7. The Internal Organization of Speech Sounds

G. N. CLEMENTS AND ELIZABETH V. HUME

Subject	Theoretical Linguistics » Pholonogy
DOI:	10.1111/b.9780631201267.1996.00009.x

0 Introduction

In recent years it has become widely accepted that the basic units of phonological representation are not segments but *features*, the members of a small set of elementary categories which combine in various ways to form the speech sounds of human languages. While features are normally construed as psychological entities, they are defined in terms of specific patterns of acoustic and articulatory realization which provide the crucial link between the cognitive representation of speech and its physical manifestation.

The wide acceptance of feature theory results from the fact that it offers straightforward explanations for many potentially unrelated observations. For example, since features are universal, feature theory explains the fact that all languages draw on a similar, small set of speech properties in constructing their phonological systems. Since features are typically binary or one-valued, it also explains the fact that speech sounds are perceived and stored in memory in a predominantly categorial fashion. Moreover, since phonological rules apply to feature representations, it accounts for the observation that phonological rules typically involve “natural classes” of sounds, that is, classes that can be uniquely defined in terms of a single conjunction of features. It also offers explanations for many generalizations in the domains of language acquisition, language disorders, and historical change, among others. Feature theory has emerged as one of the major results of linguistic science in this century, and has provided strong confirmation for the view that languages do not vary without limit, but reflect a single general pattern which is rooted in the physical and cognitive capacities of the human species.[1]

But while much research has been devoted to the questions, What are the features, and how are they defined?, it is only recently that linguists have begun to address a third and equally important question, How are features organized in phonological representations? Earlier theoreticians tended to think of phonemes as unstructured sets of features, or “feature bundles” in Bloomfield's well-known characterization. In accordance with this view, later work in the Jakobsonian and generative traditions treated segments as feature columns with no internal structure. In this approach, phonological sequences were typically characterized as two-dimensional feature matrices, as we illustrate below for the word *sun*:

(1)

	s	ʌ	n
syllabic	−	+	−
voice	−	+	+
nasal	−	−	+
coronal	+	−	+
etc.			

In this view, a phoneme (or phonemic unit) is simply a column of features. Since phonemes follow each other in strict succession, such models can be regarded as *linear*.

The matrix formalism has strong arguments in its favor: it is conceptually simple, it is mathematically tractable, and it imposes powerful constraints on the way features can be organized in representations. In spite of its advantages, however, it has become apparent that this model (as well as other models in which phonemes are viewed as strictly sequential feature bundles) has two important inadequacies.

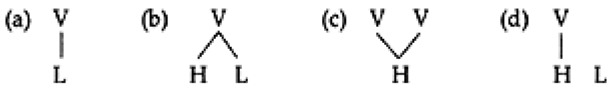
First, in such models all features defining a phoneme stand in a bijective (one-to-one) relation; thus, each feature value characterizes just one phoneme, and each phoneme is characterized by just one value from each category. It follows as a strict prediction that features cannot extend over domains greater or lesser than a single phoneme. However, there is considerable evidence that this prediction is incorrect. Simple and dramatic examples demonstrating "nonlinear" – i.e., nonbijective – relations among features can be drawn from tone languages. For example, in some tone languages, two or more tones may "crowd" onto a single syllable, forming contour tones (i.e., rising and falling tones). In many tone languages, single tones "stretch" or extend over several syllables, and in some, tones "float" in the sense that they are not associated with any particular tone-bearing unit in the representation. Tones are also found to constitute independent "tone melodies" in abstraction from the consonant and vowel sequences on which they are realized. (For discussion of these and other properties, see, e.g., Pike 1948, Welmers 1962, Goldsmith 1976, and Pulleyblank 1986.)

It was earlier thought that nonlinear relations among features of this sort are restricted to a small set of prosodic or suprasegmental speech properties, including tone, stress, and intonation. However, it has been convincingly demonstrated that segmental properties, too, show comparable behavior, if on a more limited scale. For example, in many languages the feature [nasal] may take up only part of a segment, giving rise to pre- and post-nasalized stops such as [^{n}d] and [d^{n}]; and in some languages it regularly spreads across more than one segment or syllable, establishing domains of nasal harmony (see, e.g., Bendor-Samuel 1970; Lunt 1973; Anderson 1976). Similarly, in languages with vowel harmony, features such as [back], [round] and [ATR] (advanced tongue root) have the ability to extend across many syllables at a time (see, e.g., Welmers and Harris 1942; Carnochan 1970; Vago 1980). Other segmental features also show nonlinear properties, as we shall see in the later discussion.

Problems such as these offered a direct challenge to linear theories of phonological representation, and led to the development of alternative, nonlinear frameworks.[2] The earliest of these were the theory of long components developed by Harris (1944) (see also Hockett 1942, 1947 for a similar approach) and the theory of prosodic analysis developed by J. R. Firth and his collaborators after World War II (see, e.g., Firth 1948, the Philological Society 1957, and Palmer 1970). A more recent and still evolving approach is the theory of dependency phonology developed by J. Anderson, C. Ewen, and their associates (for a general overview and fuller discussion, see Anderson and Ewen 1987 and also chap. 17, this volume).

Perhaps the most influential of these frameworks at the present time – and the one we will be primarily concerned with here – is an approach emanating from the theory of autosegmental phonology developed in the 1970s and early 1980s. In autosegmental phonology, as first presented by Goldsmith (1976, 1979a, 1979b), features that are observed to extend over domains greater or lesser than the single segment are extracted from feature matrices and placed on separate "channels" or tiers of their own. Thus tones, for example, are represented on a separate tier from vowel and consonant segments, where they are able to function in a partly autonomous fashion. Elements on the same tier are sequentially ordered, while elements on different tiers are unordered and related to each other by means of association lines which establish patterns of alignment and overlap. Since associations between tones and tone-bearing units are not necessarily one-to-one, we may find other types of linking, as shown below (H = high tone, L = low tone, and V = any tone-bearing unit, such as a vowel or syllable):

(2)



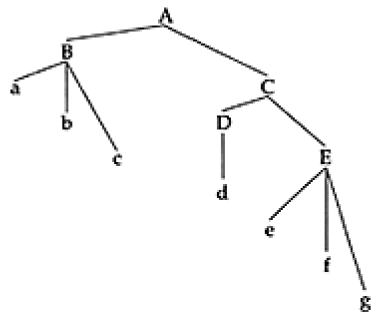
Only (2a) involves a one-to-one relation between tones and tone-bearing units of the sort admitted in linear theories. (2b) shows a vowel linked to two tones, constituting a falling tone, (2c) displays two vowels sharing a single tone, and (2d) illustrates a floating tone. Multitiered representations of this type can be extended to other features showing complex patterns of alignment, such as nasality and harmonically-operating vowel features (Goldsmith 1979a; Clements 1980; Clements and Sezer 1982).

A second problem inherent in a matrix-based approach is its implicit claim that feature bundles have no internal structure. Each feature is equally related to any other, and no features are grouped into larger sets, corresponding to traditional phonetic classes such as “place” or “manner” of articulation. This claim is an intrinsic consequence of the way the representational system is designed.[3] Some linguists, however, have proposed to classify phonological features into taxonomic categories. While they have not usually assigned any status to such categories in phonological representations themselves, they have sometimes suggested that they may have a cognitive status of some sort. Thus, while Jakobson and Halle (1956) group segmental features into “sonority” and “tonality” features on strictly acoustic grounds, they suggest that these classes form two independent “axes” in language acquisition. Chomsky and Halle (1968) classify features into several taxonomic classes (major class features, cavity features, etc.), but suggest that “ultimately the features themselves will be seen to be organized in a hierarchical structure which may resemble the structure we have imposed on them for purely expository reasons” (1968, p. 300). The most extensive earlier proposal for grouping features together into larger classes, perhaps, is that of Trubetzkoy (1939), whose “related classes” of features are defined on both phonetic and phonological principles. To take an example, the features of voicing and aspiration fall into a single related class on phonetic grounds, as they are both realized in terms of laryngeal activity, independently of the oral place of articulation. But these features also function together phonologically, in the sense that they frequently undergo neutralization as a unit (see further discussion in section 2.3), or exhibit tight patterns of mutual implication. Trubetzkoy assigns such classes of features to separate “planes” of structure, and relates their independent cognitive (psychological) status to their phonetic and functional relatedness, stating that “the projection of distinctive oppositions (and thus also of correlations) sometimes onto the same and sometimes onto different planes is the psychological consequence of just those kin relationships between the correlation marks on which the classification of correlations into related classes is made” (p. 85). These pregnant suggestions did not undergo immediate development in the Jakobsonian or generative traditions, as we have seen. However, Trubetzkoy's conception can be viewed as an important precursor of the model which we examine in greater detail below.

There is, indeed, a considerable amount of evidence that features are grouped into higher-level functional units, constituting what might be called “natural classes” of features in something very like Trubetzkoy's notion of “related classes.” For example, in contemporary English, /t/ is often glottalized to [tʼ] in syllable-final position. In some contexts, glottalized [tʼ] loses its oral occlusion altogether, yielding the glottal stop that we observe in common pronunciations of words like mitten [mi$^{?}$n]. In certain Spanish dialects, the nonlaryngeal features of /s/ are lost in syllable coda position, leaving only aspiration behind: *mismo* [mi^{h}mo] “same”. In cases like these, which have been discussed by Lass 1976, Thráinsson 1978, and Goldsmith 1979b, the oral tract features of a segment are lost as a class, while laryngeal features such as glottalization and aspiration remain behind. Similarly, in many languages all place features function together as a unit. In English, the nasal segment of the prefixes *syn-* and *con-* typically assimilates in place of articulation to the following consonant, where it is realized as labial [m] before labials (*sympathy, compassion*), alveolar [n] before alveolars (*syntax, condescend*), velar [ŋ] before velars (*synchronize, congress*), and so forth. In such processes, all features defining place of articulation function as a unit, suggesting that they have a special status in the representation (Goldsmith 1981; Steriade 1982; Mohanan 1983). To describe processes such as this, traditional linear theory requires a rule mentioning all the features designating place of articulation, i.e., [coronal], [anterior], [distributed], [back], etc. But such a rule is no more highly valued than one that involves any arbitraily-selected set of features, including those that never function together in phonological rules.

