there is therefore no reason to assume a priori that they must be covered by formally similar theories. Whether the theories are or are not similar is a contingent matter to be settled in the light of the evidence, and the evidence, as far as we can tell, indicates that they are not formally similar and that the structure of phonology is best thought of as that of a deductive system.

Syntax/semantics as practiced in the 1980s is primarily concerned with the conditions that the deep structure, surface structure, and logical form of a sentence must satisfy.[3] These include conditions peculiar to each level as well as conditions across levels. But the representation of a sentence at each level encodes information about the sentence (thematic role assignments, binding relations, sequential order, relative scope of operators, meaning, and so on) that is distinct from what a speaker must know in order to articulate a token.

As now understood, each of these three representations is assembled from words and other items stored in the lexicon in a manner instructively similar to the assembling of pieces of a three-dimensional jigsaw puzzle. Just as in the case of the jigsaw puzzle, the overriding considerations are whether—and how—the pieces fit together and that at the end there be no holes in the assembled shape nor any pieces left over. But the order in which the pieces are assembled does not matter. To extend the analogy, we may think of the relationship among the three representations as being like the relationship among the different faces of a tetrahedron. The three representations, like the separate faces of the tetrahedron, are distinct from each other and abstractable from the whole. Yet, like the separate faces of the tetrahedron, they share elements and thereby impose limits on each other. One might even go so far as to say—though at the price of oversimplification—that the sharing of elements is expressed by the celebrated single transformation Move α, whereas the distinctness of the representations is expressed by the fact that Move α

must respect conditions peculiar to each level of representation, such as the θ-Criterion and the Empty Category Principle. If this picture of the interrelation among syntactic representations is correct, then clearly there is no theoretically significant ordering among the principles that govern the interconnections among the three representations and there is no interesting sense in which any of the representations is "derived" from any of the others through a sequential application of rules and intermediate representations, just as there is no significant ordering among the mutual constraints between the faces of a tetrahedron and no derivation with intermediate forms of one face from another.[4]

Phonology, on the other hand, is primarily concerned with the connections between surface forms that can serve as input to our articulatory machinery (and to our auditory system) and the abstract underlying forms in which words are stored in memory. Whereas syntax is concerned with the relations among representations that encode different types of information requiring different types of notation, phonology is concerned with the relationship between representations that encode the same type of information—phonetic information—but do so in ways that serve distinct functions: articulation and audition, on the one hand, and memory, on the other. Since underlying phonological representations of words are stored in speakers' permanent memory, whereas phonetic surface representations are generated only when a word figures in an actual utterance, there is a clear and theoretically significant sense in which underlying representations are prior to surface representations, a sense that justifies thinking of the surface form as "derived" from the underlying form. This fact in turn brings up the question of the manner in which surface representations are derived from underlying representations. The answer clearly is to be decided by looking at the actual contingent evidence rather than by reflecting on a priori logical or methodological necessities. In particular, there is no a priori reason to hold that the derivations must be subject to a theory formally similar to the theory appropriate for syntax/semantics. We therefore turn next to an examination of some of the relevant evidence.

## 2. The Synchronic Evidence

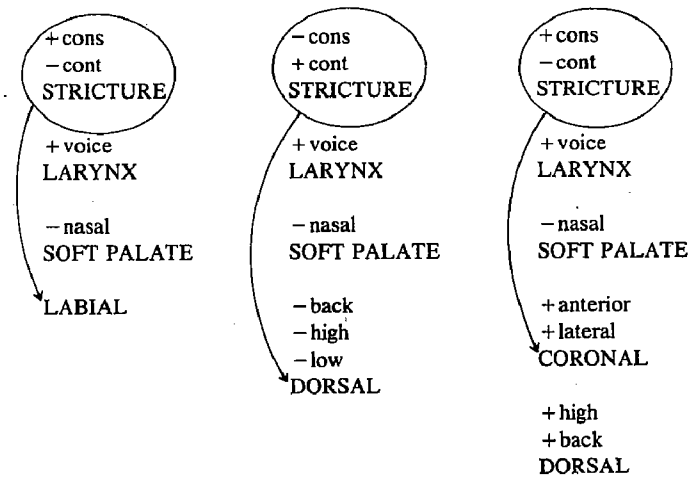
The phonological surface representation must encode how a word is pronounced. It must serve as input to our articulatory machinery.



As a first approximation we shall assume—in conformity with a well-supported tradition in phonology—that the representations required for the articulation of different words are given in the form of stipulations of discrete sound segments concatenated in the order in which they must be produced. Thus, the English word *bell* is represented by a sequence of three symbols, of which the first stands for the plosive [b], the second for the vowel [ɛ], and the third for the lateral [l].

An important result of the research of the last fifty years has been to establish the proposition first advanced by Jakobson (1938) that speech sounds are composite entities constituted by complexes of binary phonetic features such as voicing, nasality, and aspiration. As a first approximation we may think of this as an interpretation of the alphabetic symbols of the phonetic alphabet. Thus, instead of the sequence [bɛl] we write (1). (It should be noted that the feature system employed here is that developed by Sagey (1986) with further modifications due to Halle (1992) and differs quite markedly from feature systems utilized in earlier publications by the present authors.)

(1)




The representation in (1) encodes the information that enables a speaker to produce the sound sequence [bɛl]; that is, (1) specifies the vocal tract gymnastics necessary for uttering the word *bell*. This vocal tract gymnastics is performed by a small number of movable structures—in effect,

six—known as *articulators*, which are represented in (1) by capital letters. Each articulator has a small repertoire of distinct (linguistic-phonetic) behaviors known as *features*, which are represented by lowercase letters. These behaviors select between binary sets of options, represented by coefficients whose value is either *plus* or *minus*. Thus, we can move the DORSAL (tongue body) articulator either backward toward the rear wall of the pharynx (in response to the feature specification [+back]) or forward, away from the pharynx wall (in response to the feature specification [−back]). The absence of any other feature specification reflects the fact that in producing speech sounds no indication of degree of movement is linguistically significant.

