The making of cognitive science

Essays in honor of George A. Miller

Edited by

WILLIAM HIRST

Department of Psychology, Graduate Faculty of the New School for Social Research

*The right of the
University of Cambridge
to print and sell
all manner of books
was granted by
Henry VIII in 1534.
The University has printed
and published continuously
since 1584.*

CAMBRIDGE UNIVERSITY PRESS

Cambridge
New York   New Rochelle   Melbourne   Sydney

---

12 The immanent form of phonemes

Morris Halle

The role of linguistic knowledge in phonetics

Perhaps the most intriguing aspect of language is its dual nature. On the one hand, an utterance is an acoustical signal produced by readily observable gymnastics of the human vocal tract: the lips, tongue, soft palate, larynx, and so on, that is, the anatomical structures that make up the upper end of our digestive and respiratory tracts. On the other hand, an utterance always involves knowledge of a special kind, for it is only by virtue of this knowledge that the physical signal that strikes our ears has meaning. For the person who does not know English, the sounds made by someone speaking English are just that – meaningless noises. Moreover, such a person would not only fail to understand an English utterance; he would also experience great difficulty if he were asked to reproduce a segment of spoken English. For example, a person who does not know English is unlikely to be able to reproduce the eight-word sequence "the noun phrase that I am now uttering," and will, of course, have no idea of the meaning of this phrase. By contrast, a speaker of English would understand the meaning of the phrase and would find little difficulty in repeating the eight words that make up "the noun phrase that I am now uttering." Knowledge of language thus affects aspects of linguistic behavior that at first appear to be quite mechanical, such as the ability to reproduce a short phrase.

The study of the production and perception of spoken utterances has been the province of the science of phonetics. Among the questions phonetics has been trying to answer are, not surprisingly, questions concerning the neurophysiological organization of the speaking process. Phoneticians want to understand precisely what sort of gymnastics a fluent speaker of English engages in in producing an English utterance and how this gymnastics differs from as well as resembles the gymnastics engaged in by speakers of Japanese, Javanese, Arabic, or Kwakiutl. Phoneticians also want to know how this vocal tract gymnastics is

This work was supported in part by the Center for Cognitive Science, MIT.

anatomically structured and controlled. Although the speaking process has been subject to serious physiological inquiry for almost a century and a half, our understanding of this aspect of language is still rather rudimentary: We know much less about how we speak than about how cockroaches walk, fish swim, or monkeys reach for objects in their visual field. Some will no doubt attribute this disparity in our knowledge to the fact that we are limited in the type of experimentation to which we can subject humans. It seems to me, however, that a much more serious impediment to progress has been the failure of phoneticians to take adequate advantage of a large body of information that is accessible to study, namely, the linguistic knowledge that – as previously noted – is intimately involved in the production of every utterance.

In what follows, I attempt to illustrate how some aspects of this knowledge have been utilized to draw inferences about the way in which the gymnastics of the speaking process is controlled. I present evidence and arguments for a specific organization of the speaking process, and I spell out specific implications that this organization appears to have for the motor physiology of speaking.

Phonological representations are three-dimensional

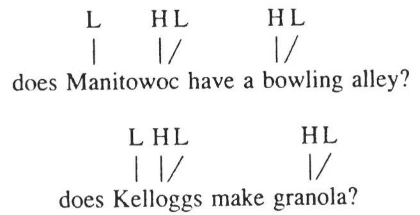
I assume that naive speakers are correct in their belief that every utterance is a sequence of words and that every word is a sequence of speech sounds. Moreover, I assume that speech sounds or phonemes are complexes of binary distinctive features of the sort discussed in Jakobson, Fant, and Halle (1952) and illustrated in (1).

(1)

	p	*b*	*m*	*k*	*g*	ŋ
labial	+	+	+	–	–	–
dorsal	–	–	–	+	+	+
nasal	–	–	+	–	–	+
voiced	–	+	+	–	+	+

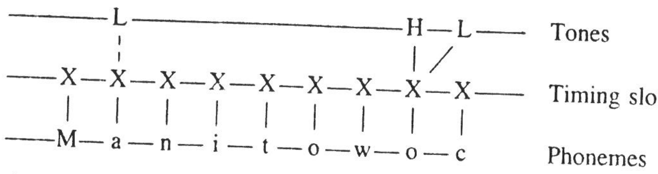
It was suggested by Jakobson that speech sounds are complexes of such distinctive features and nothing else. To Jakobson the formula meant that, for the speaker, speech sounds are not unanalyzable entities, as might be suggested by the letters of the alphabet with which speech sounds are represented in writing; rather, each speech sound is a complex of properties such as those represented in (1). As evidence for the validity of this proposition, I cited in Halle (1978) the fact – brought to my attention by Lise Menn – that English speakers have no difficulty forming plurals of foreign nouns ending in phonemes that do not exist in English. That English speakers form the plural of the German names Bach with /s/ in "boots," rather than with /z/ as in "cows" or with /Iz/ as in "bushes," can be explained only if it is assumed that the rule for forming English plurals is formulated in terms of features rather than phonemes and that English speakers are able to analyze phonemes into their features. If the English plural rule had been stated in terms of phonemes, it could not have included a phoneme that is not part of the language, and if speakers could not analyze phonemes into their component features, there would be no explanation for their ability to form the plural of words that are not part of their language and that contain sounds that are not English. In sum, there is reason to assume that words are represented in the speaker's memory in the form of feature matrices of the kind illustrated in (1).