In response to this problem, a general model of feature organization has been proposed in which features that regularly function together as a unit in phonological rules are grouped into *constituents* (Clements 1985; Sagey 1986; see also Hayes 1986a for a related approach). In this approach, segments are represented in terms of hierarchically-organized node configurations whose terminal nodes are feature values, and whose intermediate nodes represent constituents. Instead of placing features in matrices, this model arrays them in the manner of a Calder mobile, as shown below:

(3)



Unlike the tree diagrams familiar in syntactic theory, terminal elements (here, feature values) are unordered and placed on separate tiers, as we suggest in the diagram by placing them on separate lines. This organization makes it possible to express feature overlap, as in standard autosegmental phonology. All branches emanate from a *root node*, (A), which corresponds to the speech sound itself. Lower-level *class nodes* (*B, C, D, E*) designate functional feature groupings, which include the laryngeal node, the place node, and others to be discussed below.

In this model, association lines have a double function. They serve first to encode patterns of temporal alignment and coordination among elements in phonological representations, as in autosegmental phonology (cf. (2)). The importance of this function will be seen later in the discussion of contour segments (section 1.3), length (section 1.4), and multilinked nodes (section 2.1), for example. In addition, as shown in (3), they group elements into constituents, which function as single units in phonological rules. The immediate constituents of such a grouping are sister nodes, and both are daughters, or dependents, of the higher constituent node; in (3), for example, *D* and *E* are sisters, and daughters (or dependents) of *C*. Notice also that if *D* is (universally) a daughter of *C*, the presence of *D* in a representation will necessarily entail the presence of *C*, a relationship that will take on some importance in the later discussion.

This approach to feature organization makes it possible to impose strong constraints on the form and functioning of phonological rules. In particular, we assume the following principle:

(4) Phonological rules perform single operations only.

This principle predicts, for example, that a phonological rule might affect the set of features *d, e, f*, and *g* in (3) by performing a single operation on constituent *C*; however, no rule can affect nodes *c, d*, and *e* alone in a single operation, since they do not form a constituent. In general, a theory incorporating this principle claims that *only feature sets which form constituents may function together in phonological rules*. Since the set of features that form constituents is a very small proportion of all the logical possibilities, this claim represents a strong empirical hypothesis regarding the class of possible phonological rules.

One further principle is required in order to maintain this claim in its most general form. We state it as follows:

(5) Feature organization is universally determined.

According to this principle, the manner in which feature values are assigned to tiers and grouped into larger constituents does not vary from language to language. Obviously, if feature organization could freely vary, the theory would make no crosslinguistic predictions. However, there is much reason to believe that feature organization is universal, since the same feature groupings recur in language after language.[4]

We further assume that principle (5) projects the same feature organization at al levels of derivation, from underlying to surface structure. This means that phonological rules cannot have the effect of creating novel types of feature organization. Rules that would produce ill-formed structures are often assumed to be subjected to further conventions which have the effect of preserving the wellformedness of the representation (e.g., node interpolation, Sagey 1986). Thus, the feature hierarchy operates as a template defining well-formedness across derivations. Further principles constraining the form and organization of phonological rules will be discussed as we proceed.

We may now give a preliminary answer to the question, How are features organized?

(6) (a) Feature values are arrayed on separate tiers, where they may enter into nonlinear (nonbijective) relations with one another;
(b) Features are at the same time organized into hierarchical arrays, in which each constituent may function as a single unit in phonological rules.

A model having these general properties has been called a "feature geometry."[5] On these assumptions, the empirical task of feature theory is that of determining which nodes to recognize, and how these nodes are organized.

1 Simple, Complex, and Contour Segments

We now develop a theory of feature organization in more detail. We first take up the question of gross segmental structure, focusing on the characterization of simple, complex and contour segments. Here and elsewhere, our discussion of particular examples will be necessarily brief and incomplete, and the reader is urged to consult our sources for fuller discussion.

1.1 Articulator-based Feature Theory

Central to the current development of feature theory is the idea that speech is produced using several independently functioning articulators. These articulators – comprising the lips, the tongue front, the tongue body, the tongue root, the soft palate, and the larynx – may define a single, primary constriction in the vocal tract, or may combine to produce several constrictions at the same time. Since the articulators play a fundamental role in the organization of segment structure, it has been proposed that they should be represented by nodes of their own in phonological representations, arrayed on separate tiers (Sagey 1986; Halle 1988). Among these nodes, *labial, coronal*, and *dorsal* are defined in terms of oral tract articulations, as stated below (Sagey 1986, p. 274):

(7)

Labial: involving the lips as an active articulator

Coronal: involving the tongue front as an active articulator

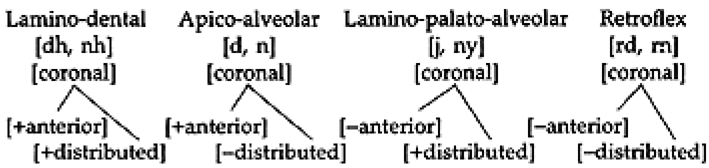
Dorsal: involving the tongue body as an active articulator

The articulator features are also called "place" features, because they link under the place constituent in the feature hierarchy.[6]

Unlike most other features, [labial], [coronal], and [dorsal] are treated as privative (one-valued), rather than binary. This is because phonological rules do not appear to operate on the negative values of these categories. For example, while many rules involve labial assimilation, there are few if any rules of nonlabial assimilation, turning, e.g., [p] to [t] in the context of a nonlabial sound ([t, č, k], etc.), or rules of nonlabial dissimilation, changing, e.g., nonlabial [t] to labial [p] next to a nonlabial. These observations follow directly from the assumption that articulator features are one-valued: if [-labial] has no existence in the theory, then of course no rule can carry out an operation on [-labial] sounds.[7]

Other features are either *articulator-bound*, in the sense that they depend on a specific articulator for their execution, or *articulator-free*, in the sense that they are not restricted to a specific articulator. Articulator-bound features, when present, further prescribe the specific nature of the constriction formed by a given articulator. Such features are located under the appropriate articulator node. Thus, for example, the articulator-bound features [anterior] and [distributed] are linked under the coronal node, where they distinguish anterior from posterior coronals and apical from laminal coronals, respectively. We illustrate these distinctions with a system of minimal contrasts found in many Australian languages (Dixon 1980; his digraphs represent simple sounds in all cases):

(8)



The placement of the features [anterior] and [distributed] directly under the coronal node is motivated by several observations. First, these features are relevant only for coronal sounds. Thus no noncoronal sounds are minimally distinguished by these features, nor do these features define natural classes including noncoronal sounds (Steriade 1986; Sagey 1986); these observations follow directly from the treatment of [anterior] and [distributed] as dependents of the [coronal] node, since the presence of either feature in a segment entails the presence of [coronal]. Second, this analysis correctly predicts that if one segment assimilates to another in coronality, it necessarily assimilates [anterior] and [distributed] at the same time. This prediction is supported by rules of coronal assimilation in languages as diverse as English (Clements 1985), Sanskrit (Schein and Steriade 1986), Basque (Hualde 1988b), and Tahltan (Shaw 1991).

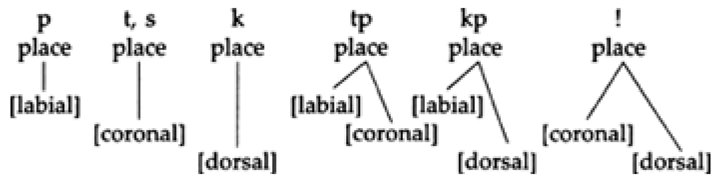
Articulator-free features designate the degree of stricture of a sound, independent of the specific articulator involved, and so are sometimes called *stricture* features. For example, [+continuant] sounds are those that permit continuous airflow through the center of the oral tract, regardless of where the major stricture is located. The features [±sonorant], [±approximant], and [±consonantal] also lack a designated articulator. Most writers place the articulator-free features higher in the hierarchy than articulator features; we will examine evidence supporting this view below.

1.2 Simple and Complex Segments

If features can be considered the atoms of phonological representation, feature complexes constituting segments may be considered the *molecules*. We now consider molecular structure in more detail. Drawing on terminology introduced by Sagey, we can distinguish between simple, complex, and contour segments. A *simple* segment consists of a root node characterized by at most one oral articulator feature. For example, the sound [p] is simple since it is uniquely [labial].

A *complex* segment is a root node characterized by at least two different oral articulator features, representing a segment with two or more simultaneous oral tract constrictions. This analysis receives striking support from Halle's observation (1983) that we can find doubly articulated complex segments involving all possible pairs of oral articulators, as defined by the articulator features *labial, coronal*, and *dorsal*. For example, the labio-coronal stop [tp] of Yeletnye is formed by simultaneous closure of the lips and tongue front (Maddieson and Ladefoged 1988), the labiovelar stop [kp] of Yoruba by simultaneous closure of the lips and tongue body (Ladefoged 1968), and the alveolar click [!] of South African Bantu and Khoisan languages by simultaneous closure of the tongue front and tongue body (Ladefoged and Traill 1984). Representations of several simple and complex segments are given below, showing relevant structure only. (Recall that nodes on different tiers are unordered with respect to each other. We disregard the distinction between major and minor articulations, to be discussed in section 3.5.)

(9)



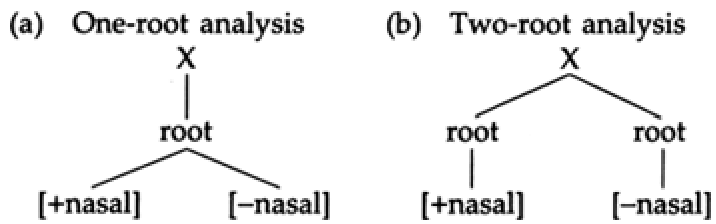
It will be appreciated that Halle's original observation follows directly from the articulator-based model on the assumption that complex segments are formed by the free combination of oral articulator features: since there are only three of these, we should find exactly the three combinations illustrated above.

1.3 Contour segments

Parallel to the treatment of contour tones, depicted in (2b) above, multitiered feature representations allow the direct expression of *contour segments*, that is, segments containing sequences (or "contours") of different features. The classical motivation for recognizing contour segments is the existence of phonological "edge effects," according to which a given segment behaves as though it bears the feature [+F] with regard to segments on one side and [-F] with regard to those on the other (Anderson 1976). Commonly proposed candidates for such segment types include affricates and prenasalized stops.

There are currently two main views on how such segments can be characterized, as suggested by the following figure, representing prenasalized stops (irrelevant structure has been omitted):

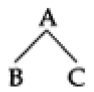
(10)



In the one-root analysis (10a), contour segments are characterized by a sequence of features linked to a single higher node (Sagey 1986); in this view, a prenasalized stop such as [^{n}d] is represented as a single root node characterized by the sequence [+nasal] [-nasal],[8] and an affricate such as [ts] is represented as a root node characterized by the sequence [-continuant] [+continuant]. This analysis assumes that only terminal features, not class nodes, may be sequenced in a given segment. Notice, however, that even with this constraint, a large number of theoretically possible but nonoccurring, complex segments are predicted, bearing such sequences as [+voice] [-voice] or [-distributed] [+distributed].