Most features can be actualized only by a single articulator. Thus, [nasal] is always implemented by the SOFT PALATE, [voicing] by the LARYNX. This fact is encoded graphically in (1) by grouping the different features and placing them above their articulators. These *articulator-bound* features contrast with *articulator-free* STRICTURE features such as [consonantal] and [continuant], that is, with features that can be executed by any one of the LABIAL, CORONAL, or DORSAL articulators. Since the choice of the appropriate articulator reflects a linguistically relevant distinction, this choice must be encoded in the representation of the sound. We have indicated this graphically in (1) by means of the arrows connecting [consonantal, continuant] to the articulators appropriate to each sound. It is obvious that not every articulator is—or need be—actively involved in the production of every sound. For example, the tongue blade (CORONAL) and tongue body (DORSAL) articulators play no role in the production of the English consonant [b]. This is encoded in (1) by omitting mention of these articulators in the representation of [b]. In similar fashion, a given articulator, though active in the production of a particular sound, may not execute in a linguistically significant way all features of which it is capable. For example, when producing consonants, English speakers do not deliberately round the lips or spread them: the feature [round], a behavior of the LABIAL articulator, therefore does not figure in the representation of any English consonant, even of one like [b] that requires active involvement of the LABIAL articulator.

As noted, (1) is a surface representation of the English word *bell* in that in principle it provides the information needed by a speaker to produce this word correctly.[6] In addition to surface representations such as (1) words also have abstract underlying representations (that is, representations that encode the form in which words are stored in memory).



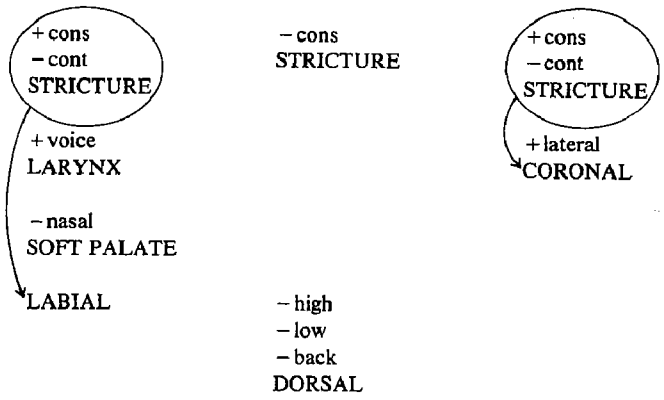
We must now elaborate on this. Utterances are, to a first approximation, sequences of word tokens produced one after another. But speakers can produce an utterance only if they know the words of which it is composed. But what does it mean for a speaker to know a word? At a minimum, it means that the speaker knows that a given sequence of speech sounds is a word in his or her language. For instance, speakers of English know the words [boy] and [bel] but not [naʕar] and [paʕamon]. It is for this reason that under normal circumstances [boy] and [bel] may figure in their utterances, but not [naʕar] and [paʕamon].

However, this sort of knowledge is not innate. It must be acquired and retained—as is obvious from the fact that children raised in different language communities acquire different lexicons. Thus, learning a language involves—among other things—registering a long list of words in memory. There is good reason to assume that speakers represent words in their memory by means of a code that is directly related to ways of producing linguistic sounds and that words are stored in memory as sequences of items in such a code. Specifically, *boy* and *bell* are not represented in memory by a numerical code where some arbitrary numeral like 797 stands for the former and some other arbitrary numeral like 2593 for the latter. Rather, the symbols in memory stand in a direct relation to the production of sounds, so that, for instance, *boy* and *bell* are both represented by three (complex) symbols, of which—in these two examples—the first are the same, and the other two different.

Not all of the information required for producing a word phonetically is needed by speakers for storing the word in memory and for retrieving it when the occasion arises, because a significant fraction of that information is predictable through general rules and principles that govern the pronunciation of English and that are also part of the speaker's knowledge of his or her language. For example, in English all vowels and the lateral [l] are invariably [+voiced] and [−nasal]. Moreover, the behavior of the tongue body (DORSAL) in the lateral is governed by special rules that depend on the phonetic context. Finally, in vowels the articulator-free features [consonantal] and [continuant] are universally implemented by the DORSAL articulator. Moreover, vowels are always [+continuant]. When information recoverable through these rules is eliminated from (1), the word *bell* is represented as in (2).

If, as implied above, memory storage and search time are at a premium in the case of language, then (2) will serve as an effective under-

(2)




lying representation of the word *bell*, since in (2) information that is retrieved through the general rules of the language is systematically omitted. It is important to notice that the omitted information is absolutely crucial for the correct phonetic realization of the word and that it must therefore be accessible to the speaker. Thus, both representations (1) and (2) play a significant role in accounting for our ability to speak.[7]

The rules supplying the information missing in the underlying representation must be applied in a definite order. As evidence, consider the English rules of syllabification and of stress assignment. Both of these rules provide information essential for the production of English words and account for the fact that this information is completely predictable and must therefore not appear in the underlying representations. In the overwhelming majority of English words stress is assigned to the (ante)-penultimate syllable S* if the following syllable S** has a nonbranching core; otherwise, stress is assigned to S**. (For details, see Chomsky and Halle (1968) and Halle and Vergnaud (1987).) Since stress assignment thus depends on whether or not certain syllables have a branching core, stress cannot be assigned until the word has been syllabified. But the syllable structure of an English word in turn is totally predictable from the sounds that compose the word. In short, both syllable structure and stress are predictable; therefore, they do not appear in the underlying representation but are introduced into the surface representation as a result of the application of certain rules. But the rules assigning syllable structure and those assigning stress are distinct rules since they affect different aspects of the representation and do not always operate in tandem. Moreover, the stress rules must apply after the rule of syllable



structure assignment since the stress rule requires information that is not present until syllable structure has been assigned.

The rules discussed so far are rules that add features omitted from underlying representations. But underlying representations can not only differ from surface representations in containing fewer specified features. They can also differ in assigning different values (coefficients) to features present in both. This difference between the two representations is ultimately a consequence of the fact that like all physical systems the individual articulators are subject to inertia and their movements are influenced by their earlier positions and movements and by simultaneous movements and positions of other articulators. Although these contextual effects have their origin in mechanical factors, they achieve certain articulatory optimizations that are brought about in different ways by different languages. They are thus not mere effects of the physics or physiology involved. They are brought about through the application of language-specific rules—rules that speakers acquire in the course of their linguistic maturation and that are part of their knowledge of their language.

A typical example of this sort is the process of English colloquial speech that turns intervocalic [t] and [d] into a voiced flapped stop. The main effect of flapping is to eliminate the distinction between /t/ and /d/ in certain contexts; as a result, utterances (words) that differ in their underlying representations become phonetically indistinguishable, as illustrated in (3).

(3) plotting—plodding
wetting—wedding
butting—budding

In many dialects flapping takes place on some occasions and not on others. In some dialects, however, flapping is institutionalized so that it is applied consistently by speakers and failure to flap is perceived in such dialects as affectedness or as "putting on an act." We shall assume here that in such dialects flapping is in part due to a rule that somewhat informally is stated in (4).