Research of the last decade has shown that this is only a partial picture of the actual situation. For example, all languages utilize variations in the fundamental pitch of the voice to give melodic shape to their utterances. Thus, in English, utterances are pronounced with quite different melodies (pitch curves) when they are used as a response to a neutral question than when they are intended to convey surprise, dismay, or other emotions. It has been established that pitch curves represent sequences of discrete tones. Like tone sequences in a song, the tone sequences encountered in spoken language frequently can be spread over an arbitrary number of syllables. Thus, the melody appropriate for asking a question in English remains essentially unchanged, without regard to the number of syllables over which it must be spread. This means, in effect, that as in the musical score of a song, we are dealing with two parallel sequences of discrete entities: tones and speech sounds or "phonemes," as illustrated in (2), which has been adapted from Pierrehumbert (1980).



(2)

We note that in (2) not every phoneme is linked to a tone, nor is a single tone linked to a phoneme. It is well known that in the usual case tones are linked only to vowels, and that in English only certain vowels in a phrase are supplied with a tone, as shown in (2). As a result of work by Goldsmith (1979), Williams (1976), Liberman (1975), McCarthy (1979), Pulleyblank (1986), Levin (1985), and others, we have learned a great deal about the formal apparatus that is required to deal with this type of information. An important result of this work has been a change in the nature of the relation between the phonemes and the tones. Rather than link the tones to the phonemes directly, as was done in (2), it was found necessary to establish a somewhat more indirect relationship between the two sets of segment sequences that are now connected by being linked to a sequence of timing slots, as illustrated in (3).

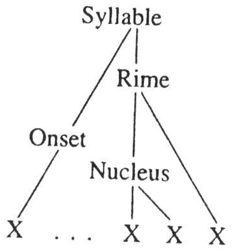


(3)

We have in (3) three parallel lines of units: the tones, the timing slots, and the phonemes. It is an elementary fact of solid geometry that two parallel lines define a plane. It is, therefore, possible to think of (3) as consisting of two half-planes that intersect in the line constituted by the timing slots. The phonemes represented by complexes of distinctive features, as in (1), are contained in the bottom half-plane of (3), and the tones represented by a different set of distinctive features are represented in the top half-plane. The obvious question suggested by a representation such as (3) is whether there are additional half-planes above and beyond the two given in (3) and what function such half-planes might perform.

Perhaps the most striking result obtained by utilizing the three-dimensional format just sketched has been John McCarthy's treatment of the discontinuous morphemes of the Semitic languages, an old conundrum of linguistic theory. We now briefly discuss the problem and present McCarthy's solution.

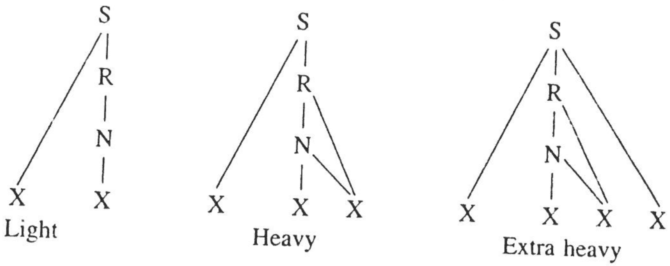
Every language has constraints on what strings of phonemes constitute well-formed sequences in that language. Thus, for example, speakers of English will usually agree that "blick," "snill," and "trun" might be words of English, whereas "lbick," "nsill," and "rtun" might not. It has been discovered that the domain over which sequential well-formedness is defined in all languages is the *syllable*. Languages, of course, differ as to what types of syllables they allow, but in all languages a sequence of well-formed syllables constitutes a well-formed word or utterance. Work by Steriade (1982), Levin (1985), and others has shown that syllables have internal constituent structure of the sort illustrated in (4).



(4)

The syllable is thus a complex of nested binary constituents. It is composed of the rime, which may or must be preceded by one or more timing slots linked to consonants. The rime itself is composed of the nucleus, which must or may be followed by one or more timing slots linked to consonants. Finally, the nucleus may or may not be branching; it must, however, dominate a timing slot linked to a vowel or phoneme of high sonority. Different phonemes or phoneme sequences are admitted in different positions in the syllable. For example, in English /ž/, as in "rouge," is admissible in the rime but not in the onset, whereas tones in English are admitted in the nucleus but not elsewhere. These restrictions can readily be expressed if the syllable structure is represented on a separate half-plane that intersects the half-planes of (3) – that is, the half-planes carrying information about the phonemes and the tones – in the line of timing slots.

The structure of the syllable in classical Arabic, as well as of many other Semitic dialects, is of the fairly simple variety given in (5).



(5)

Heavy syllables have either a branching rime or nucleus, whereas in an extra-heavy syllable an extra consonant is added to a heavy syllable.

Arabic shares an interesting property with all Semitic languages, as well as with a number of non-Semitic languages such as the American Indian language Yokuts (Archangeli, 1984). In these languages the syllabification of a word is determined not by the phonemes that happen to compose the word, but rather by its morphological structure. To illustrate, I cite in (6) some forms from McCarthy (1982a) epitomizing the formation of the so-called broken plurals of Arabic.