In the two-root analysis (10b), contour segments consist of two root nodes sequenced under a single skeletal position. In this view, a prenasalized stop can be represented as a sequence of two root nodes, characterized as [+nasal] and [−nasal] respectively (Clements 1987; Piggott 1988; Rosenthall 1988). This analysis assumes a constraint that universally forbids branching structure under the root node, which can be stated as follows:

(11) The No Branching Constaint:
Configurations of the form



are ill-formed, where A is any class node (including the root node), A immediately dominates B and C, and B and C are on the same tier.

This statement is based on a proposal by Clements (1989b), generalizing a more specific version proposed by Piggott (1988).[9] Even with this constraint, however, further principles are required to express the fact that not every sequence of root nodes constitutes a possible contour segment (for suggestions, see Rosenthall 1988).

In an important study of nasal spreading phenomena, Steriade (1991) presents extensive evidence in favor of a two-root analysis of pre- and post-nasalized stops, while taking an important step toward constraining the class of potential contour segments. Steriade proposes that contour segments (including released stops, in her analysis) should be analyzed as sequences of what she terms "aperture nodes." She recognizes three kinds of aperture nodes, defined phonetically as follows:

(12)

A_0 = total absence of oral airflow (as in oral and nasal stops)
A_f = degree of oral aperture sufficient to produce a turbulent airstream (as in fricatives and the second phase of affricates)
A_{max} = degree of oral aperture insufficient to produce a turbulent airflow (as in oral sonorants and the release phase of stops)

Steriade suggests that such "aperture nodes" can be incorporated into the phonological feature model as root nodes characterized by the appropriate feature values. Inspiring ourselves freely on this proposal, we suggest the following root-node interpretation (others are possible):[10]

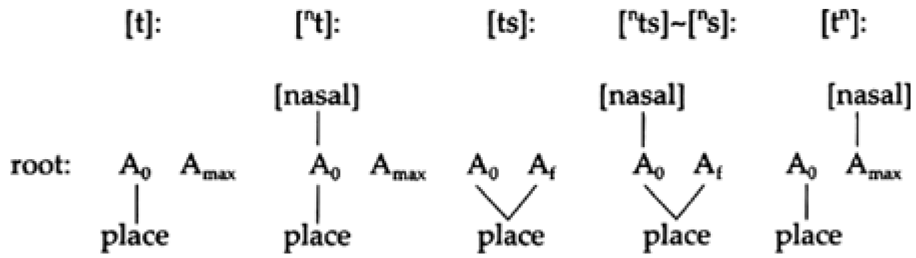
(13)

A_0 = a root node characterized as [−continuant, −approximant]
A_f = a root node characterized as [+continuant, −sonorant]
A_{max} = a root node characterized as [+continuant, +sonorant]

This model allows only two types of segment-internal sequences: A_0 A_b, defining segments with a stop phase followed by fricative release (i.e., affricates), and A_0 A_{max}, defining segments with a stop phase followed by abrupt maximal release (all other released stops).

Given these assumptions, and continuing to use the shorthand notation in (13), we may represent oral (released) stops, affricates, prenasalized stops, and prenasalized affricates, respectively, by the partial configurations shown in (14) (only relevant structure is shown; we place the feature [nasal] above the root nodes for convenience):

(14)



The first two figures show plain and prenasalized stops, respectively. The prenasalized stop [^{n}t] differs from its oral counterpart [t] only in having the feature [nasal] attached to its noncontinuant root node A_0; it is thus formally analyzed as nasal closure followed by (maximal) oral release. The third figure shows an oral affricate, analyzed as a noncontinuant root node A_0 followed by a (homorganic) fricative release A_f.

The fourth figure provides a single representation for the prenasalized segments often transcribed as [^{n}ts] and [^{n}s]. The latter, usually described as a prenasalized fricative, is here represented just like the former, as an affricate with [nasal] closure. This analysis assumes that the articulatory difference between [^{n}s] and [^{n}ts] is always nondistinctive, being determined by language-particular principles of phonetic implementation. The last segment type shown in (14), the postnasalized stop [t^n], can quite naturally be characterized as a [t] with maximal [nasal] release. Ordinary nasal stops such as [n] (not shown here) differ from it only in that [nasal] is associated with both the closure and release nodes.

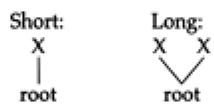
In Steriade's proposal, true contour segments are restricted to the class of stops and affricates, as shown above. Other major classes, such as fricatives, liquids, and vocoids, have only a single root node, and thus cannot be phonologically pre- or post-nasalized.[11] Steriade's system, then, is a highly constrained one which restricts contour segments to just a few, well-attested types; it remains to be seen whether other complex segment types sometimes proposed in the literature, such as short diphthongs, should be added to the inventory.

Affricates are less well understood than prenasalized stops. While “edge effects” have been convincingly demonstrated for pre- and post-nasalized stops, they are much less evident for affricates. Indeed, in some languages, such as Basque, Turkish, and Yucatec Mayan, affricates show “anti-edge effects,” behaving as stops with respect to following segments and/or as fricatives with respect to preceding segments. The formal analysis of affricates remains an unresolves question at the present time (see Hualde 1988a, Lombardi 1990, and Steriade 1991 for several different proposals).

1.4 Length

Speech sounds may be long or short. Phonological length (or quantity) can be defined as *bipositionality* on the tier representing phonological quantity, whether this is taken as the CV- or X-skeleton in the sense of McCarthy (1981, 1985), Clements and Keyser (1983), Prince (1984), and others, or the weight unit tier in the sense of Hyman (1985).[12] In all these approaches, a long consonant or vowel is represented as a root node linked to two units of quantity, as shown in (15):

(15)



A surprising result of this analysis is that we no longer have a uniform way of reconstructing the traditional notion "segment". Thus a complex segment such as [ts] consists of one node on the quantity tier and two on the root tier, while a long consonant such as [t:] consists of two nodes on the quantity tier and one on the root tier. On which tier is segmenthood defined? Neither choice seems fully appropriate. The apparent paradox may simply reflect the fact that we are dealing with different kinds of segmentations on each tier. It might be more useful to distinguish between "melodic segments" defined on the root tier and "metric segments" defined on the skeleton; in this way an affricate would consist of two melodic segments linked to one metric segment, and so forth.

2 Phonological Processes

A classical problem in phonological theory is that of determining the class of elementary phonological processes which map underlying representations into surface representations. Standard generative phonology provided a rich vocabulary for stating phonological rules, but as Chomsky and Halle themselves pointed out (1968, chap. 9), it did not provide an intrinsic way of distinguishing plausible, crosslinguistically attested rules from highly improbable ones (for further discussion of this point, see, e.g., Clements 1976, Goldsmith 1981, and McCarthy 1988). In reaction to this problem, the theory of natural phonology developed a set of criteria for distinguishing between "natural processes" and "learned rules" (Stampe, 1980), but did not provide a formal basis for the distinction.

In this section we take up this issue from the perspective of hierarchical feature representation. We show that a small number of elementary rule types and organizational principles project to a large class of "natural" rule types, while excluding rare or unattested ones.

2.1 Assimilation

Perhaps the most widely recurrent type of phonological rule is *assimilation*. Standard generative phonology characterized assimilation in terms of feature copying, according to which one segment copies feature specifications from a neighboring segment. In the present model, in contrast, assimilation rules are characterized as the association (or "spreading") of a feature or node F of segment A to a neighboring segment B, as shown below (dashed lines indicate association lines added by rule.):

(16)

This approach represents the phonological counterpart of the articulatory model of assimilation assumed by the French phonologist Grammont, who writes (1933, p. 185): *"L'assimilation* consiste dans l'extension d'un ou de plusieurs mouvements articulatoires au delà de leur domaine originaire. Ces mouvements articulatoires sont propres au phonème agissant; le phonème agi, en se les appropriant aussi, devient plus semblable à l'autre." In this view, an assimilation involves one or more articulatory movements extending their domain from an affecting segment (*phonème agissant*), or trigger, to an affected segment (*phonème agi*), or target. There is considerable phonological support for such a view, as we shall see in section 2.1.2.

2.1.1 Assimilation Types

As a basis for the discussion, it will be useful to distinguish various assimilation types. One distinction depends on the nature of the affected segment. If the rule spreads only feature(s) that are not already specified in the target, it applies in a *feature-fiiling* mode. This common pattern can be regarded as the unmarked (or default) mode of assimilation. If the rule applies to segments already specified for the spreading feature(s), replacing their original values, the rule applies in a *feature-changing* mode.

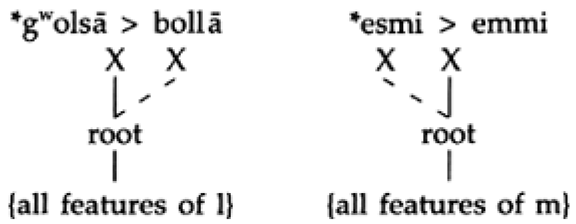
We can also distinguish different types of assimilation according to the identity of the spreading node. If the root node spreads, the affected segment will acquire all the features of the trigger. In the feature-changing mode, this result, often called *complete* or *total assimilation*, gives the effect of deletion with compensatory lengthening. For example, in a well-known sound change in the Lesbian and Thessalian dialects of Ancient Greek, [s] assimilates to a preceding or following sonorant, perhaps passing through an intermediate [h] (Steriade 1982; Wetzels 1986; Rialland 1993):

(17)

*gʷolsā	>	bollā	"council"
*awsōs	>	awwōs	"dawn"
*esmi	>	emmi	"I am"
*naswos	>	nawwos	"temple"

These assimilations illustrate both cases of schema (16), where F = the root node:

(18)



The spreading root node replaces the root node of [s], which is deleted by convention.

If a lower-level class node spreads, the target acquires several, but not all of the features of the trigger (*partial* or *incomplete* assimilatio). We have already mentioned an example from English, involving the assimilation of place features in the prefixes *syn-* and *con-*. Similar rules occur in many languages, and have the general form shown in (19):

(19)

Finally, in *single-feature* assimilation, only a terminal feature spreads. Common types include vowel harmony, voicing assimilation, and nasal assimilation. Many examples have been presented from the earliest literature in nonlinear phonology onward (see, e.g., representative papers in van der Hulst and Smith 1982, and Aronoff and Oehrle 1984).

Assimilation rules provide a powerful criterion for answering the question, How are features organized?, since any feature or feature set that assimilates as a unit must constitute a node on an independent tier of its own. We now consider an important empirical prediction of the spreading model of assimilation.