(4) $\begin{bmatrix} -\text{cont} \\ \text{CORONAL} \end{bmatrix} \rightarrow [+\text{voiced}]$

in env. $\begin{bmatrix} -\text{cons} \\ \text{stressed} \end{bmatrix} \text{——} \begin{bmatrix} -\text{cons} \\ \text{unstressed} \end{bmatrix}$

As a result of (4), underlying voiceless /t/ in certain environments is phonetically implemented as voiced; or, put differently, in certain contexts /t/ has a different specification for the feature [voice] in underlying representation than in surface representation.

A striking feature of many Canadian dialects of English is the implementation of the diphthongs [ay] and [aw] as [ʌy] and [ʌw] in position before voiceless consonants. We exemplify the contrasts in (5) and give an informal statement of the rule responsible for them in (6).

| | | | | | |
|---|---|---|---|---|---|
| (5) | a. | r[ay]z | r[ʌy]ce | r[aw]se | m[ʌw]se |
| | | tr[ay]be | tr[ʌy]pe | cl[aw]d | cl[ʌw]t |
| | b. | r[ayD]ing | wr[ʌyD]ing | cl[awD]ed | sh[ʌwD]ed |

(6) $[-\text{cons}] \rightarrow [-\text{low}]$ in env. $\underset{\text{stressed}}{\_\_\_}$ $[-\text{voiced}]$

A fact of special interest is that in most Canadian dialects that are subject to both rules (4) and (6), (6) applies only to words with underlying /t/, not to words with underlying /d/. We have exemplified this in (5b).

We can predict this result if we assume that (6) is ordered before (4) and that application of a phonological rule is subject to Principle (7).

(7) Phonological rules are ordered with respect to one another. A phonological rule R does not apply necessarily to the underlying representation; rather, R applies to the derived representation that results from the application of each applicable rule preceding R in the order of the rules.

There are Canadian dialects in which (6) does not apply in words of the type illustrated in (5b) but in which the pronunciation of the words in (5a) is the same as their pronunciation in other dialects. As noted in Chomsky and Halle (1968), these dialects differ from the others in that (4) is ordered before (6), rather than after it.

It is worth noting that Principle (7) was not needed to account for the order in which the rules of syllabification and stress assignment are applied in English. That ordering did not need to be explicitly stipulated. It could be achieved by the simple proviso that a rule applies whenever conditions for its application are satisfied. Principle (7) is needed if conditions for the application of more than one rule are satisfied simultaneously. The order of application then—as the Canadian example shows—becomes a language-specific matter. The validity of Principle (7) in phonology and its absence from syntax/semantics is one revealing manifestation of the fact that the representations treated by phonology differ in nature from those treated by syntax/semantics.



Rule ordering is one of the most powerful tools of phonological descriptions, and there are numerous instances in the literature where the ordering of rules is used to account for phonetic effects of great complexity. Until and unless these accounts are refuted and are replaced by better-confirmed ones, we must presume that Principle (7) is correct. If we are right in doing so, then Principle (7) is also one of the major features that distinguish syntax/semantics from phonology.

We have presented instances where the surface representation is derived from the underlying representation by the application of several ordered rules. It is, of course, possible to account for all of the empirically observed facts of phonology without rule ordering. Since the number of words stored in the memory of a fluent speaker is relatively small (hardly ever exceeding 100,000 items), it is in principle possible to account for the pronunciation of each word by a separate rule. Such an approach, however, would be grossly implausible since it would exclude rules like (4) and (6), which speakers clearly know—as shown, for example, by the fact that when presented with written words they have not encountered before, speakers pronounce these in conformity with (4) and (6). If every word were acquired with its own rule of pronunciation and if speakers knew no phonological rules, then speakers would not know how to pronounce words they had not previously encountered and there would be no reason to expect them to pronounce new words in a way that corresponds systematically to the way they pronounce other words.

Another logical possibility where Principle (7) would play no role is that all rules apply to underlying representations and that the relation between the underlying representations and the surface representations is therefore never mediated by derivations made up of intervening forms. We believe that this possibility is rather implausible in the light of the following sort of evidence. Consider the second Canadian dialect mentioned above, in which the contrast between *riding* and *writing* is systematically eliminated—that is, the dialect in which, according to the account presented above, rule (4) is ordered before rule (6). If rules were applied to the underlying representation only, then instead of (4) we would need a rule such as (8) to account for the facts.

(8) [−cons] → [−low] in env. ___ [−voiced]
stressed

*but not* in env. ___ t V

This rule is more complex than (4) since it includes an exception stated in the "but not" clause. The inclusion of this clause is motivated solely by the theoretical decision to drop Principle (7). But note that Principle (7) purports to be a universal principle, in other words, a principle of universal phonology. It should thus be viewed as something that does not have to be acquired but is part of the innate endowment of potential speakers. Rules like (4), (6), and (8) must be learned separately. Complex rules with exception clauses are evidently more difficult to discover in a random corpus than are exceptionless rules.[8] Thus, the hypothesis that rules like (8) are acquired rather than rules like (6) under the guidance of Principle (7) is much more difficult to reconcile with the known ease and rapidity with which children learn to speak their dialect, and that hypothesis is therefore much less plausible.

It has been known at least since Chomsky and Halle (1968) drew attention to this fact that the strict linear order of rules implicit in Principle (7) is not maintained everywhere. These deviant rule orderings are predictable in the sense that they occur only when specific conditions are met; they are therefore not violations of the principle of linear rule order but rather extensions of the principle.

The most important of these extensions are the following three. First, if A and B are two rules, and the conditions for the application of A include all the conditions for the application of B, but not vice versa (in other words, if the application of A is subject to more conditions than the application of B), then A is ordered before B, and B cannot apply to any string to which A has applied. This type of *disjunctive* rule ordering has been studied by Kiparsky (1973); see also Myers (1985), Halle and Vergnaud (1987), and Mahajan (1988).

Second, the order of application of some rules is determined by the internal constituent structure of words. This is the famous *cyclic* order of rule application, which has provoked some of the most ingenious work in modern phonology.[9]

Third, every phonological rule must be assigned to one or several blocks or *strata*, and the strata to which a given rule is assigned determine whether it applies cyclically to the immediate constituents of a word or whether it applies only once to the entire word.[10]



When fully specified so as to incorporate these three extensions, Principle (7) is exceptionless.