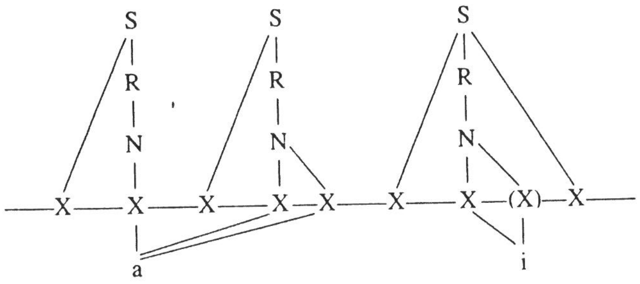
(6)

jundab	janaadib	"locust"
sultaan	salaatiin	"sultan"
duktar	dakaatir	"doctor"
safarjal	safaarij	"quince"
maktab	makaatib	"office"
miktaaḥ	mafaatiiḥ	"key"
nuwar	nawaawir	"white flower"
9andaliib	9annaadil	"nightingale"

The first thing to observe about these examples is that whereas the singular forms are either bi- or trisyllabic, the plural forms are uniformly trisyllabic. Moreover, the structure of the syllables in the plural is fixed. The first syllable is light, the second syllable is heavy, and the third syllable, which always ends with a consonant, has a vowel that is identical in length with that of the second syllable in the singular. This is a typical instance of what is meant by morphology-driven syllabification; that is, syllable structure that is imposed not by the phoneme composition of the word, but by the fact that the word belongs to a particular grammatical category, the plural in the case under discussion.

Nor is syllable structure all that is determined by the fact that the form is a plural noun. Note that the vowel pattern in all plural forms is the same: /i/ in the last syllable, /a/ in the first two syllables. The vowel pattern of the plural is thus totally unrelated to that of the singular: It is determined not by the noun that is pluralized, but rather by the fact of pluralization. This leaves only the consonants to signal the identity of the pluralized noun: to tell us that we are speaking of "doctors" or "sultans" rather than of "nightingales" or "quince"; everything else in the word is determined by the morphology of the word, by the fact that it is a noun belonging to a particular inflectional class. The distribution of the consonants, moreover, is severely restricted: They occur only in those positions in the word where consonants are admitted by the syllable structure, and there are precisely four such positions in every plural form. If the word has more than four consonants in the singular, the extra consonants are omitted, as shown in (6) by the nouns meaning "quince" and "nightingale."

We illustrate in (7) the three-dimensional template of the Semitic broken plural forms.

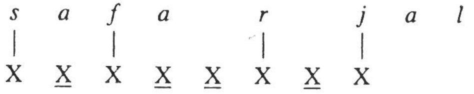


(7)

The question that is posed by the template is clearly that of the way the consonants of the base noun are to be inserted. Since we have already established the need for three half-planes – one to represent syllable structure, another to represent the tones, and a third to represent the phonemes – our first move would be to represent the consonants on the same plane as the vowels. It can, however, be readily seen that this would make it impossible for us to link the consonants to the appropriate timing slots without crossing the lines linking the vowels to their timing slots. The crossing of linking lines, however, must be prohibited if one wishes to preserve the logical coherence of the representation (Sagey 1986). In light of these considerations, it has been proposed that the consonants – that is, the phonemes of the base noun – should be represented on a separate plane intersecting the line of timing slots. Moreover, it was proposed by McCarthy that in all cases the phonemes of the base noun are linked to the slots of the template by the general convention (8).

> Link unlinked phonemes to empty timing slots from left to right and one for one subject to the constraint that the linking results always in well-formed syllables. (8)

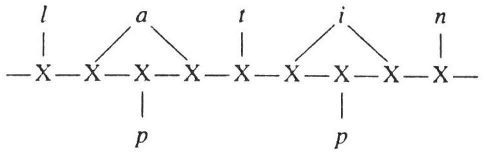
In the case of the broken plurals, this means that only the consonant phonemes of the base noun can be linked, since, as already noted, all timing slots that may be occupied by vowels are prelinked in the template (7). We illustrate this procedure in (9) with the derivation of the plural for the noun meaning "quince." Underlining indicates that a timing slot is prelinked in (9).



(9)

What is especially significant here is the case in which phonemes cannot be linked, either because their linking would create an ill-formed syllable or, as in the case of all vowels and the last consonant of the stem, because there are no more empty timing slots to which they might link, they do not appear in the phonetic realization of the form. Put differently, the vowels of the base noun cannot be linked because all timing slots that admit vowels are prelinked in the template (7), whereas the last consonant of the base noun cannot be linked because all timing slots where consonants are admissible have already been preempted, and these phonemes are omitted in the output because only phonemes that are linked to timing slots can be pronounced.

There is good corroborative evidence for the "psychological reality" of this three-dimensional phonological representation. Perhaps the most convincing evidence known to me is the ability of children to learn various "secret" languages that consist of the insertion of extraphonetic material into the original word. For example, consider a secret language in which the word "Latin" is recoded as "lapatipin." Given the formalism developed here, the recoded word appears as in (10).

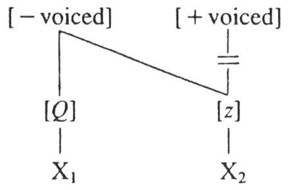


(10)

Formally, this type of language requires the following rule: In every syllable, insert before the rime a copy of the rime vowel followed by the consonant /p/. Much more complicated secret languages have been studied by McCarthy (1982b) and Yip (1982). It is quite difficult to imagine an alternative account of this type of language deformation without making use of essential aspects of the three-dimensional representations that have been described here. The fact that naive speakers can readily master the distortions exemplified here suggests rather strongly that they have access to representations of this type or their equivalent.