2.1.2 Multilinked Nodes

In the spreading model of assimilation, an assimilation rule always gives rise to multilinked nodes in its output. Notice, for example, that in the output of a total assimilation rule a single root node is linked to two skeletal positions, as shown in (18). It will be recalled that this type of representation is identical to the one proposed earlier for underlying long segments (section 1.4). The spreading model of assimilation predicts, therefore, that geminates derived by assimilation rules should be formally indistinguishable from underlying geminates with regard to later stages of a phonological derivation.

This prediction appears to be correct. Hayes (1986b) shows that geminates created from dissimilar segments by assimilation rules share special properties with underlying (i.e., monomorphemic) geminates, which are not displayed by sequences of identical consonants occurring at the boundary between different morphemes or words. One of these is *geminate inseparability*, according to which "true" geminates (i.e., those exhibiting multilinked structure) cannot be separated by epenthetic vowels. This property can be illustrated by rules of epenthesis and assimilation in Palestinian Arabic, as first described by AbuSalim (1980). In words containing clusters of three or more consonant positions, an epenthetic vowel [i] is inserted before the final two: (C)CCC → (C)C[i]CC. This rule is illustrated in (20a). The rule does not apply if the leftmost two positions correspond to a monomorphemic geminate, as shown in (20b). Epenthesis also applies within clusters separated by word boundaries (20c), even if the flanking consonants are identical (20d). (20e) illustrates epenthesis between the definite article /l-/ and a following cluster. By an independent rule, /l-/ totally assimilates to a following coronal consonant, giving surface geminates as in [š-šams] "the sun" from underlying /l-šams/. Geminates created by this rule cannot be separated by epenthesis; instead, contrary to the regular pattern, the epenthetic vowel is inserted to their left (20f).

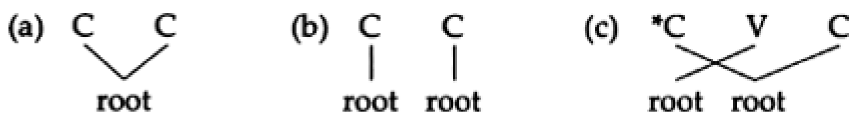
(20)

(a)	ʔak[i]l-kum	"your food" /ʔakl-kum/
	ʔib[i]n-ha	"her son" /ʔibn-ha/
(b)	sitt-na	"our grandmother" (*sit[i]t-na)
	ʔimm-na	"our mother" (*ʔim[i]m-na)
(c)	ʔakl [i] mniiḥ	"a good food"
	walad [i] kbiir	"a big boy"
(d)	samak [i] kbiir	"a big fish"
(e)	l-walad l[i]-kbiir	"the big boy" /l-kbiir/
(f)	l-walad [i] z-zɣiir	"the small boy" (* z[i]-zɣiir])

In this paradigm, the inseparable geminates are just those that are monomorphemic (20b), or heteromorphemic and created by assimilation (20f).

This pattern can be explained on the assumption that "true" geminates, whether underlying or created by assimilation, have the multilinked structure shown in (21a), while "accidental" geminates created by concatenation across boundaries have separate root nodes, as shown in (21b). The failure of epenthesis to apply in true geminates can be explained by the fact that the insertion of an epenthetic vowel into the linked structure would create a violation of the constraint against crossed association lines (discussed in section 2.5), as shown in (21c):[13]

(21)



Not only total, but partial assimilation gives rise to multilinked nodes, as was shown in (19). Partially assimilated clusters should therefore show inseparability effects just as full geminates do. This prediction is also well supported by the evidence. For example, in Kolami, clusters that have undergone place assimilation are impervious to a later rule of epenthesis that would otherwise be expected to break them up (see Steriade 1982, after Emenau 1955). Thus, epenthesis normally inserts a copy of the stem vowel between the first two members of a CCC cluster, as shown by the first column in (22a). However, it fails to apply in homorganic clusters created by place assimilation, as shown in (22b).

(22)

		past	present	UR (root)
(a)	"break"	kinik-tan	kink-atun	/kink/
	"make to get up"	suulup-tan	suulp-atun	/suulp/
	"sweep"	ayak-tan	ayk-atun	/ayk/
(b)	"boil over"	ponk-tan	pong-atun	/pong/
	"bury"	min(t)-tan	mind-atun	/mind/

This behavior may be explained by attributing a multilinked place node to the assimilated clusters; as in the case of geminates, epenthesis would create an illformed structure.[14] Similar effects have been cited from Tamazight Berber (Steriade 1982), Sierra Popoluca (Clements 1985), and Barra Isle Gaelic (Clements 1986), among other languages.

The multiple linking of partially assimilated clusters can be demonstrated in other ways as well. For example, partly assimilated clusters show the same sort of inalterability effects that are found in true geminates, according to which certain types of rules that ordinarily affect the feature content of single segments fail to apply to otherwise eligible segments with linked structure; see Hayes (1986a, 1986b) and Schein and Steriade (1986) for discussion and examples from a variety of languages. Again, phonological rules are frequently restricted to apply only to members of partially assimilated homorganic clusters; such rules can be simply formulated by making direct reference to the linked place nodes (see Kiparsky 1985, Clements 1985, and Hume 1991 for examples from Catalan, Sierra Popoluca, and Korean, respectively). Another argument for linked structure can be cited from the many languages which restrict intervocalic consonant clusters to geminate consonants (if present in the language) and homorganic clusters; in such cases we may say that intervocalic clusters may only have one place node (Prince 1984, p. 243). In sum, there are many independent types of evidence supporting the spreading theory of assimilation, which taken together provide a strong source of support for nonlinear feature representation as outlined above.

2.2 Dissimilation and the OCP

We now consider dissimilation, the process by which one segment systematically fails to bear a feature present in a neighboring (or nearby) segment. Dissimilation rules are also common across languages, and should receive a simple formal expression in the theory.

Traditionally, dissimilation has been stated in terms of feature-changing rules of the type [X] → [-F] / ___ [+F]. However, this approach cannot be adopted in the present framework, since many features that commonly undergo dissimilation ([coronal], [labial], [dorsal], etc.) are one-valued. Instead, dissimilation can be expressed as an effect of delinking, according to which a feature or node is delinked from a segment; the orphaned node is then deleted through a general convention. A later rule may insert the opposite (typically, default) value. (See Odden 1987, McCarthy 1988, and Yip 1988 for examples and discussion.)

While dissimilation can be formally expressed as delinking, we must still explain why delinking so commonly has a dissimilatory function. A rather elegant answer comes from the Obligatory Contour Principle (OCP), originally proposed in work on tone languages to account for the fact that sequences of identical adjacent tones, such as HH, are widely avoided in both underlying and derived representations (Leben 1973). In later work, McCarthy extended this principle to the segmental phonology to explain why so many languages avoid sequences of identical (or partly identical) segments (McCarthy 1986). He stated this principle in its most general form as follows (McCarthy 1988):

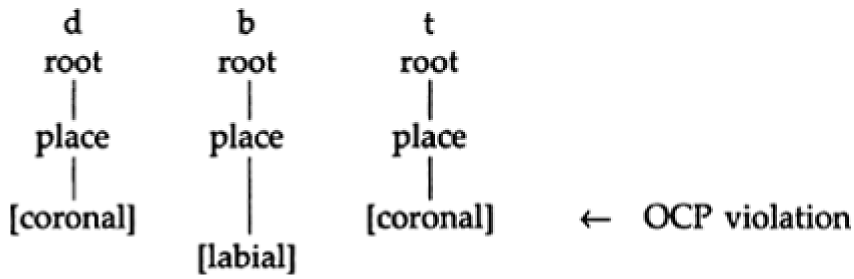
(23) Obligatory Contour Principle (OCP):
Adjacent identical elements are prohibited.

By this statement, the OCP applies to any two identical features or nodes which are adjacent on a given tier. Its empirical content is threefold: it may prohibit underlying representations which violate it, it may "drive" or motivate rules which suppress violations of it, and it may block rules that would otherwise create violations of it (see McCarthy 1981, 1986, 1988; Mester 1986; Odden 1988; Yip 1988, 1989; Clements 1990b, 1993, for examples and discussion). A direct consequence of the OCP is that dissimilatory delinking should be a preferred process type across languages, since it has the effect of eliminating OCP violations.

We will illustrate the OCP with a well-known example from Classical Arabic (Greenberg 1950; McCarthy, in press). In this language, consonantal roots are subject to strict constraints. First, within such roots, no two consonants can be identical. Thus hypothetical roots such as /bbC/, /Cbb/ and /bCb/ are ill-formed, where C is any consonant.[15] Furthermore, roots containing homorganic consonants strongly tend to be excluded; thus hypothetical roots like /bmC/, /Cbm/, and /bCm/, with two labial consonants, are totally absent (see McCarthy, in press, for qualifications and fuller discussion).

Consider now how these constraints can be accounted for by the OCP. The constraint against identical adjacent consonants follows directly from statement (23), applied at the root tier. The constraint against homorganic consonants also follows from (23), applied in this case to articulator features. Consider the following (partial) representation of the ill-formed root*/dbt/, with homorganic initial and final consonants:

(24)



The illformedness of this representation is due to the violation of the OCP on the [coronal] tier, as shown by the arrow. Crucially, the two occurrences of [coronal] are adjacent on their tier, even though the segments they characterize, /d/ and /t/, are nonadjacent on the root tier. This is because the inter-vening consonant, /b/, is characterized by a [labial] node, which, lying on a tier of its own, is unordered with respect to [coronal]. This result crucially presupposes that articulator features are arrayed on separate tiers, as proposed earlier.

Thus in Arabic, and in many other languages, the OCP generates a pervasive pattern of dissimilation involving identical and homorganic consonants. In many languages, OCP violations are resolved in other ways as well, such as the merger or assimilation of adjacent identical nodes (Mester 1986), the blocking of syncope rules that would otherwise create OCP violations (McCarthy 1986), and the insertion of epenthetic segments, as in the English plural formation rule which inserts a vowel between two coronal sibilants in words like *taxes, brushes* (Yip 1988). This evidence, taken cumulatively, suggests that dissimilation is just one of several stratagems for reducing or eliminating OCP violations at all levels of representation. Like assimilation, dissimilation (and other delinking rules) provide a criterion for feature organization: any delinked node must occur on a tier of its own.