## 3. The Diachronic Evidence

Further evidence for the psychological reality of ordered rules (and hence for derivations) in phonology is provided indirectly by the phenomenon of diachronic sound change. Research on sound change began in the nineteenth century as an attempt to account for the observation that in Sanskrit, Greek, Latin, and a number of other languages cognate lexical items exhibit widespread and systematic phonetic resemblances. It had been suggested by Sir William Jones, a high official in the British civil service in India, that the striking resemblances among cognate words in these languages, spoken in widely separate geographic locales, cannot have arisen by accident and that the only plausible explanation for them is that these languages all descend from a common protolanguage. Nineteenth-century linguistics adopted this proposition and devoted its major and best efforts to displaying in detail the phonological regularities that link the different Indo-European languages to their protolanguage.

By the end of the nineteenth century the phonological system of the Indo-European protolanguage had been reconstructed in a surprisingly convincing way. A crucial aspect of this reconstruction was the postulation of "sound laws" relating earlier stages of the language to later stages.

Consider for instance the first part of Grimm's Law, surely one of the most securely established of all "sound laws," which accounts for phonetic correspondences between the words of Germanic on the one hand and those of the other Indo-European languages, such as Greek, Sanskrit, Latin, and Baltic, on the other. The "law" consists of three distinct parts, of which the first, which is of special interest here, can be stated formally as in (9a); the evidence for it is found in correspondences such as those in (9b).

(9) a. $\begin{bmatrix} -\text{cont} \\ -\text{voiced} \end{bmatrix}$ → [+cont] except after obstruent

b.

| Germanic | Greek | Latin | Sanskrit | Baltic | |
|---|---|---|---|---|---|
| fot | pod | ped | pad | ped | 'foot' |
| θre | tri | trē | tray | tri | 'three' |
| xund | kun | kan | švan | sun | 'dog' |
| naxt | nukt | nokt | nakt | nakt | 'night' |

Formally this "law" is indistinguishable from a phonological rule such as (10a), which accounts for the fact that English [p t k] must be aspirated in the words in (10b) but unaspirated in the words in (10c).

(10) a. $\begin{bmatrix} -\text{cont} \\ -\text{voiced} \end{bmatrix} \rightarrow$ [+asp] at the beginning of a stressed syllable

b. pill, till, kill

c. spill, still, skill, soapy, naughty, shaky

(9a) and (10a) have exactly the same format and differ only in the features indicated to the right of the arrow—that is, in the features affected and in the respective contexts in which the rules apply.

This formal similarity could be viewed as a mere coincidence. However, there is a much more plausible explanation, namely, that the addition of phonological rules to a language is the main mechanism responsible for phonetic change. According to this explanation, lawlike phonetic change occurs when speakers add a new rule to their language. The character of the diachronic "sound law" then follows trivially from the character of the added rule, since it simply reflects the latter's operation. On this view, then, the first part of Grimm's Law given in (9a) describes a diachronic change of forms brought about by the fact that later speaker/hearers had (9a) in their phonology whereas earlier speaker/hearers did not.[11]

If we accept this explanation—and the arguments in its favor are very strong—then information about diachronic linguistic change yields information about the rules in the synchronic phonology of certain speaker/hearers.

The question now arises whether such information can also tell us anything about rule ordering.

In order to answer this question, it is necessary to recall that there is a second part to Grimm's Law, which can be formally stated as in (11a) and which accounts for the correspondences between Germanic and the other Indo-European languages illustrated in (11b).

(11) a. $\begin{bmatrix} -\text{cont} \\ -\text{asp} \end{bmatrix} \rightarrow$ [−voiced]

b.

| Germanic | Greek | Latin | Sanskrit | Baltic | |
|---|---|---|---|---|---|
| two (Eng) | duo | duo | dva | divi (Latv) | 'two' |
| yoke (Eng) | dzugon | yugum | yugan | yungas (Lith) | 'yoke' |



Grimm's Law thus produced the two sets of changes illustrated in (12): those in (12a) are due to rule (9a), and those in (12b) are produced by rule (11a).

(12) a. p → f  t → θ  k → x

b. b → p  d → t  g → k

Is there any reason to believe that these two sets of changes were ordered so that the set in (12a) applied before the set in (12b)?

Bloomfield (1933, 368) thought that there was, and his reasons are interesting:

> ... it is clear that in pre-Germanic time, the Primitive Indo-European [b, d, g] can have reached the types of Primitive Germanic [p, t, k] only *after* Primitive Indo-European [p, t, k] had already been changed somewhat in the direction of the types of Primitive Germanic [f, θ, h]—for the actual Germanic forms show that these two series of phonemes did not coincide.

Bloomfield assumed rightly that if a language had first undergone the change (12b) and then the change (12a), the effect would have been to turn both [p] and [b] into [f], [t] and [d] into [θ], and [k] and [g] into [x], contrary to known facts about Germanic. In the quoted passage Bloomfield was, of course, talking about diachronic ordering, not about ordering of rules in a synchronic Germanic phonology. However, if we assume that the mechanism of rule addition is responsible for the diachronic facts, then Bloomfield's considerations can be turned into reasons for holding that (12a) was ordered before (12b) in the synchronic phonology of Germanic speakers, since, by the same reasoning, the reverse order would also have had the false consequences just described.

So it would seem that evidence from language change does show that the two parts of Grimm's Law must be ordered in the phonology of Germanic. Unfortunately, the evidence, as it stands, is inconclusive. It does not rule out another possibility, and it is noteworthy that neither Bloomfield—nor to our knowledge any other student of sound change—ever entertained it. This possibility is that both sound changes apply to underlying representations directly. Viewed synchronically, this possibility comes down to a denial of Principle (7), at least for rules that bring about linguistic change. Under that hypothesis such rules would not be ordered at all. Since no rule would then have any effect on the input to any other, that would be compatible with the facts that led

Bloomfield to order (12a) before (12b). In other words, these facts tell us how Grimm's Law/rules are ordered in the phonology of Germanic *if they are ordered*, but they do not tell us *that* they are ordered. We therefore also need evidence demonstrating that rules responsible for diachronic change abide by Principle (7); in other words, that they do not apply exclusively to underlying representations.

Such evidence is provided by Verner's Law, formally stated as (13).[12]

(13) [+cont] → [+voiced] after unstressed vowel

Verner's Law is generally believed to have come into the language after Grimm's Law (9a). The evidence adduced for this ordering is that Verner's Law applies not only to the continuant /s/ (which Germanic inherited unchanged from proto-Indo-European) but also to continuants that have appeared as a result of (9a). That evidence, conjoined now in the familiar way with the hypothesis that "sound laws" are the effect of the addition of phonological rules, unlike the earlier evidence, does constitute a conclusive argument for the view that in the phonology of Germanic, Verner's Law operated after (9a). This is so because the new evidence shows that (13) must apply to some outputs of (9a), whereas the evidence used by Bloomfield showed that (12b) may not apply to outputs of (12a). It was evidence against one way of ordering rules, not for ordering them in a certain way.