The immanent structure of phonemes

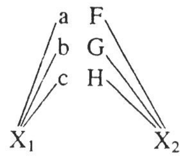
A common phonological process is ''feature assimilation,'' a process whereby a given value of a feature is spread from one phoneme to one or more adjacent phonemes. For example, the rule of implementing the *s*-suffix of the English plural discussed in the section ''The role of linguistic knowledge in phonetics'' must include a subpart specifying that if the stem ends with a [– voiced] sound, and the suffix is actualized as ([– voiced]) [*s*], whereas in all other contexts the morpheme appears as ([+ voiced]) [*z*], for we pronounce [*s*] in ''books,'' ''boots,'' ''loops,'' ''coughs,'' and ''sixths,'' but [*z*] in ''roads,'' ''groves,'' ''cans,'' ''ways,'' and so on. It has therefore been assumed that the basic plural suffix is the voiced /*z*/. After a voiceless obstruent this suffix becomes the voiceless /*s*/. In the three-dimensional notation developed in the preceding section, this fact can be formalized by spreading the feature [– voiced] from the stem segment to that of the suffix while simultaneously delinking its [+ voiced] feature. We illustrate this in (11).



(11)

where *Q* and *z* stand for the feature complexes, excluding voicing, associated with the last stem consonant and the suffix, respectively; X_1 and X_2 for the two timing slots; the dotted arrow for the spreading of the feature [voiced] from X_1 to X_2; and the = sign for the delinking of [+ voiced] from X_2.

Many languages distinguish short and long vowels and/or short and long consonants. (The latter are frequently referred to as ''geminate'' consonants.) In the three-dimensional notation previously developed, such ''long'' phonemes are naturally represented by linking two consecutive timing slots to every feature in a given feature complex, as illustrated in (12).



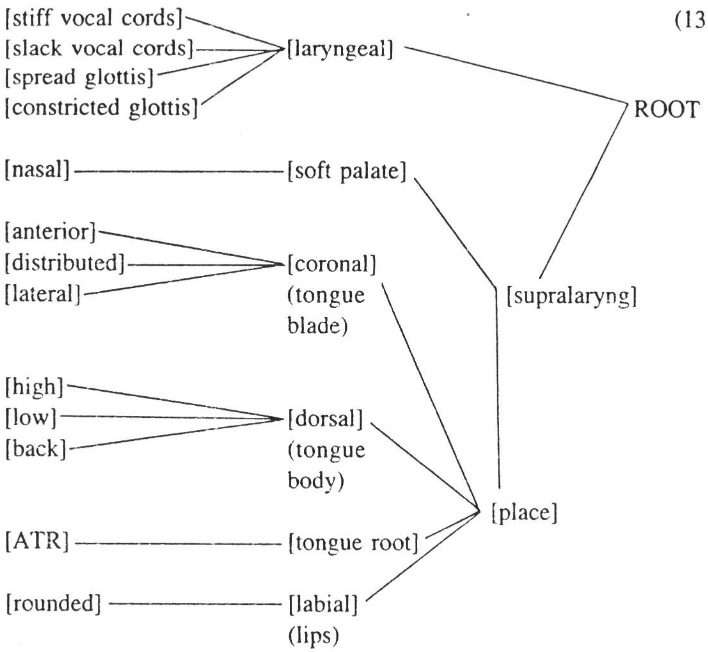
(12)

This type of notation implies, on the one hand, that adjacent timing slots may share any number of features in common, and, on the other hand, that the complexity of shared features increases in direct proportion to their number. When actual assimilation processes are examined, it is readily seen that both implications are incorrect. The markedness of sequences with shared features that are the result of assimilation processes is quite unrelated to their number. There do not seem to be severe restrictions on the sharing of a single feature or of *all* features in a complex. By contrast, the sharing of feature subsets composed of two or more features, yet less than the entire complex, is subject to extremely heavy constraints; a few such multiple assimilations appear to be quite common, but the large majority are never encountered. For example, the sharing of the two features [anterior] and [distributed] is not uncommon, whereas the simultaneous assimilation of the feature pair [nasal] and [round] is unattested to.

These restrictions have been the subject of a number of recent studies (Mascaro, 1983; Mohanan, 1983; Clements, 1985; Sagey, 1986). The chief result of these investigations has been to attribute internal structure to the feature complexes. Specifically, features subsets that are readily assimilated are grouped together into hierarchically superordinate classes; some of the classes, in turn, are grouped together in still higher hyperclasses, which, as proposed by Mohanan, are grouped under a single ROOT node. In (13) I have illustrated this hierarchization of the feature complex, modifying somewhat the proposal made in Sagey (1986).[1]

In (13) the set of terminal features represented in the column on the left is organized into a hierarchy of superordinate classes represented by the labeled nodes in the tree shown to the right of the terminal features. If, following Mohanan (1983), assimilation processes are restricted to single nodes in the tree structure (13) (including those of the terminal features), almost all of the attested and none of the unattested, assimilatory processes are accounted for. In other words, given Mohanan's restriction, we expect single features to be assimilated. We also expect sharing of each of the feature subsets dominated by the different

(13)



class nodes. Thus, our model leads us to expect assimilation of the four terminal features dominated by the class node [laryngeal] or of the nine terminal features dominated by the class node [place], but not to expect the simultaneous assimilation of [nasal] and [round], since these two features are not dominated by a single class node.