2.3. Neutralization

Another common process type is neutralization, which eliminates contrasts between two or more phonological features in certain contexts (Trubetzkoy 1939). We are concerned here with neutralization rules which are neither assimilations nor dissimilations. Common examples include rules of debuccalization (Clements 1985; McCarthy 1988; Trigo 1988) which eliminate contrasts among oral tract features; rules of devoicing, deaspiration, and/or deglottalization, which eliminate contrasts among laryngeal features (Lombardi 1991); and rules of vowel height reduction which reduce or eliminate constrasts in height or [ATR] (Clements 1991). Neutralization at the level of the root node eliminates all segmental constrasts, as in the reduction of all unstressed vowels to a neutral vowel (as in English), or of certain consonants to a "default" element such as [?] (as in Toba Batak, see Hayes 1986a). Typically, neutralization rules eliminate marked values in favor of unmarked values.

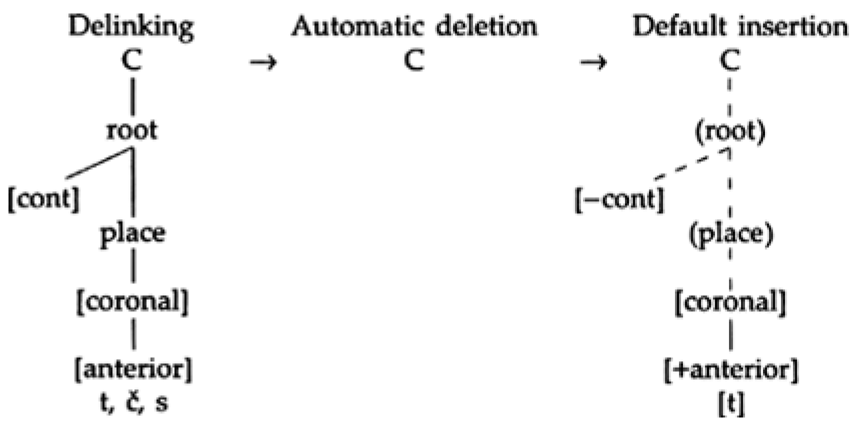
Like dissimilation, simple neutralization can be characterized in terms of node delinking. We illustrate with a particularly interesting example from Korean. In this language, the three-way phonemic contrast among plain voiceless, aspirated, and "tense" (or glottalized) obstruents is neutralized to a plain voiceless unreleased stop in final position and preconsonantally (i.e., in the syllable coda). In addition, the coronal obstruents /t t^h t′ č $\check{c}^h$ č′ s s′ / and (at least for some speakers) /h/ are neutralized to [t] in the same contexts. In faster or more casual speech styles, however, the coronals may totally assimilate to a following stop under conditions which appear to vary among speakers. The two styles are illustrated by the following examples (Martin 1951; Cho 1990; Kim 1990):[16]

(25)

	/-e/ "in"	/-kwa/ "and" slower		faster	
/pat^h/	pat^h-e	pat-k′wa	or	pak-k′wa	"field"
/os/	os-e	ot-k′wa	or	ok-k′wa	"clothes"
/čəč/	čəč-e	čət-k′wa	or	čək-k′wa	"mother's milk"
/k′oč^h/	k′oč^h-e	k′ot-k′wa	or	k′ok-k′wa	"flower"

Notice that the neutralization rule illustrated in the slower speech styles applies to the features [anterior] and [continuant] which, in the feature organization of Sagey (1986), for example, are widely separated: [anterior] is dominated by [coronal], whereas [continuant] is immediately dominated by the root node. To achieve this effect in terms of a single operation in accordance with principle (4), the rule must delink the root node of the coronal obstruent.[17] The resulting empty skeletal position is assigned the features of unreleased [t], the unmarked consonant, by default. The following derivation of the slower speech forms illustrates the analysis of neutralization; note that parenthesized nodes are automatically interpolated to preserve wellformedness (Sagey 1986). (Irrelevant structure is omitted.)

(26)



In faster speech, the default rule is preempted by a rule spreading the root node of the second consonant onto the skeletal position of the first.

Neutralization rules provide a further criterion for feature organization: since only single nodes may undergo delinking, any features that delink as a group must constitute a single node on an independent tier of its own.

2.4 Other Elementary Rule Types

The elementary rule types required for the processes described above are linking, delinking, and default insertion. A further process, deletion, can usually be decomposed into delinking followed by automatic deletion. This brief list is probably not complete. Feature-changing rules, affecting values of features such as [sonorant], [consonantal], and [continuant], are most likely required to express processes of strengthening and weakening, and nondefault feature insertion rules are sometimes needed to express the introduction of marked feature values. Other possible rule types include fusion (or merger), proposed to account for various types of feature coalescence processes (Mester 1986; Schane 1987; de Haas 1988), and fission, designed to account for diphthongization and other types of "breaking" phenomena (Clements 1989b). Among these various rule types, however, those which reorganize patterns of association among existing nodes (spreading, delinking) appear to represent the least marked case.

To summarize, the feature theory presented here assumes a small set of elementary rule types which carry out single operations on feature representations. It adopts the strong hypothesis that all genuine phonological rules fall into one of these elementary types. This result takes us a step closer to the elusive goal of characterizing the class of "natural" rules in formal terms.

2.5 Transparency and Opacity

Another classical issue in phonological theory is that of delimiting the domain within which rules may apply. It has long been known that rules may affect not only adjacent segments, but also segments that occur at some distance from each other. For example, rules of vowel harmony and assimilation typically apply from vowel to vowel, regardless of intervening consonants (see Clements and Sezer 1982, McCarthy 1984, van der Hulst 1985, and Archangeli and Pulleyblank 1989 for representative examples and analyses). Similarly, and more dramatically, many languages allow long-distance assimilations in which one consonant affects another across any number of intervening consonants at other places of articulation; languages that have been studied include Chumash (Poser 1982), Sanskrit (Schein and Steriade 1986), and Tahtlan (Shaw 1991). Dissimilatory rules, too, often operate at a distance (Itô and Mester 1986; McConvell 1988). Nevertheless, if we set aside the special case of languages with nonconcatenative morphologies (McCarthy 1981, 1985, 1989a), we find that there are important limits on how far a rule can "reach" across intervening material to affect a distant segment. In particular, it appears that assimilation rules cannot reach across "opaque" segments - segments that are already characterized by the spreading node or feature (Clements 1980; Clements and Sezer 1982; Steriade 1987a).

These limits follow, at least in part, from structural properties of the representations themselves. Of particular importance is the prohibition on crossed association lines (Goldsmith 1976), which we state in its most general form as follows:

> (27) No-Crossing Constraint (NCC)
> Association lines linking two elements on tier *j* to two elements on tier *k* may not cross.

This constraint applies as shown below, allowing representations like (28a), but ruling out those like (28b):

(28)

The NCC applies not only to underlying, but also to derived representations, where it serves as an absolute constraint blocking any rule application which would produce a violation of it. Consequently, it will prevent an assimilation rule from spreading a feature [F] across a segment already specified for [F], accounting for opacity effects of the sort described above. (An example will be given in the next section.)

3 Toward a Formal Model of Feature Organization

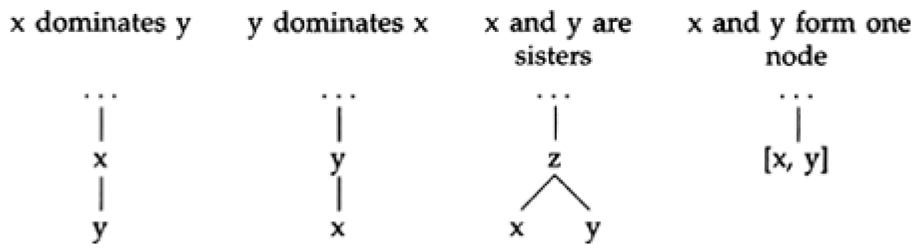
We now consider the model of feature organization in more detail. We assume a metatheoretical principle that features have minimal hierarchical organization in the absence of evidence to the contrary. We next consider what evidence to the contrary might consist of.

3.1 Evidence for Feature Organization

We have already examined several types of evidence for feature organization. The most important of these is the operation of phonological rules. Thus, if a phonological rule can be shown to perform an operation (spreading, delinking, etc.) on a given set of features to the exclusion of others, we assume that the set forms a constituent in the feature hierarchy.

Two features x and y can be grouped into constituents in four ways, as shown below:

(29)



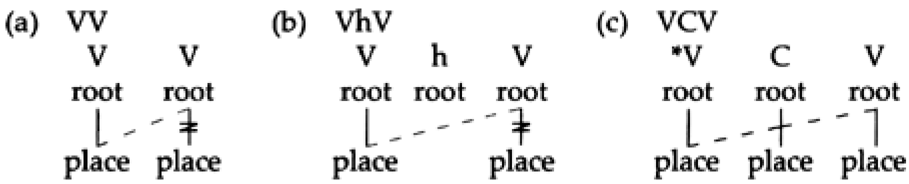
If an operation on x always affects y, but not vice versa, the first configuration is motivated. If an operation on y always affects x, but not vice versa, the second is required. If x and y can be affected independently of each other, they are each independently linked to a higher node z, as shown in the third figure. Finally, if an operation on one always affects the other, they form a single node, as shown at the right.

Another criterion for feature organization is the presence of OCP-driven co-occurrence restrictions. As the previous discussion has shown, any feature or set of features targeted by such constraints must form an independent node in the representation.

A further criterion, but one which must be used with caution, is node implication. If a node x is always linked under y in the universal feature organization, the presence of (nonfloating) x implies the presence of y. For example, since [anterior] is universally linked under the [coronal] node, we predict that all [±anterior] segments are coronal. Note, however, that not all implicational relations among nodes can be expressed in this way. For example, although all [+consonantal] segments must have a place node, the place node must be lower in the hierarchy, since place can spread independently of [consonantal] (Halle 1989; and see section 3.3.2 below). Similarly, for reasons that have nothing to do with node organization as presently conceived, all [-consonantal] segments are [-lateral] and [+sonorant], all [-continuant] obstruents are [-nasal], and all [+low, -back] vowels are [-rounded]. Thus while it is possible to capture some implicational relations directly in terms of the dependency relation, others must apparently be expressed in terms of explicit wellformedness conditions.