However, it might be objected, as long as we restrict ourselves to diachronic evidence, (13)—Verner's Law as usually stated—is not the only way to describe the facts. The changes it describes can also be described with a different, more complicated rule that applies to underlying representations, namely, (14).

(14) $$\left\{\begin{array}{ll} [+\text{cont}] & \rightarrow [+\text{voiced}] \\ \begin{bmatrix}-\text{cont}\\-\text{voiced}\end{bmatrix} & \rightarrow \begin{bmatrix}+\text{cont}\\+\text{voiced}\end{bmatrix} \end{array}\right\} \text{after unstressed vowel}$$

$$\begin{bmatrix}-\text{cont}\\-\text{voiced}\end{bmatrix} \rightarrow [+\text{cont}] \text{ except after obstruent or unstressed vowel}$$

If we knew that the changes described by Verner's Law did in fact occur historically after the changes described by Grimm's Law, then we would have a reason to prefer (13) over (14), since we would have reason to believe that (13) describes a set of changes that actually occurred and



affected the output of a law (Grimm's) that had already had its effects. But we do not know that. We have no records that bear on these facts.

We might of course appeal to the fact that (13) is simpler than (14). But simplicity by itself does not constitute evidence about what happened in history. Simplicity considerations become pertinent, however, if we remember that the central mechanism of phonological change is the addition of phonological rules. Diachronic laws are nothing but phonological rules that were added to the language at some point in its history. One of the things that distinguishes Germanic from other Indo-European languages is that speakers of Germanic added to their phonology rules that today we call Grimm's Law, Verner's Law, and so on. These laws were at one point phonological rules that were actually acquired by individual human beings in the course of their linguistic maturation. As a rule of synchronic phonology (14) is much less plausible than (13), and this for the sort of reasons already cited in connection with the formulation of the rules of Canadian English, namely, (4) and (6). We noted there that it is unlikely that children (learners) innately equipped with Principle (7) would acquire a rule containing a special exception clause (see (8)) when a functionally equivalent exceptionless rule (namely, (6)) is available. By the same reasoning, the exception clause makes it unlikely that children would be able to acquire rule (14) through exposure to ambient speech, whereas (13) would be relatively easy to acquire by children already equipped innately with Principle (7).[13]

In short, then, there are known facts about diachronic changes that are best explained as resulting from the introduction of new phonological rules in the grammar of certain speaker/hearers. When we try to specify what these phonological rules might have been, we find that the more plausible answer assumes that Principle (7) holds of these rules too, and hence that these rules too are ordered in the phonology and operate through derivations.

## 4. A Note on Recent History

Extrinsically ordered rules obeying Principle (7) much like those illustrated above were employed in a synchronic account of the phonology of a language by the great Sanskrit grammarian Pāṇini over twenty-five hundred years ago. They were assumed standardly—without much discussion—during the nineteenth century (and later) in accounts of different sound changes. (See Verner's statement of his law in note 12.)

Attempts to utilize extrinsically ordered rules in the description of synchronic rather than historical phenomena date back to the 1930s. One of the earliest is Bloomfield's (1939) paper "Menomini Morphophonemics."[14] Bloomfield describes his approach in the following much quoted passage:

> The process of description leads us to set up each morphological element in a theoretical base form and then to state the deviations from this basic form which appear when the element is combined with other elements. If one starts with the basic forms and applies our statements ... in the order in which we give them, one will arrive finally at the forms of words as they were actually spoken. Our basic forms are not ancient forms, say of the Proto-Algonquian parent language, and our statements of internal sandhi are not historical but descriptive, and appear in a purely *descriptive order*. However, our basic forms do bear some resemblance to those which would be set up for a description of Proto-Algonquian, some of our statements of alternation ... resemble those which would appear in a description of Proto-Algonquian, and the rest ..., as to content and order, approximate the historical development from Proto-Algonquian to present-day Menomini. (pp. 105–106)

It is somewhat difficult to empathize today with the belief widely held among linguists in the 1930s that principles operative in languages conceived as synchronic systems functioning autonomously were totally different from the principles operative in the historical evolution of languages. In particular, to the linguists of that day Principle (7) and derivations of the sort illustrated above seemed appropriate only to historical descriptions, not to synchronic accounts. In fact, in his book *Language* (1933) Bloomfield fully shared the views about the irrelevance of rule order in synchronic descriptions. He wrote:

> The actual sequence of constituents, and their structural order ... are a part of the language, but *the descriptive order of grammatical features is a fiction* and results simply from our method of describing the forms; it goes without saying, for instance, that the speaker who says *knives*, does not "first" replace [f] by [v] and "then" add [-z], but merely utters a form (*knives*) which in certain features resembles and in certain features differs from a certain other form (namely, *knife*). (p. 213; our italics)

As we have seen, some six years later, by the time of composing "Menomini Morphophonemics," Bloomfield had changed positions. The fact that he had done so, however, was totally ignored by the American linguistic community in the 1940s and 1950s. The article was omitted—"inadvertently," according to Hockett (1970, 494)—from



Hockett's "Implications of Bloomfield's Algonquian Studies," which was published in the issue of *Language* (24.1) dedicated to Bloomfield on the occasion of his sixtieth birthday in 1948. It is not referred to in Hockett's (1954) influential "Two Models of Grammatical Description" (which echoes the passage quoted above from Bloomfield (1933) almost verbatim);[15] nor was it reprinted in Joos's (1957) *Readings in Linguistics*. In fact, the article was so unknown in America that Chomsky tells us that he had not read "Menomini Morphophonemics" until his attention was drawn to it by Halle in the late 1950s. And thereby hangs a tale (with a moral perhaps) with which we conclude this article.

In the years immediately following World War II graduate students in linguistics were taught that words and morphemes had a number of distinct representations, each of which corresponded to a specific descriptive level. Three such levels were recognized (the morphophonemic, the phonemic, and the phonetic), and at each level the representations were composed of entities that were specific to that level (morphophonemes, phonemes, and phones). The primary focus was on discovering *the correct* phonemic and morphophonemic representations; the correct phonetic representation did not have to be discovered, since it was directly given in tokens. Implicit in this doctrine was the further assumption that at each level there was only a single representation, and it is this assumption of the standard theory of the 1950s that distinguished it fundamentally from Bloomfield's (1939) (and Pāṇini's) model. As noted, however, the fact that an alternative approach to phonological description had been tested successfully by Bloomfield was hardly known at the time, and the consensus in the 1940s was that derivations and ordered rules did not belong in synchronic accounts of the phonology of a language.