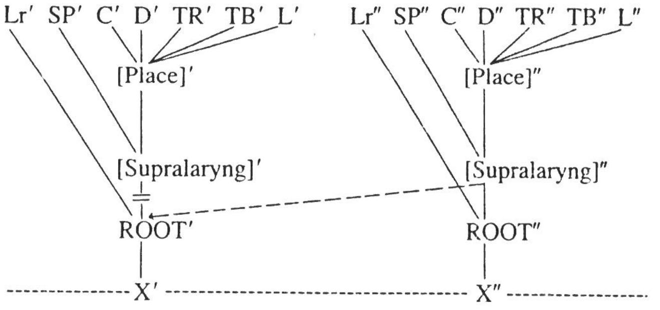
We illustrate this situation with an example discussed by Clements (1985). In the American Indian language Klamath, the phonological alternations shown in (14) have been found.

$$nl \rightarrow ll \quad nL \rightarrow lh \quad nl' \rightarrow l? \qquad (14)$$
$$lL \rightarrow lh \quad ll' \rightarrow l?$$

where $L = [l]$ with [+spread glottis]
$l' = [l]$ with [+constricted glottis]

These alternations are the result of two rules. The first rule turns a nasal into a lateral before a following lateral;[2] it spreads to the former all but the laryngeal features of the latter. The subsequent rule removes from the second consonant in the sequence all but its laryngeal features. In (15) I present the first of these two processes in the tree hierarchy notation previously developed.

(15)



In (15) each timing slot X is linked to a feature complex, as before. The complex, however, is no longer without internal structure, but rather is organized in the manner shown in (13) and (15). Formally, each feature tree is represented in a plane that is orthogonal to the line of timing slots. Because of limitations on my capabilities for graphic representation, I have omitted all terminal features that are dominated by the class nodes at the top of (15); the reader can, I hope, imagine the additional branching lines extending above the topmost nodes.

As noted by Clements, the proposed modifications affect rather markedly the form of phonological representations. Whereas up to this point phonological representations could be envisaged as consisting of a number of half-planes intersecting in a single line on which the timing slots are represented, in representations such as (15) the half-plane on which features were represented is split into a family of half-planes intersecting in lines running parallel to the central line of timing slots. Specifically, in (15) at the ROOT node, the half-plane containing the timing slots and the ROOT nodes is split into two half-planes, one containing the consecutive [Laryngeal] nodes and the other the consecutive [Supralaryngeal] nodes. A further split into two half-planes occurs at the [Supralaryngeal] node and at every other nonterminal node. Because of the already mentioned limitations on my graphic capabilities, these further half-planes are not illustrated graphically.

The Klamath process that we have characterized in (15) is one in which the first phoneme in the sequence assimilates from the second all but the terminal features dominated by the class node Lr″ ([Laryngeal]″). We implement this by connecting the node ROOT′ to the [Supralaryng]″ node while simultaneously disconnecting the ROOT′ node from the [Supralaryng]′ node. The result is two timing slots that agree in their supralaryngeal features but differ in their laryngeal features. The second rule previously mentioned applies to the representation in (15) and cuts the link between ROOT″ and [Supralaryng]″, so that at the end ROOT′ is linked to the terminal features dominated by Lr′ and [Supralaryng]″, whereas ROOT″ is linked to Lr′ but not to any supralaryngeal features.

One consequence of the Klamath rules thus is to generate a timing slot (phoneme) specified for the laryngeal features but unspecified for any of the other features. At first, it might seem that such defective timing slots should be ruled out by a condition governing the well-formedness of phonological representation. When we examine the sounds that are represented by these "defective" timing slots, we discover that a good case can be made for their defectiveness. The sounds represented by the defective timing slots are the glides [*h*] and [ʔ]. These sounds are produced by particular configurations in the larynx: The [*h*] requires that the vocal cords be spread, whereas to produce a glottal stop ([ʔ]) the vocal cords must be constricted. The only other requirement for the production of these sounds is that there should be no constrictions in the vocal tract narrow enough to impede the flow of the expiratory air stream. If we can assume that such an unconstricted state is characteristic of the vocal tract in its "neutral" position and that, in the absence of specific instructions to the contrary, the vocal tract automatically goes into this neutral position, then glides like [*h*] and [ʔ] are correctly characterized by specifying laryngeal features only and omitting specifications for all other features. As remarked by Clements, the possibility of omitting specifications of classes of features – of the sort just illustrated – is one of the arguments in favor of the hierarchical tree structure previously sketched: The structure permits us to express in a simple manner specific properties of speech sounds that could not be expressed except by ad hoc stipulations in earlier frameworks.

The hierarchical tree structure in (13) was proposed in order to facilitate the statement of phonological rules of various kinds. It should, therefore, come as a gratifying surprise that the hierarchical tree in (13) has a direct interpretation in terms of vocal tract anatomy, a fact that was also observed by Clements. In other words, the organization that we were led to impose on the basis of purely grammatical considerations turned out to be one that is directly interpretable in terms of the functional anatomy of the vocal tract.

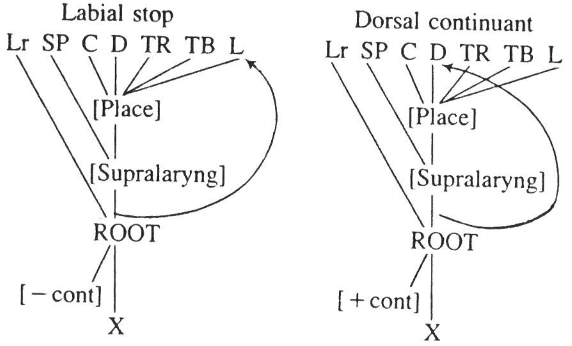
It is obvious that the six class nodes immediately dominating the terminal features in (13) represent each of the six articulators that control the shape of the vocal tract: the larynx, the soft palate, the tongue blade, the tongue body, the tongue root, and the lips. Phoneticians have always been aware of the obvious fact that speech is the result of changes in the shape of the vocal tract and that the only way in which vocal tract shape can be changed is by changing the positioning of its movable parts, that is, of articulators such as the larynx, the soft palate, the lips, and the tongue. In spite of this, articulators play only a secondary role in all major phonetic frameworks such as that of the International Phonetic Association or that of Jakobson et al. (1952). It is one of the obvious advantages of the framework in (13) that it explicitly recognizes this fundamental aspect of the speech production process. It should also be remarked that this recognition was not imposed a priori on the framework, but emerged as a consequence of the attempt to group features in a manner that was optimal for purposes of characterizing certain abstract phonological processes. Thus, we have two independent lines of evidence – one stemming from a study of the rules of phonology and the other from a study of the process of speaking – converging on a single result: the need for explicit recognition of the role of the articulators.

