

Why Phonology Is Different

Author(s): Sylvain Bromberger and Morris Halle

Source: *Linguistic Inquiry*, Vol. 20, No. 1 (Winter, 1989), pp. 51-70

Published by: The MIT Press

Stable URL: <http://www.jstor.org/stable/4178613>

Accessed: 14-04-2018 01:49 UTC

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://about.jstor.org/terms>



JSTOR

The MIT Press is collaborating with JSTOR to digitize, preserve and extend access to *Linguistic Inquiry*

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

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.

order, strict cyclicity, and so on; see, for example, Kaye and Lowenstamm (1986) and Majdi and Michaels (1987).

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 external 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,

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.

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.

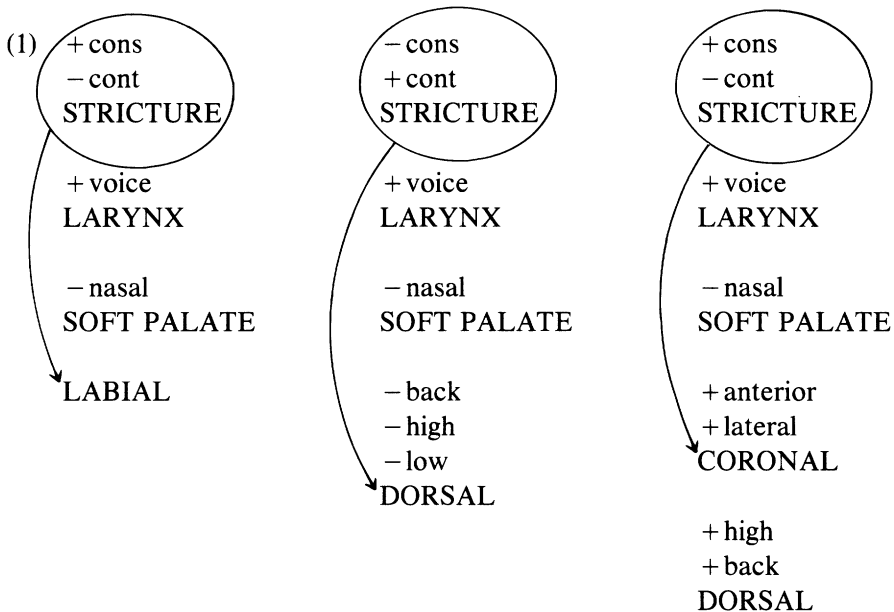
2. The Synchronic Evidence

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

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).

discrete sound segments concatenated in the order in which they must be produced.⁵ 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 [e], 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 [bel] 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 (1988) and differs quite markedly from feature systems utilized in earlier publications by the present authors.)



The representation in (1) encodes the information that enables a speaker to produce the sound sequence [bel]; 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

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.

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.

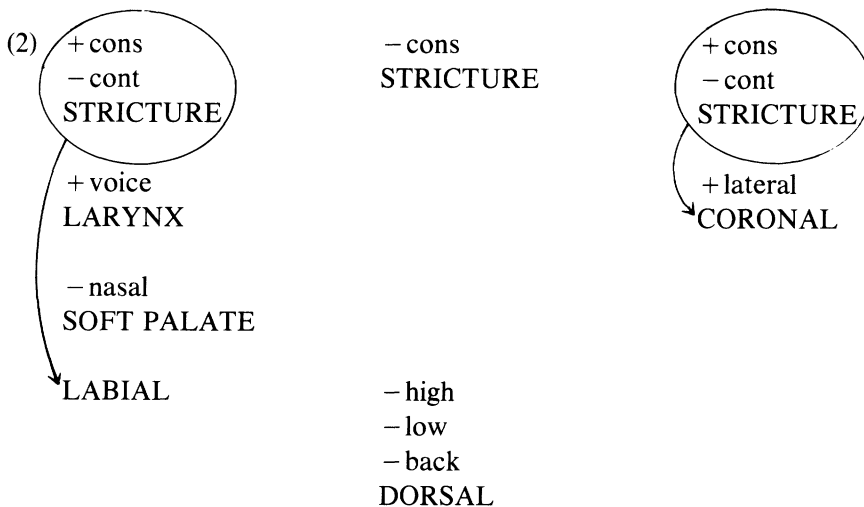
As noted, (1) is a surface representation of the English word *bell* in that in principle it provides the information needed by a speaker to produce this word correctly.⁶ In addition to surface representations such as (1) words also have abstract underlying representations (that is, representations that 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 [naʃar] and [paʃamon]. It is for this reason that under normal circumstances [boy] and [bel] may figure in their utterances, but not [naʃar] and [paʃamon].

However, this sort of knowledge is not innate. It must be acquired and retained—

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.

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 underlying 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

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 trade-off 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.

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).

$$(4) \left[\begin{array}{c} -\text{cont} \\ \text{CORONAL} \end{array} \right] \rightarrow [+ \text{voiced}] \text{ in env. } \left[\begin{array}{c} -\text{cons} \\ \text{stressed} \end{array} \right] \text{ — } \left[\begin{array}{c} -\text{cons} \\ \text{unstressed} \end{array} \right]$$

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

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

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 elim-

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).

$$(9) \text{ a. } \left[\begin{array}{l} - \text{cont} \\ - \text{voiced} \end{array} \right] \rightarrow [+ \text{cont}] \text{ except after obstruent}$$

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).

b.	Germanic	Greek	Latin	Sanskrit	Baltic	
	fot	pod	ped	pad	ped	‘foot’
	θre	tri	trēs	tray	tri	‘three’
	xund	kun	kan	švan	sun	‘dog’
	naxt	nukt	nokt	naktis	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. $\left[\begin{array}{l} -\text{cont} \\ -\text{voiced} \end{array} \right] \rightarrow [+ \text{asp}]$ 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. $\left[\begin{array}{l} -\text{cont} \\ -\text{asp} \end{array} \right] \rightarrow [-\text{voiced}]$

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.

b.	Germanic	Greek	Latin	Sanskrit	Baltic	
	two (Eng)	duo	duo	duva	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. $p \rightarrow f$ $t \rightarrow \theta$ $k \rightarrow x$
 b. $b \rightarrow p$ $d \rightarrow t$ $g \rightarrow k$

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) $\left\{ \begin{array}{l} [+cont] \rightarrow [+voiced] \\ \left[\begin{array}{l} -cont \\ -voiced \end{array} \right] \rightarrow \left[\begin{array}{l} +cont \\ +voiced \end{array} \right] \end{array} \right\}$ after unstressed vowel

$\left[\begin{array}{l} -cont \\ -voiced \end{array} \right] \rightarrow [+cont]$ 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

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 in voiced environment syllable-initially, but remained voiceless in position after stressed syllables. (1876, 114)

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

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 footnote 13.) 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:

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.

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*.

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)

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, *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 *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

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.

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.

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.
- Chomsky, N. (1988b) "Prospects for the Study of Language and Mind," paper presented at the conference "The Chomskyan Turn," Jerusalem, April 1988.
- Chomsky, N. and M. Halle (1968) *The Sound Pattern of English*, Harper and Row, New York.
- 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. (1988) "Features," ms., Department of Linguistics and Philosophy, MIT, Cambridge, Massachusetts. [To appear in *Oxford Encyclopedia of Linguistics*, 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.]
- 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.
- Kiparsky, P. (1982) "Lexical Morphology and Phonology," in Linguistic Society of Korea, ed., *Linguistics in the Morning Calm*, Hanshin, Seoul.
- 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.
- 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.

*Department of Linguistics
and Philosophy*

MIT

Cambridge, Massachusetts 02139