One further criterion for feature organization consists of transparency and opacity effects, as discussed in section 2.5. To see how these effects bear upon feature organization, let us consider the phenomenon of *laryngeal transparency* as discussed by Steriade (1987b). In a number of languages, including Acoma, Nez Perce, Arbore, and Yokuts, vowels assimilate in all features to adjacent vowels, but not to nonadjacent vowels. Exceptionally, laryngeal glides [h, ?] are transparent to this assimilation; thus in Arbore, assimilation crosses the laryngeal in examples like /(ma) beh-o/ "he is not going out" → [...boho]. This behavior can be explained on the assumption that laryngeal glides, unlike true consonants and vowels, have no distinctive oral tract features. In this view, [h] is only characterized by the laryngeal feature [+spread glottis] (Clements 1985), acquiring its oral tract phonetic characteristics from its phonetic context (Keating 1988). The transparency of laryngeal glides and the opacity of true consonants follows from the structure of their respective representations, as shown below (irrelevant details omitted):

(30)



In (30a), the place node of the first vowel spreads to the root node of the following vowel, triggering the delinking of its original place node (indicated by the z). Spreading may also take place in (30b), since the intervening [h] has no place node to block the spreading. In (30c), however, vowel-to-vowel spreading cannot take place without introducing crossed lines, in violation of the No-Crossing Constraint (27). The transparency of [h] is therfore fully predictable from the fact that it is not characterized by place features.[18]

With this background, let us consider feature organization in more detail.

3.2 The Root Node

The root node, dominating all features, expresses the coherence of the "melodic" segment as a phonological unit. There is considerable evidence in favor of a root node, which we touched on briefly in section 2. We have seen, for example, that processes of total assimilation in languages such as Ancient Greek can be expressed as the spreading of the root node from one skeletal position to another. Without the root node, such processes would have to be expressed as the spreading of several lower-level nodes at once, contrary to principle (4).

We have also seen that the different phonological behavior of short segments, contour segments, and geminate segments can be insightfully accounted for in terms of different patterns of linkage between root nodes and skeletal positions. Other evidence for the root node can be drawn from segment-level metathesis, segmental deletion, rules mapping segments to morphological template positions, and OCP effects on the root node (as just discussed in Arabic), as well as from the fact that single segments commonly constitute entire morphological formatives in their own right, while subparts of segments rarely do. All of these phenomena would be difficult to express without the root node.

Schein and Steriade (1986) and McCarthy (1988) propose to assign a special status to the root node by allowing it to bear the major class features, which we take to be [sonorant], [approximant], and [vocoid] (the terminological converse of [consonantal]). The unity of these features derives from their role in defining the major sonority classes, obstruent, nasal, liquid, and vocoid. Given these features, sonority rank is a simple function of positive feature values (Clements 1990a).

(31)

	[sonorant]	[approximant]	[vocoid]	sonority rank
obstruent	−	−	−	0
nasal	+	−	−	1
liquid	+	+	−	2
vocoid	+	+	+	3

The assignment of the sonority features directly to the root node predicts that they can never sprread or delink as a class independently of the root node as a whole. This prediction seems largely correct, though see Kaisse (1992) for proposed cases of [consonantal] ([vocoid]) spreading. Assuming it is generally true, we have the following representation of the root node:

(32)

$$\begin{bmatrix} \text{sonorant} \\ \text{approximant} \\ \text{vocoid} \end{bmatrix}$$

Christdas (1988) and Clements (1990a) propose that sonority features are present and fully specified in underlying representation, at least to the extent necessary to "drive" the process of core syllabification and account for root structure constraints.[19]

Piggott (1987) proposes that [nasal] attaches under the root node on a tier of its own. In Sagey's model (1986), [nasal] links to the root through an intervening soft palate node, representing its articulator.

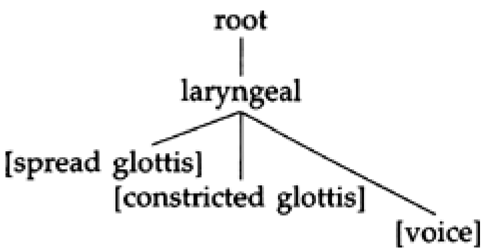
3.3 The Feature Organization of Consonants

3.3.1 The Laryngeal Node

Primary motivation for a laryngeal node comes from the fact that laryngeal features may spread and delink not only individually, but as a unit. For example, in Proto-Indo-Iranian, voicing and aspiration spread bidirectionally as a unit from voiced aspirates to adjacent obstruents (Schindler 1976). In the Shapsug dialect of West Circassian, the distinction among voiceless aspirated, plain voiced, and glottalic (ejective) stops and fricatives is lost in preconsonantal position, the surface phonation of the cluster as a whole being determined by its final member (Smeets 1984). Similarly, in Korean, as noted in section 2.3, the three-way lexical contrast among plain, aspirated, and "tense" (glottalized) obstruents is neutralized to a plain unreleased type in syllable coda position.

To express these facts, we assign the laryngeal features to separate tiers and group them under a laryngeal node, which links in turn to the root node:

(33)

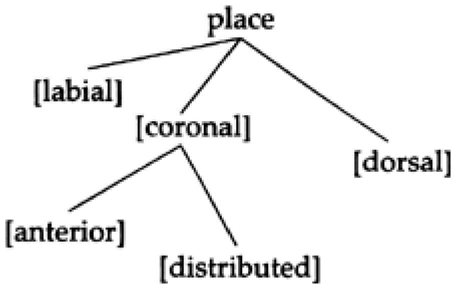


It may be preferable to characterize voicing by the features [stiff vocal cords] and [slack vocal cords] (Halle and Stevens 1971). Bao (1990) suggests that [stiff] and [slack] may form one constituent under the laryngeal node, and [spread] and [constricted] another.

3.3.2 The Place Node

In rules of place assimilation, the oral tract place features [labial], [coronal], and [dorsal] and their dependents spread as a single unit, independently of stricture features such as [continuant], [vocoid], and [sonorant]. We may capture this fact by grouping them under a single place node, as illustrated in (34).

(34)



Nasals commonly assimilate to the place of articulation of following stops. Although it is rarer to find nasal assimilation before fricatives and approximants, a number of cases of this type have been reported. In Chukchi, underlying /ŋ/ assimilates to the place of articulation of following consonants, including fricatives, liquids, and glides (Bogoras 1922; Krause 1980; Odden 1987; v is a bilabial continuant):[20]

(35)

təŋ-əɬʔ-ən	"good" /tEŋ -/
tam-pera-k	"to look good"
tam-vairgin	"good being"
tam-waɣərɣ-ən	"good life"
tan-t^{s}ai	"good tea"
ten-leut	"good head"
tan-ran	"good house"
ten-yəɬqət-ək	"to sleep well"

In each case, the nasal assimilates to the place, but not the stricture of the following consonant. Similar examples of place assimilating independently from stricture features can be cited from Yoruba (Ward 1952), Catalan (Kiparsky 1985), and the Yongding dialect of Chinese (Dell 1993), among many others.

Besides spreading, the place node can be delinked, accounting for debuccalization processes such as t > ? and s > h (McCarthy 1988). Note that debuccalized sounds are always realized as [−consonantal] ([+vocoid]) glides. This fact follows directly from the standard definition of [+consonantal] ([−vocoid]) segments as sounds produced with a radical obstruction in the midsaggital region of the vocal tract (Chomsky and Halle 1968). Sounds without oral place features can have no such obstruction, and so are necessarily nonconsonantal.

It will be noted that on this analysis, if the place node of a nasal is delinked, the feature [+nasal], which links to the root node, should remain behind. This prediction receives some support from patterns of sound change in Malay dialects observed by Trigo (1988, 1991). The evolution of final stops, fricatives, and nasals in two dialects is summarized below:

(36) (a) p, t, k > ?
(b) s, f, h > h
(c) m, n, ŋ > N (a placeless nasal glide)

The first two sets of changes (a, b) represent standard examples of debuccalization. We might suppose that the glottal features of [?] and [h] were present redundantly in stops and fricatives, respectively, at the point when delinking took place, accounting for their presence in the debuccalized forms. Crucially, set (c) shows that when a nasal is debuccalized, [+nasal] is left behind. This gives direct evidence that the place node is delinked, since if the root node were delinked instead, [+nasal] should have been delinked with it. The simplest account of these phenomena (though not necessarily the historically correct one) is that the place node was deleted in all cases. Laryngeal features and [+nasal] are not affected, and the resulting segment is shifted to a [+vocoid] glide due to its lack of a place node.

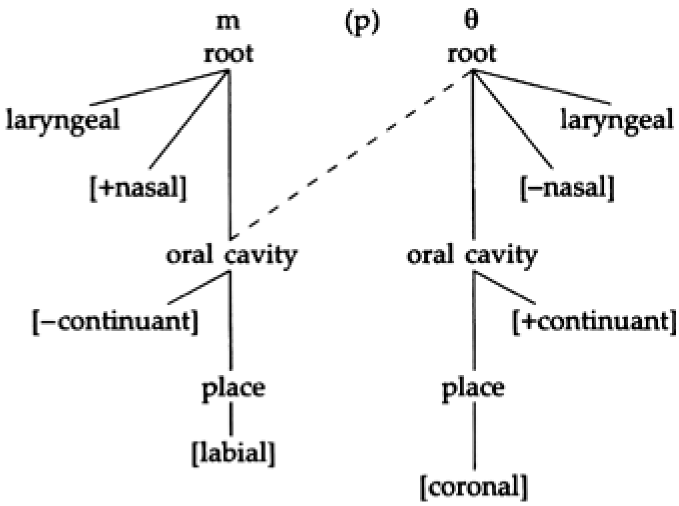
3.3.3 The Oral Cavity Node

In some presentations of feature geometry, the place node links directly to the root node. However, recent work has brought to light evidence in favor of an *oral cavity* node intervening between the place node and the root node, dominating place and [±continuant] nodes. This constituent corresponds to the articulatory notion “oral cavity constriction,” and characterizes it as a functional unit in the phonology.

We illustrate this node with the process of intrusive stop formation (ISF) found in many varieties of English (Clements 1987). By this process, words like *dense* and (for some speakers) *false* acquire a brief, intrusive [t] at the point of transition from the nasal or lateral to the following fricative, making them sound similar to *dents* and *faults*. The intrusive element always has the same place of articulation as the consonant on its left, as we see in further examples like *warmth* [...m^{p}θ] and *length* [...ŋ^{k}θ]. Phonetic studies show that the intrusive stop is shorter by a small but significant margin than the underlying stop found in words like *dents* and *faults* (Fourakis and Port 1986).