An obvious articulatory difference between features that has not been taken explicit account of in previous frameworks is that between features such as [nasal], [high], [round], or [stiff vocal cords], which can be executed only by a particular articulator, and features like [continuant] or [consonantal], which may be implemented by a number of different articulators. The framework in (13) includes features of the former kind, but none of the latter. Clearly, it is necessary to indicate how features such as [continuant] and [consonantal] are dealt with in the theoretical framework under discussion. I follow here the treatment proposed in Sagey (1986).

Sagey shows that it is not possible to represent the feature [continuant] on the Place node because in many languages place of articulation is assimilated without simultaneous assimilation of the feature [continuant]. For example, in Sanskrit words, the final /*s*/ optionally assimilates in place of articulation to the following obstruent but retains its [+continuant] character regardless of whether the following obstruent is [+cont] or [−cont]. For similar reasons, it is impossible to represent the feature [continuant] on any of the nodes hierarchically subordinate to Place. Sagey concludes, therefore, that [continuant] must be represented in (13) on a node that is superordinate to Place, and we follow Sagey in representing the feature on the ROOT node, the topmost node in the hierarchy.

The feature [continuant] is universally restricted to [+consonantal] phonemes. In many, perhaps even most, languages the [continuant] feature that controls the degree of closure in the vocal tract is implemented by a single articulator. Thus, in English we have labial, coronal, and dorsal (=velar) consonants, that is, consonants where the [continuant] feature is executed by the lips, tongue blade, tongue body, or tongue root, respectively. The problem, therefore, is how to link the [continuant] feature represented on the ROOT node to the articulator that actually executes the feature. Sagey's proposal is to supplement the notation by introducing a special pointer that indicates the articulator in question. We illustrate this in (16).

It was noted previously that each of the seven class nodes dominating the different groups of terminal features in (13) represents a specific articulator. There is (as yet) no direct anatomical interpretation for the three remaining nodes in (13), that is, for the nodes labeled "Place," "Supralaryng," and "ROOT." These may be thought of as higher-level controls in the neurological organization of the vocal tract. Implicit in (13), then, is the empirical hypothesis that in the



(16)

production of speech the tongue and lips together are controlled by a single center, the one labeled "Place." Analogous hypotheses are implied by the nodes Supralaryng and ROOT. It hardly needs saying that at present these hypotheses are nothing but speculations, but they indicate directions for potentially fruitful neuroanatomical explorations.

Pursuing this line of reasoning further, one may regard (13) as a circuit diagram in which the different class nodes represent binary switches. One can then imagine that the intention to articulate a particular sound results in a specific configuration of on and off positions of the switches in the circuit. An immediate consequence of this way of looking at (13) is that whenever a class node is in an off position, no current can flow past it to any of the nodes below it in the circuit, so that the state of the switches lower in the hierarchy becomes irrelevant. The situation is equivalent to that resulting from delinking a class node from the nodes it dominates.

When an articulator node switch is thrown into the off position, we assume that no instructions are transmitted to the articulator. In this case, the articulator will persist in the position to which it was instructed to proceed during the preceding timing slot, or it may continue to move in the direction of that position, subject to the effects of gravity and other external forces that might be acting on the articulator, or, as suggested previously, the articulator may move to its neutral position. In any event, the behavior of an articulator under these circumstances is determined by physiological or physical factors that are independent of the speaker's linguistic intentions.

There are further implications of looking at (13) as a switching circuit. Perhaps the most striking of these implications emerge when we examine from this point of view the production of the two major classes of speech sounds: the consonants and the vowels.

Vowels in all languages are executed with the tongue body; that is, they always involve the dorsal articulator. The lips usually also participate in the production of the vowels, although there are languages (the examples usually cited since Trubetzkoy, 1939, are from the Caucasian languages: Adyge, Kabardian, Abaza), in which the lips play no role in differentiating vowels. The participation of the tongue blade in the production of vowels is quite rare; the "retroflex" vowels of certain Dravidian languages and of English seem to be the major attested examples of vowel sounds that are coronal. Vowel production may or may not involve active participation of the soft palate. A minority of languages have nasal vowels; all languages have nonnasal vowel. Moreover, there are languages (e.g., certain African languages) in which the various laryngeal features are fully exploited in the production of vowels, as discussed in Halle and Stevens (1971).