The prevailing wisdom was challenged in Chomsky's (1951) Master's thesis, *The Morphophonemics of Modern Hebrew*. In this early study Chomsky explicitly dissents from the proposition that utterances have single representations at each of the descriptive levels. Rather, he assumes that at least some levels consist of a set of representations generated by extrinsically ordered rules. Chomsky describes the morphophonemic level as follows:

> Beginning with a sequence of morphemes ... each statement of the ... grammar specifies certain changes which must be undergone by any sequence of a certain shape. It will appear that an order is imposed on the statements relative to certain criteria of simplicity. Thus the statements are ordered so as to present a maximally simple grammar. (p. 4)

In fact, the ordering of the statements is a central objective of Chomsky's investigation; he says:

> ... this investigation is limited in that only one "dimension" of simplicity is considered, viz. ordering. (p. 5)

In the version of Chomsky's thesis published in 1979 there is no reference to the fact that like the rules in Bloomfield's "Menomini Morphophonemics" some of the synchronic rules of Modern Hebrew are identical with well-known sound changes; for example, MR 34 is identical with the rule of Postvocalic Spirantization (see Brockelmann (1916, 84)), whereas MR 28 is identical with Vowel Reduction (see Brockelmann (1916, 61)). As a student of Semitic languages, Chomsky was of course fully aware of these parallels between synchronic and diachronic rules. Unlike most linguists of that period he was not concerned about confusing synchronic and diachronic descriptions and viewed the parallels between the two types of rules as evidence in support of his proposed analysis (Chomsky (personal communication)). He assumed that sound changes are due to the addition of phonological rules, and as a consequence it did not seem to him at all strange that some sound changes should survive as synchronic rules for long periods of time.

Chomsky's treatment of the segholates offers another example in which the historical evolution of forms receives a synchronic interpretation so that a form such as [melek] is derived from underlying [malk]. Chomsky reports that this replaces an earlier account where [melek] rather than [malk] was the basic underlying form from which the different surface variants were derived. Chomsky made the change at the suggestion of the late Yehoshua Bar-Hillel, who was one of the few people to study the rather forbidding text of *The Morphophonemics of Modern Hebrew* in considerable detail. Bar-Hillel pointed out to Chomsky that the assumption that [malk] is the underlying form led to a simpler account than the alternative that had figured in the earlier version that Bar-Hillel was reading. He also noted that this account paralleled the known historical evolution of the language.

In 1951 Chomsky thus was independently led to the same conclusions that Bloomfield had reached twelve years earlier. It is a matter of some puzzlement that none of Chomsky's teachers at the University of Pennsylvania drew his attention to Bloomfield's paper and suggested that he take account of it at least by including it in his bibliography. It is idle at this distance in time to speculate about the reasons for this oversight. In any event, as noted above, Chomsky learned of the existence of Bloomfield's paper only in the late 1950s, many years after submitting his Master's thesis.[16]



Chomsky (1988a) notes that his work on the phonology of Modern Hebrew naturally led him to explore whether some of the devices he had used there might also have a use in syntax. Such a project was especially attractive at that time as phonology was then widely viewed not only as the most advanced branch of the field but also as a model for all other linguistic domains to follow. It took two decades of intensive research for Chomsky to conclude that the syntax of a language does in all likelihood not include a system of extrinsically ordered rules (ordered transformations). Since, as we have tried to suggest in section 1, the subject matter of phonology is intrinsically different from that of syntax, the consequences of this conclusion for phonology are far from self-evident: whether and how the principles-and-parameters approach of Chomsky (1981) should be extended is an empirical question. None of the arguments and facts that led Chomsky to this radical change in position with regard to syntax has any detectable bearing on the structure of phonological theory. By contrast, there is much evidence of the sort adduced above in support of the view that in phonology extrinsically ordered rules play a major role. In the absence of evidence to the contrary, it would therefore be a mistake to try to eliminate such rules from phonology. To construct phonology so that it mimics syntax is to miss a major result of the work of the last twenty years, namely, that syntax and phonology are essentially different.

## *Notes*

This article is a modified version of the paper presented on April 14, 1988, in Jerusalem at the symposium "The Chomskyan Turn." The authors gratefully acknowledge the support for work on the paper provided by the Van Leer Foundation, Jerusalem, the Center for Cognitive Science, MIT, and the Center for the Study of Language and Information, Stanford University. For critical discussion and advice we are indebted to N. Chomsky, J. Goldsmith, J. Harris, N. Hornstein, M. Kenstowicz, P. Kiparsky, J. McCarthy, K. P. Mohanan, D. Pesetsky, D. Pulleyblank, and J.-R. Vergnaud.

1. We deliberately eschew in this discussion the use of "declarative rules" and "procedural rules" in characterizing the differences between syntax/semantics on the one hand and phonology on the other hand. That terminology, which carries a number of associations from the domain of computational linguistics, strikes us as unhelpful.