Traditional accounts have sometimes viewed ISF as involving an anticipation of the orality of the fricative on the preceding nasal; however, such accounts do not explain why this process may also apply after laterals. A unified account of ISF is possible if we view it as involving a lag of the oral cavity constriction of the nasal or lateral into the following fricative. In pronouncing a word like *warmth*, for example, speakers prolong the labial occlusion of the [m] into the [θ], producing a “hybrid” segment having the labial closure of the [m] but all other features of the [θ], in other words, a [p]. This process can be formalized as a rule spreading the oral cavity node rightward onto the root node of the fricative. We illustrate its effect below:

(37)



In the derived representation, [θ] bears two oral cavity nodes in succession, the labial stop node of the [m] followed by the coronal continuant node of the [θ]. Thus the two oral cavity nodes form a “contour” across the [θ], in just the same way that two tones may form a contour across a single vowel in the case of rising and falling tones (see (2b) above). In this analysis, the intrusive stop [p] is not a full segment in its own right, but results from the partial overlap of the oral cavity node of [m] with the other features of [θ]. The fact that the intrusive stop is significantly shorter in duration than an underlying stop can be explained in terms of the fact that it constitutes part of a contour segment.[21]

It will be noted that the derived segment [^{p}θ] has the internal structure of an affricate under the root node, in that it consists of a stop [p] followed by a fricative [θ]. Recall, however, that the No Branching

Constraint (11) prohibits such branching structure under the root. In line with our earlier discussion, we assume that (37) is automatically converted into a structure with two root nodes by the appropriate repair convention.[22]

For alternative proposals which bundle [continuant] into a single constituent with the place features, equally consistent with the analysis of intrusive stop formation proposed here (though differing in other empirical predictions), see Selkirk 1990 and Padgett 1991.

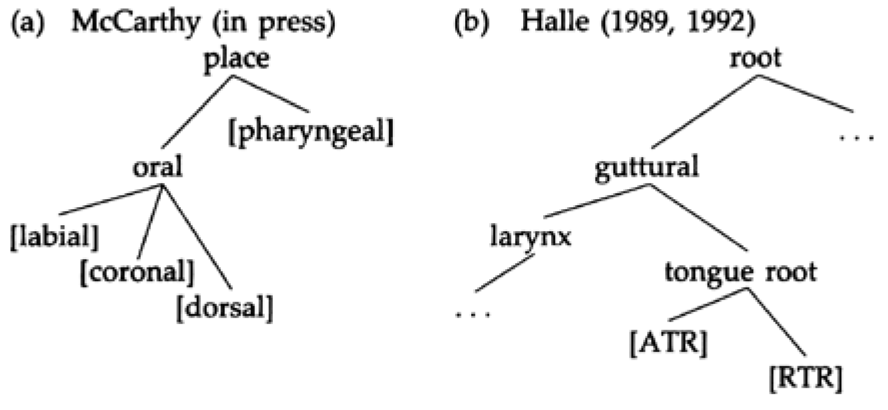
3.3.4 The Pharyngeal (or Guttural) Node

In many languages, we find that glottal, pharyngeal, and uvular sounds define a natural class, often referred to as "gutturals." For example, in Classical Arabic many rules and constraints are defined on the [+approximant] subclass of these sounds consisting of the laryngeals [h ʔ], the pharyngeals [ħ ʕ], and the uvular continuants [χ ʁ] (Hayward and Hayward 1989; McCarthy 1989b, in press); thus with very few exceptions, no roots may contain two sounds of this group. The class of "guttural" sounds can be characterized by the feature [guttural] (Hayward and Hayward) or [pharyngeal] (McCarthy).

While this feature is now established beyond reasonable doubt, its exact status and relation to other features is still uncertain. McCarthy points out that it cannot be an articulator feature on a par with [labial], [coronal], etc., since it cannot be defined in terms of the movement of any single articulator. Rather, what the sounds of this class have in common is that they are articulated in a continuous region of the vocal tract, extending approximately from the upper pharynx to the larynx, inclusively.

There are currently two main theories of how the feature [guttural] (or [pharyngeal]) is to be integrated into the feature hierarchy. McCarthy suggests that [pharyngeal] links under the place constituent together with the oral tract place features [labial], [coronal], and [dorsal]. A potential problem for such an analysis is the phenomenon of "guttural transparency," according to which guttural sounds, and no others, may be transparent to rules spreading vowel place features. In Tigre, for example, the underlying / ə / of the prefix /tə-/ assimilates to the following [a] across the guttural [ʔ] in words like ta-ʕārafa "he visited", but does not assimilate across the uvular [q] in words like tə-qābala "he met", since [q], though a guttural sound, is [dorsal] as well as [pharyngeal]. To accommodate such facts, McCarthy proposes to group the oral place features into a single "oral" constituent which forms a sister to [pharyngeal]. This conception is illustrated in (38a), after McCarthy (in press).

(38)



The Tigre rule may be expressed as the spreading of the oral node from the root vowel to the prefix vowel; spreading will be blocked by the No-Crossing Constraint (27) just in case the intervening consonant bears any oral feature such as [dorsal].

An alternative view, illustrated in (38b), is proposed by Halle (1989, 1992). Halle argues that since the guttural sounds are not defined by the activity of any single articulator, they should not be assigned an articulator node on a par with [labial], [coronal], and [dorsal]. Instead, he proposes to group the gutturals under a higher-level "guttural" node, which groups the laryngeal articulator ("larynx") and its dependent features on the one hand, and the tongue root articulator and its dependent features [ATR] (Advanced Tongue Root) and [RTR] (Retracted Tongue Root) on the other. In this model, the Tigre assimilation rule can be expressed as ordinary spreading of the place node, and the oral node is not needed.[23]

While both of these conceptions are consistent with the guttural transparency phenomenon, they make substantially different predictions in other respects. Given that laryngeal features are sufficient to characterize the laryngeals [h ?], McCarthy's model does not straightforwardly predict that these segments pattern with the gutturals, unless, following McCarthy, we allow them to bear the redundant specification [pharyngeal]. Halle's model predicts that obstruents with distinctive laryngeal features such as [+voiced] or [+spread glottis] can potentially pattern with the gutturals by virtue of their guttural node. Perhaps the central difference, however, regards their claims concerning possible spreading and delinking rules. McCarthy's model predicts that we should find rules spreading or delinking [pharyngeal] together with the oral tract place features, while Halle's predicts rules that spread or delink laryngeal and tongue root features as a unit. To date, no fully conclusive evidence has been brought to bear on these predictions.

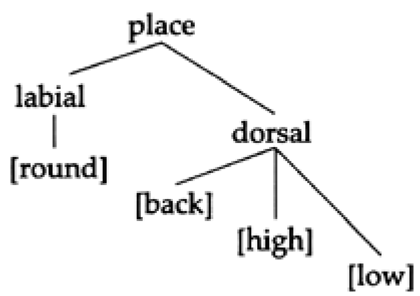
3.4 The Feature Organization of Vocoids

We now consider the feature organization of vocoids, that is, vowels and glides. A long-standing issue in phonological theory has been the extent to which consonants and vocoids are classified by the same set of features. While most linguists agree that they share such features as [sonorant], [nasal], and [voiced], at least at the level at which nondistinctive feature values are specified, there has been much less agreement regarding the extent to which features of place of articulation and stricture are shared. The articulator-based framework of feature representation, as described above, has made it possible to offer a more integrated approach to this problem. In this section we first outline two approaches inspired by this general framework, and then consider some of the differences between them.

3.4.1 An Articulator-based Model

In the earlier of these approaches, Sagey (1986) retains the SPE features [high], [low], [back], and [round]. She integrates them within the articulator-based framework by treating them as articulator-bound features, linked under the appropriate articulator node. Thus [back], [high], and [low], as features executed by the tongue body, are linked under the dorsal node, and [round], as a feature executed by the lips, is assigned to the labial node, as shown below:

(39)



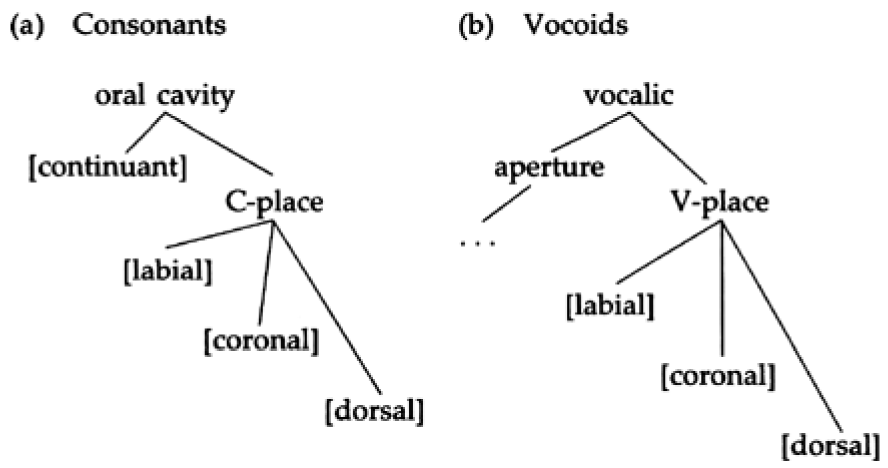
In this model, all consonants and vocoids formed in the oral tract are characterized in terms of an appropriate selection from the set of articulator nodes and their dependents, although coronal, reserved for retroflex vowels, is usually nondistinctive in vocoids. One of the central predictions of this model is that the set [back], [high], and [low], as features of the dorsal node, has a privileged status among subsets of vowel features, in that it alone can function as a single phonological unit.

3.4.2 A Constriction-based Model

A second approach, emanating from work by Clements (1989a, 1991, 1993), Herzallah (1990), and Hume (1992), proposes to unify the description of consonants and vocoids in a somewhat different way. This model is based on the preliminary observation that any segment produced in the oral tract has a characteristic *constriction*, defined by two principal parameters, constriction degree and constriction location. Since vocal tract constrictions determine the shape of the acoustic signal and thus constribute directly to the way in which speech is preceived, they can be regarded as constituting the effective goal of articulatory activity.

Given their centrality in speech communication, it would not be surprising to find that constrictions play a direct role in phonological representation itself. This is the view adopted by the model under discussion, which proposes to represent constrictions by a separate node of their own in the feature hierarchy. The parameters of constriction degree and location are also represented as separate nodes, which link under the constriction node. This type of organization was already proposed for consonants above, in which the constriction itself is represented by the oral cavity node, constriction degree by the [±continuant] node, and constriction location by the place node; this conception is summarized in (40a). A parallel structure can be assigned to vocoids, as shown in (40b). In this figure, the constriction of a vocoid is represented by its vocalic node, its constriction degree by an aperture node, and its constriction location by a place node. As in the case of consonantal constrictions, these nodes have no intrinsic content, and receive their intepetation by virtue of the feature values they dominate. In these figures, place nodes of consonants and vocoids, which occur on different tiers, are designated as "C-place" and "V-place," respectively.

(40)



The aperture node dominates vowel height features, represented by the ellipsis, which are discussed further in section 3.4.5 below.