In the production of consonants, there is a marked difference between the set of features executed by the larynx and the soft palate, on the one hand, and those executed by the tongue blade, tongue body, and lips, on the other hand, that is, by the articulators dominated by the class node Place. As noted previously, in many languages, when a consonant is produced, only one of these Place articulators is active in the production of consonants. Thus, in English, when a labial consonant is produced, only the lips are actively involved and no role is played by the tongue body and blade. And the same – *mutatis mutandis* – goes for consonants produced with the tongue body, like [*k*, *g*, ŋ], and those produced with the tongue blade. This complementarity does not extend to nasalization or to features executed by the larynx: Consonants are or are not nasal, regardless of which role the three place articulators play in their production, and the same is true for voicing and aspiration. In (16), this fact was expressed by means of a pointer connecting the ROOT node with a particular articulator.

I should note here that the situation with English consonants is not universal. There are many languages (e.g., Russian) in which consonants involve the simultaneous participation of two articulators, and in Kinyarwanda (Sagey, 1984) there are consonants produced with simultaneous activation of three Place articulators. Formally, such sounds are represented in the fashion illustrated in (16), except that instead of a single pointer there are multiple pointers from the ROOT node to the different articulators involved.

When a particular articulator is linked to the ROOT node by the pointer, the articulator executes the feature(s) specified at the ROOT. The articulator may or may not also execute some of the terminal features it dominates. For example, in English, none of the terminal features is specified for labial or dorsal (= velar) consonants, but for coronal consonants the terminal features [anterior], [distributed], and [lateral] must be specified.

By distinguishing phonemes in which a place articulator has specified terminal features from those in which it does not, we can account for the well-known fact that in vowel harmony languages the consonants, for which the terminal features are not specified, are transparent to the propagation of such terminal features as

[back], [round], and [advanced tongue root]. It has been observed that in harmony processes (Kiparsky, 1985) the harmonic feature cannot be propagated across a phoneme for which the feature in question has been specified. Since consonants do not normally block backness harmony and roundness harmony in languages like Finnish or Turkish, it must be assumed that these features are not specified for consonants. Support for this analysis comes from the treatment in Turkish of the special palatal stops for which the feature of backness must be specified: Such stops block the propagation of backness harmony (Clements & Sezer, 1982).

The terminal features tell us what configuration a given articulator is to assume during the production of a particular sound. Articulators differ in terms of the variety of configurations they are capable of assuming. The soft palate is capable only of being lowered or raised. By contrast, the tongue body and the larynx have a varied repertory of configurations. This difference, of course, is explained primarily on anatomical grounds: The muscles controlling the tongue allow a great variety of configurations, whereas those of the soft palate do not. In our formalism, these differences are reflected in the number of terminal features dominated by a particular articulator.

It is, of course, obvious that articulators come to occupy one position or another only as a consequence of some muscular action. In Halle (1983) I have made concrete suggestions about how particular feature specifications are translated into activities of particular muscles, and I have tried to show that what little electromyographic evidence there is does not radically contradict the proposed account. The theoretical framework that I have sketched in this section is considerably richer than that underlying the attempt in Halle (1983). It provides, therefore, many new openings for further investigations into the questions raised there.

Notes

1. In line with results of Halle and Stevens (1971), the feature [voiced] utilized in the preceding discussion, is replaced in (13) with the feature pair [stiff vocal cords] and [slack vocal cords].
2. A lateral is an *l*-type sound.

References

Archangeli, D. (1984). Underspecification in Yawelmani phonology and morphology. Ph.D. dissertation, MIT.

Clements, G. N. (1985). The geometry of phonological features. In C. Ewan & T. Anderson (Eds.), *Phonology Yearbook*, Vol. 2 (pp. 223–52). Cambridge University Press.

Clements, G. N., & Sezer, E. (1982). Vowel and consonant disharmony in Turkish. In H. Van der Hulst & N. Smith (Eds.), *The Structure of Phonological Representations* (Part II, pp. 213–55). Dordrecht: Foris Publications.

Goldsmith, J. (1979). *Autosegmental Phonology*. New York: Garland Press.

Halle, M. (1978). Knowledge unlearned and untaught: What speakers know about the sounds of their language. In M. Halle, J. Bresnan, & G. A. Miller (Eds.), *Linguistic Theory and Psychological Reality* (pp. 294–303). Cambridge, Mass.: MIT Press.

Halle, M. (1983). On distinctive features and their articulatory implementation. *Natural Language and Linguistic Theory, 1*, 91–105.

Halle, M. (1985). Speculations about the representation of words in memory. In V. Fromkin (Ed.), *Phonetic Linguistics* (pp. 101–14). Orlando, Fla.: Academic Press.

Halle, M., & Stevens, K. N. (1971). A note on laryngeal features. *MIT RLE Quarterly Progress Report*, No. 101, 198–213.

Jakobson, R. (1938). Observations sur le classement phonologique des consonnes. Reprinted in *Selected Writings*, Vol. I (pp. 273–9). The Hague: Mouton.

Jakobson, R., Fant, G., & Halle, M. (1952). *Preliminaries to Speech Analysis*. Technical Report #13. Cambridge, Mass.: MIT Acoustics Laboratory.

Kahn, D. (1976). Syllable-based generalizations in English phonology. Ph.D. dissertation, MIT.

Kiparsky, P. (1985). Some consequences of lexical phonology. In C. Ewan & J. Anderson (Eds.), *Phonology Yearbook*. Vol. 2 (pp. 85–138). Cambridge: Cambridge University Press.

Levin, J. (1985). A metrical theory of syllabicity. Ph.D. dissertation, MIT.

Liberman, M. (1975). The intonational system of English. Ph.D. dissertation, MIT.

Mascaro, J. (1983). Phonological levels and assimilatory processes. Unpublished paper, Autonomous University, Barcelona.