2. For the purpose of this article we limit our use of "derivation" and of "ordering" to nondegenerate cases, that is, derivations of more than one step and orderings of more than one rule.
3. We set aside here issues surrounding the need to assume Logical Form as an autonomous level of representation; see, for example, Williams (1988). Nothing in what follows requires that we take a stand on that issue.
4. J.-R. Vergnaud has drawn our attention to the fact that in accounting for constructions with parenthetical phrases such as *John is not—what I'd call—a great lover* versus **I would not call what John is a great lover*, it may be necessary to assume that the Surface Structure representation is derived from the Deep Structure representation. Vergnaud notes, however, that to the best of his knowledge there do not exist outside of phonology derivations where the application of a pair of rules or principles must be extrinsically ordered, and it is the existence of this type of derivation and of intermediate representations that is at issue here. In short, the issue is not whether representations themselves can be meaningfully ordered but whether the rules or principles applicable to them are ordered prior to any application (and whether the rules ever generate intermediate representations).
5. For present purposes we restrict attention exclusively to the articulatory aspect of language and ignore the auditory interpretive system. The role of memory in the interpretation of utterances is obviously very different from its role in production, but we believe that here again words must be stored in maximally succinct form in order to expedite the search. See also footnote 7.
6. For expository reasons (1) has been simplified by omitting a number of features and other phonetic properties such as sonorant, stress, pitch, and length that would have to be included in a full surface representation of the word.
7. John McCarthy has objected to our attributing the requirement of nonredundant underlying representations to memory limitations. He notes that whatever evidence we have on this matter argues that memory is freely available but that word recognition is hard. Phonology must therefore provide "lots of different ways to get from speech back to the lexical entry," and this retrieval process is most effectively accomplished if the lexical entry is stored in the least redundant form so that there are numerous ways of getting back from phonetic surface to stored entry. We agree with McCarthy's points about the relevance of retrieval requirements. But we believe that memory limitations probably also play a role. However, when talking about optimizing memory storage, we must distinguish between what is required in order to maximize the number of words that can be stored and what is required in order to store a particular word. We believe that there is probably an upper (but very high) limit to the number of representations that can be memorized and in addition an upper (relatively low) limit to the complexity of any representation that can be stored or is likely to be stored on the basis of a few exposures. There may be a tradeoff between the two limits—that is, the simpler the representations, the more of them can be stored. We know too little about this to say anything more. But the fact that matters for our purpose—and that is relatively uncontroversial—is that our ability both to store and to retrieve representations is increased when what must be stored comprises fewer elements.
8. Norbert Hornstein has rightly pointed out to us that this argument is based on an assumption that needs independent justification, namely, on the assumption that it is easier to learn rules (4) and (6) *and their relative ordering* than it is to learn rules (4) and (8) and nothing about their ordering. All other things being equal, learning three things must be harder than learning two things, but all other things are not equal here. Note that (8) is a rule of great complexity when stated fully, that is, when formalized in the full phonological notation. Moreover, the elimination of Principle (7) would require not only the replacement of (6) by the more complicated rule (8) but also the replacement of a host of other relatively simple rules by rules of greater complexity. Consider also that our examples deal with the ordering of only pairs of rules but that a real phonology involves ordering of triplets, quadruplets, quintuplets, and so on. The added complexity in such cases renders the replacement rule totally untransparent, if not unstatable. Finally, there is no reason to believe that these added complexities share general properties that can be encoded in a principle that is available to a learner in the way in which rule ordering is available to a learner equipped with Principle (7). Without some such principle it is unlikely that a learner would discover the exception clauses. Thus, the evidence available to us at this time suggests that the answer to Hornstein's question is that a theory based on rule ordering is more plausible than one based on complicated contextual restrictions. We are grateful to Hornstein for drawing our attention to this issue, which we had previously overlooked.



9. We cannot consider this in detail here, but see, for example, the discussion of the English Stress Rule in Chomsky and Halle (1968). This cyclic rule order has played a major role in discussions of the theory of Lexical Phonology. See especially Pesetsky (1979), Kiparsky (1982), Halle and Mohanan (1985), Halle and Vergnaud (1987), and Halle (1987b). The concept of "strict cycle" in phonology that has resulted from these discussions is, in our opinion, one of the most intriguing and profound results of modern phonological investigations.
10. See Halle and Mohanan (1985) and Halle (1987b).
11. Although this idea is all but self-evident today, it took linguists almost three-quarters of a century to accept the fact that "sound laws" are nothing but phonological rules. The reason for this was that the status of phonological rules in speakers' knowledge of their language was not properly understood until relatively recently. Thus, as Halle (1987a) has argued, Schuchardt's opposition to the "neogrammarian" doctrine of the exceptionless functioning of the "sound laws" was founded on his belief that speakers' knowledge of the phonology of their language consists exclusively of the knowledge of words and that phonological rules play no role in it.
12. Verner's own formulation reads:

> IE *k*, *t*, *p* first became *h*, θ, *f* everywhere [by virtue of Grimm's Law—SB/MH]; the voiceless fricatives that arose in this fashion as well as the voiceless fricative *s* inherited from IE were subsequently voiced word-internally in voiced environment, but remained voiceless in position after stressed syllables. (1876, 114)

See also Saussure (1949, 200–202) and Bloomfield (1933, 357–358).

13. Paul Kiparsky has observed that the above account assumes that at the stage where Verner's Law entered the language the first part of Grimm's Law (that is, (9a)) was still part of the phonology of the language. It is conceivable that this assumption is incorrect and that the effects of Grimm's Law had become lexicalized by the time

Verner's Law entered the language. Though it is far from conclusive, there is some evidence militating against the lexicalization of the effects of Grimm's Law. As indicated in (9a), this part of Grimm's Law was contextually restricted so as not to apply in position after obstruents. As a result, the first part of Grimm's Law did not eliminate voiceless stops from the language altogether but only restricted their distribution, and this fact would have to be reflected formally in the phonology, by means of a rule much like (9a). Moreover, (9a) predicts that voiceless obstruents at the beginning of Germanic suffixes should alternate between stop and continuant depending on whether or not these suffixes are attached to stems that end in an obstruent. That prediction is borne out by the behavior of the participial suffix /t/, which regularly alternated in the predicted way. In view of these facts it seems to us somewhat unlikely that the effects of Grimm's Law were lexicalized by the time Verner's Law came into the language.

14. Very similar in approach is Swadesh and Voegelin's (1939) paper on Tübatulabal. It is difficult at this distance in time to establish whether Bloomfield influenced Swadesh and Voegelin, whether the latter influenced Bloomfield, or whether the ideas were developed independently.

15. According to Hockett, a model with extrinsically ordered rules and derivations (which in Hockett's paper is referred to by the initials IP) has been rejected by some workers in favor of a model that expressly violates Principle (7) (the latter approach is labeled IA)

> because of a feeling of dissatisfaction with the 'moving-part' or 'historical' analogy implicit in IP. At the very least, these analogies seem to imply the necessity of making certain decisions in a possibly arbitrary way. Critics of IP would prefer to circumvent such decisions altogether. For example,... if it be said that the English past-tense form *baked* is 'formed' from *bake* by a 'process' of 'suffixation', then no matter what disclaimer of historicity is made it is impossible not to conclude that some kind of priority is being assigned to *bake*, as against either *baked* or the suffix. And if this priority is not historical, what is it? Supporters of IP have not answered that question satisfactorily. (p. 211)

16. Noam Chomsky has remarked that our presentation of the positions of structuralist phonology—both American and Praguian—fails to bring out their empiricist and antimentalistic foundations. For structuralists phonemes are defined as similarity classes of phones, and morphophonemes as similarity classes of phonemes: all phones of a given phoneme therefore had to share a specific set of phonetic properties that distinguished them from the phones of any other phoneme. In Bloomfield's "Menomini Morphophonemics" the relation between morphophonemes and phonemes and/or phones was conceived in a radically different way: morphophonemes were related to phonemes or to phones by means of rules that "translate" (or map) sequences of morphophonemes into sequences of phonemes/phones. In effect, then, in spite of his frequently professed antimentalism, here Bloomfield viewed the two kinds of sequences as equally real (mental) representations of the words, phrases, or sentences of a language. Moreover, on this view there is no longer an a priori (definitional) requirement that the set of phonemes/phones that correspond to a given morphophoneme share some distinguishing set of properties.