McCarthy, J. (1979). Formal problems in Semitic phonology and morphology. Ph.D. dissertation, MIT.

McCarthy, J. (1982a). A prosodic account of Arabic broken plurals. In I. R. Dihoff (Ed.), *Current Approaches to African Linguistics*. Dordrecht: Foris Publications.

McCarthy, J. (1982b). Prosodic templates, morphemic templates, and morphemic tiers. In H. Van der Hulst & N. Smith (Eds.), *The Structure of Phonological Representations* (Part I, pp. 191–223). Dordrecht: Foris Publications.

Mohanan, K. P. (1983). The structure of the melody. Unpublished paper, University of Singapore.

Pierrehumbert, J. (1980). The phonology and phonetics of English intonation. Ph.D. dissertation, MIT.

Sagey, E. Walli. (1984). On the representation of complex segments and their formation in Kinyarwanda. Unpublished paper, MIT.

Sagey, E. (1986). The representation of features and relations in nonlinear phonology. Ph.D. dissertation, MIT.

SPE = Chomsky, N., & Halle, M. (1968). *The Sound Pattern of English*. New York: Harper & Row.

Steriade, D. (1982). Greek prosodies and the nature of syllabification. Ph.D. dissertation, MIT.

Steriade, D. (1986). A note on coronal. Unpublished paper, MIT.

Trubetzkoy, N. (1939). *Grundzuege der Phonologie*. Second edition (1958). Goettingen: Vandenhoeck and Ruprecht.

Van der Hulst, H., & Smith, N. (1982). *The Structure of Phonological Representations*. Dordrecht: Foris Publications.

Yip, M. (1982). Reduplication and CV skeleta in Chinese secret languages. *Linguistic Inquiry, 13*, 637–61.

This postscript was part of the original article, but the editor excluded it from the printed volume

Postscriptum

September 1, 1986

Dear George,

In the preceding I have tried to convey how (my) conceptions of the nature of phonemes have changed since you and I first discussed these matters a third of a century ago (in the MIT swimming pool, I seem to recall). At that time -- and during the twenty year following -- I believed that utterances are strings of phonemes, just as sentences on a printed page might be viewed as strings of letters. What distinguished letters from phonemes was that phonemes were bundles of binary features chosen from a narrowly limited universal set, whereas letters are further unanalyzable entities.

During the past twelve years this picture has undergone radical transformations, primarily as a result of a number of MIT doctoral dissertations on which I had the good fortune to be the official advisor. What all these dissertations had in common was that they not only pointed out serious flaws in the existing theoretical framework but also proposed attractive modifications of the framework designed to cure the documented inadequacies.

The first of these was the dissertation by Mark

Liberman completed in 1974. Mark proposed a framework for dealing with various rhythmic properties of language. This framework ultimately became the basis for what to-day is known as <u>metrical phonology</u>. Liberman's dissertation was followed in 1976 by John Goldsmith's dissertation, which laid the foundations for <u>autosegmental phonology</u> of which the fundamental idea is that a phonological representation cannot consist of a single string of entities, but must allow for at least two parallel strings, one representing the tones and the other the phonemes. Then came John McCarthy's (1979) dissertation, where it was shown on the basis largely of data from Semitic languages that it is necessary to separate formally the time slot or interval occupied by a given phoneme from the phonetic features of the phoneme. McCarthy's suggestion amounted in effect to transforming phonological representations from two-dimensional into three-dimensional objects. Specifically, the representation now consists of a number of half-planes intersecting in a single line. The time intervals or slots are represented on the line of intersection and the phonetic content of the phonemes is represented on the planes intersecting in the line of timing slots. This is the chief topic of section B.

Finally, during the academic year just past phonological theorizing has been much influenced by

Clements' (1985) paper on the geometry of features and by Sagey's (1986) dissertation, which significantly modifies and extends Clements' proposals. The effect of these proposals on the phonological representation is to convert at least one of the intersecting half-planes into a family of half-planes as described in sec. C above. Thus, the phonological representation of an utterance is now a three-dimensional object where certain half-planes intersect in the line of timing slots, but others intersect in a line parallel to the time line.

To me the most striking thing about this evolution is that while it has resulted in a much more abstract representation, certain aspects of the representation have become much more concrete, directly relatable to specific aspects of the process involved in giving acoustic shape to the utterance represented. As noted in section C in order to express perspicuously various assimilatory processes found in different languages the world over it is necessary to group different terminal features into a number of classes. The fact that these groupings recur in language after language and that a subset of these corresponds to the grouping we obtain if we list terminal features by the articulator that executes them impresses me as a result of fundamental importance: it reflects the connection that exists between the abstract structure of phonology and the concrete physiological apparatus (our lips, tongue, larynx etc.) by which this abstract structure is externalized and physically communicated. While it will take a number of years to work out this framework in proper detail, it seems to me pretty clear that much of interest is to be expected. It is obvious already that the framework provides the requisite means for dealing with a number of complex phonological problems; e.g., it makes possible an essentially simple and transparent treatment of all types of vowel harmony phenomena. As just noted the framework has a simple interpretation in terms of the different articulators involved in producing speech, and though no psycho-acoustic experiments have as yet been suggested to test the model, there is good reason to suppose that once the model is properly understood it will mark a new era in research on the phonic aspect of language.

I am looking forward to your reaction to the paper which though short on anecdote and gossip represents a good portion of my intellectual autobiography during the decade that almost imperceptibly has passed since we last cooperated in the series of seminars organized at MIT to mark the centennial of the telephone.

With all best wishes,

Morris