The proposition that phonology should deal with mental representations—that is, with facts that go beyond physical and directly observable events (classified by the linguist)—was not one that linguists were ready to accept in the 1940s and early 1950s when naive forms of positivism were almost universally taken for granted. As a consequence, Bloomfield's paper was treated as a curious experiment—not to say, indiscretion—that did not merit extensive discussion.

## *References*


Akmajian, A. and F. Heny
1975 *An Introduction to the Principles of Transformational Syntax*, MIT Press, Cambridge, Massachusetts.

Baker, C. L.
1978 *Introduction to Generative-Transformational Syntax*, Prentice-Hall, Englewood Cliffs, New Jersey.

Bloomfield, L.
1933 *Language*, Holt, New York.

Bloomfield, L.
1939 "Menomini Morphophonemics," *Etudes dédiées à la mémoire de M. le Prince N. S. Trubetzkoy = Travaux du Cercle linguistique de Prague* 8, 105–115.

Brockelmann, C.
1916 *Semitische Sprachwissenschaft*, G. J. Goschen'sche Verlagshandlung, Berlin.

Chomsky, N.
1951 *Morphophonemics of Modern Hebrew*, Master's thesis, Department of Linguistics, University of Pennsylvania. Published in 1979 by Garland, New York.

Chomsky, N.
1981 *Lectures on Government and Binding*, Foris, Dordrecht.

Chomsky, N.
1988a "A Personal View," paper presented at the conference "The Chomskyan Turn," Jerusalem, April 1988, published version "Linguistics and Adjacent Fields: a Personal View," in Asa Kasher, ed. *The Chomskyan Turn* Basil Blackwell, Cambridge, MA and Oxford 1991, pp. 3–25.

Chomsky, N.
1988b "Prospects for the Study of Language and Mind," paper presented at the conference "The Chomskyan Turn," Jerusalem, April 1988, published version "Linguistics and Cognitive Science: Problems and Mysteries," in Asa Kasher, ed. *The Chomskyan Turn* Basil Blackwell, Cambridge MA and Oxford, 1991, pp. 26–53.

Chomsky, N. and M. Halle
1968 *The Sound Pattern of English*, Harper and Row, New York. Paperback edition published in 1991 by MIT Press, Cambridge, MA.

Halle, M.
1962 "Phonology in a Generative Grammar," *Word* 18, 54–72.

Halle, M.
1987a "Remarks on the Scientific Revolution in Linguistics 1926–1929," in K. Pomorska et al., eds., *Language, Poetry and Poetics: The Generation of the 1890s: Jakobson, Trubetzkoy, Majakovskij*, Mouton de Gruyter, Berlin.
Halle, M.
1987b "On the Phonology-Morphology Interface," ms., Department of Linguistics and Philosophy, MIT, Cambridge, Massachusetts.
Halle, M.
1992 "Phonological Features," in W. Bright, ed. *International Encyclopedia of Linguistics*, vol. 3, 207–212. Oxford University Press, Oxford and New York.
Halle, M. and K. P. Mohanan
1985 "Segmental Phonology and Modern English," *Linguistic Inquiry* 16, 57–116.
Halle, M. and J.-R. Vergnaud
1987 *An Essay on Stress*, MIT Press, Cambridge, Massachusetts.
Hockett, C. F.
1954 "Two Models of Grammatical Description," *Word* 10, 210–231.
Hockett, C. F., ed.
1970 *A Leonard Bloomfield Anthology*, Indiana University Press, Bloomington, Indiana.
Jakobson, R.
1938 "Observations sur le classement phonologique des consonnes," in Proceedings of the 3rd International Congress of Phonetic Sciences, Ghent. [Also in *Selected Writings* I, Mouton, The Hague and Berlin, 1971, pp. 272–279.]
Joos, M., ed.
1957 *Readings in Linguistics*, American Council of Learned Societies, Washington, D.C.
Kaye, J. and J. Lowenstamm
1986 "Compensatory Lengthening in Tiberian Hebrew," in L. Wetzels and E. Sezer, eds., *Studies in Compensatory Lengthening*, Foris, Dordrecht.
Kiparsky, P.
1973 "'Elsewhere' in Phonology," in S. Anderson and P. Kiparsky, eds., *A Festschrift for Morris Halle*, Holt, Rinehart and Winston, New York, pp. 93–106.
Kiparsky, P.
1982 "Lexical Morphology and Phonology," in Linguistic Society of Korea, ed., *Linguistics in the Morning Calm*, Hanshin, Seoul, pp. 3–91.
Mahajan, A.
1988 "Hindi Stress," ms., Department of Linguistics and Philosophy, MIT, Cambridge, Massachusetts.
Majdi, B. and D. Michaels
1987 "Syllable Structure, Gemination and Length in Iraqi Arabic," paper presented at the 62nd Annual Meeting of Linguistic Society of America.
Myers, S.
1985 "The Long and the Short of It: A Metrical Theory of English Vowel Quantity," in *Papers from the 21st Regional Meeting of the Chicago Linguistic Society*, University of Chicago, Chicago, Illinois, 275–288.



Perlmutter D. and S. Soames
1979 *Syntactic Argumentation and the Structure of English*, University of California Press, Berkeley, California.
Pesetsky, D.
1979 "Russian Morphology and Lexical Theory," ms., Department of Linguistics and Philosophy, MIT, Cambridge, Massachusetts.
Sagey, E. C.
1986 *The Representation of Features and Relations in Non-Linear Phonology*, Doctoral dissertation, Department of Linguistics and Philosophy, MIT, Cambridge, Massachusetts.
Saussure, F. de
1949 *Cours de linguistique générale*, Payot, Paris.
Swadesh, M. and C. F. Voegelin
1939 "A Problem in Phonological Alternation," *Language* 15, 1–10.
Verner, K.
1876 "Eine Ausnahme der ersten Lautverschiebung," *Zeitschrift für vergleichende Sprachwissenschaft* 23, 97–130.
Williams, E.
1988 "Is LF Distinct from S-Structure?" *Linguistic Inquiry* 19, 135–146.