A further innovation of this model is that the features [labial], [coronal], and [dorsal], occuring under the V-place node in vocoids, are sufficient, by themselves, to distinguish place of articulation in vowels, and replace the traditional features [back] and [round]. In order to fulfill this new and expanded role in the theory, they must be redefined in terms of constrictions rather than articulator movements as such. This can be done as follows, to a first approximation (compare the definitions given earlier in (7)):[24]

(41) Liabial: incolving a constriction formed by the lower lip
Coronal: involving a constriction formed by the front of the tongue
Dorsal: involving a constriction formed by the back of the tongue (= the dorsum, cf. Ladefoged 1982, p. 281)

These statements, valid for consonants and vocoids alike, define constriction location in terms of the active articulator involved. Since all segments with oral tract constrictions are formed by the lips or the tongue body, all are characterized by at least one of these three features. As far as vocoids are concerned, rounded vocoids are [labial] by these definitions, front vocoids are [coronal], and back vocoids are [dorsal]. Central vocoids satisfy none of the definitions in (41), and are thus treated as phonologically placeless. These features appear sufficient to characterize all phonologically relevant properties of constriction location in vocoids, and make the features [back] and [round] superfluous (Clements 1989b, 1991b, 1993).

A constriction-based model incorporating the definitions in (41) makes a number of different predictions from Sagey's model regarding the phonological behavior of vocoids. First, the constriction-based model predicts that front vowels can form a natural class with coronal consonants, and back vowels with dorsal consonants, while Sagey's model predicts that all vowels form a natural class with dorsal consonants and no others. Second, the constriction-based model predicts that the aperture features, the V-place features, or the aperture and V-place features together can function as single units in phonological rules, while Sagey's model predicts that only the dorsal features [high, back, low] can do so. Third, the constriction-based model predicts that dorsal consonants (or at least "plain" dorsals with no secondary articulation, see below) will be transparent to rules spreading any two or more vowel features, while Sagey's model predicts that dorsal consonants are opaque to such rules, which must spread the dorsal node. Fourth, the constriction-based model predicts that not only dorsals but *all* ("plain") consonants will be transparent to rules spreading lip rounding together with one or more vowel features, while Sagey's model predicts that all intervening (supralaryngeal) consonants will be opaque to such rules, which must spread the place node. We examine these predictions in turn.

3.4.3 Natural Classes of Consonants and Vowels

The constriction-based model predicts that we should find a natural class corresponding to each of the oral tract place features, as shown below:

(42) [labial]: labial Consonants; rounded or labialized vocoids
[coronal]: coronal consonants; front vocoids
[dorsal]: dorsal consonants; back vocoids

Each of these classes is, in fact, well documented in the literature. Of these, the first is the least controversial, since it has been recognized and discussed since the early studies of Reighard (1972) and Campbell (1974); see Selkirk (1988), Capo (1989), and Clements (1990b, 1993) for further examples. As both of the models under consideration account for this class, we will consider here only the other two.

The interaction of coronal consonants and front vowels is covered by the surveys in Clements 1976, 1990b, 1993; Pulleyblank 1989; Hume 1992; Blust 1992; and references therein. For example, in many laguages, velar and/or labial consonants become coronal, and anterior coronals become posterior, before front vowels. This process, sometimes termed palatalization, may be better termed *coronalization* since the resulting sound, though coronal, is not necessarily either palatal or palatalized (Mester and Itô 1989, who attribute the term to Morris Halle and Alan Prince). While the appearance of coronal consonants in the context of front vowels has sometimes been treated in terms of automatic linking conventions or similarly arbitrary mechanisms, it can be viewed as a straightforward case of assimilation if front vowels are treated as [coronal] themselves; see the references above as well as Broselow and Niyondagara 1989; Mester and Itô 1989; and Lahiri and Evers 1991 for further discussion. We take a closer look at coronalization (and palatalization) rules in section 4.

In parallel fashion, vowels are fronted next to coronal consonants in a number of languages. The triggering consonant may be, but is not necessarily palatal. In Maltese Arabic, for example, the vowel of the imperfective prefix is always predictable. It is generally realized as a copy of the stem vowel, as shown in the second column of (43a) (the perfective stem is given to the left for comparison). However, when the following consonant is a coronal obstruent, the prefix vowel is systematically realized as the high front vowel [i], as shown in (43b).

(43) perfective imperfective

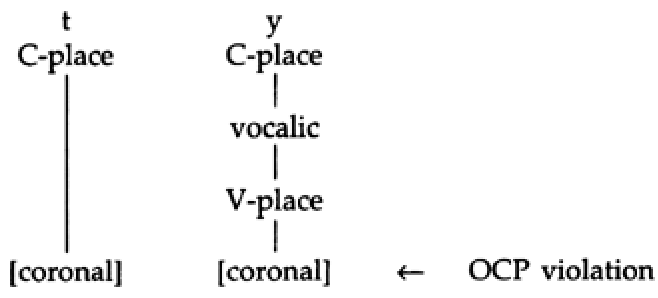
(a)	kotor	yo-ktor	"abound"	
	rifed	yi-rfed	"support"	/rifid/
	ʔasam	ya-ʔsam	"break"	
	ħebel	ye-ħbel	"rave"	
(b)	ʃorob	yi-ʃrob	"drink"	
	dalam	yi-dlam	"grow dark"	
	žabar	yi-žbor	"collect"	
	seħet	yi-sħet	"curse"	

Hume (1992) points out that the pattern in (43b) cannot plausibly be attributed to default rules, and argues that it results from the spreading of [coronal] from the consonant to the vowel.

Co-occurrence constraints also reveal the special relation of front vowels and coronal consonants. In Cantonese, for example, among other patterns, if the onset and coda of a given syllable are both coronal, any non-low vowel must be one of the front vowels [i e ü ö] (Cheng 1989). Thus, while the words [tit] "iron", [tüt] "to take off", and [tön] "a shield" are well-formed, words like *[tut], *[tsot], *[sut] are excluded. Here, then, a vowel flanked by two [coronal] consonants assimilates their coronality.

Korean has a particularly interesting dissimilatory constraint, with further implications for feature organization. In underlying representations, coronal obstruents do not occur with front glides in word-initial syllables, nor do front glides occur with front vowels; thus, syllables containing such sequences as *ty, *sy, *cy, *yi are systematically excluded (Clements 1990b, 1993, after Martin 1951). This pattern can be understood as an OCP-driven constraint against occurrences of [coronal] in successive segments, meeting the conditions just stated. This analysis is supported by parallel OCP-driven constraints involving labiality (*pw, *mw, *wu, etc.). In Korean, then, it appears that the OCP applies "cross-categorially" to rule out sequences of consonants and vocoids having identical occurrences of the features [coronal] and [labial]. Note, however, that the OCP as stated in (23) only applies to nodes which are adjacent, and hence located on the same tier. To extend the OCP to Korean (and similar cases in other languages), Hume (1992) proposes that each articulator feature of a given category should be assigned to the same tier whether it characterizes a consonant or a vocoid.[25] This proposal assigns the following structure to sequences like /ty/.

(44)



Since both instances of [coronal] lie on the same tier, they trigger the OCP as stated in (23).

There is also considerable evidence that back (but not front) vocoids and dorsal consonants form a natural class, defined by [dorsal]. For example, in the Khoisan languages of southern Africa, only back vowels may occur after velar and uvular consonants, including clicks (Traill 1985). Assuming that all clicks have a [dorsal] component (Sagey 1986; Bradlow 1992), we may view this as a syllable structure constraint spreading [dorsal] from the consonant to the vowel. We find a dissimilatory process at work in the historical development of French, where velar and labial consonants were deleted in intervocalic position when flanked on either side by one of the rounded (i.e., labiovelar) vowels [u o]; examples of velar deletion include Lat. *fagu* > *fau* (MFr. *fou*) “crazy”, Lat. *ruga* > *rue* “street” (Clements 1990b, 1993, after Bourciez and Bourciez 1967). We may regard this deletion process as OCP-driven on the assumption that velars and back rounded vowels share the feature [dorsal]; were front vowels dorsal they should have triggered the deletion, too. Further examples of back vowel/velar consonant interaction are discussed in Clements (1990b, 1993), Herzallah (1990), Blust (1992), and Dell (1993).[26]

In sum, phonological rules offer considerable evidence for the natural classes of labial, coronal, and dorsal consonants and vocoids as stated in (42). This result supports a unified account of place in consonants and vowels, in which [labial], [coronal], and [dorsal] do double duty for consonants and vocoids, allowing the standard features [back] and [round] to be eliminated. (For other recent proposals to unify the feature characterization of consonants and vowels within comparable frameworks, see Pulleyblank 1989 and Gorecka 1989.) Let us now examine the internal structure of vocoids in more detail.

3.4.4 The V-place Node

In the constriction-based model, as we have seen, vocalic constrictions are defined in terms of the parameters of location (place) and degree (aperture). As Odden (1991) particularly has pointed out (though from a somewhat different perspective), there is considerable phonological evidence for a division of vowel features into these general categories.

As far as place is concerned, Odden offers evidence from several languages that the features of backness and roundness, i.e., our [labial], [coronal], and [dorsal], function as a single unit. For example, Eastern Cheremis has the vowel set /i ü u e ö o a/, in which both backness and rounding are distinctive, as well as neutral vowel /ə/. Word-final /e/ assimilates in backness and roundness, but not height, to the first preceding non-neutral vowel if it is labial; thus /e/ surfaces as [o] after [u, o] and [ö] after [ü, ö]. Examples are given in (45).

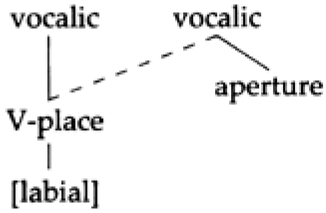
(45)

kit-še	**“his hand”**
ergə-že	**“his boy”**
šužar-že	**“his sister”**
surt-šo	**“his house”**
üp-šö	**“his hair”**
boz-šo	**“his wagon”**
šör-žö	**“its milk”**

There is good reason not to analyze vowel assimilation as two separate rules, one spreading backness and another roundness, since, as Odden points out, both rules would apply under exactly the same conditions and have exactly the same set of exceptions. The pattening of backness and roundness together to the exclusion of vowel height argues that these features form a single constituent, which we take to be the vocalic place (or V-place) node (Odden's node labels are somewhat different).

In the spirit of Odden's analysis, we assume that the rule of vowel assimilation spreads the V-place node, of a [labial] vowel rightward onto a final mid vowel unspecified for a place node, i.e., in the unmarked, feature-filling mode, as shown in (46). If the spreading place node has a dependent [coronal] feature as in the case of [ü ö], this feature will spread as well. Since the aperture node is not linked under V-place, it is not affected. (Note that intervening consonants are not specified for vocalic and V-place nodes, and so will not block the rule.)

(46)



Other examples are cited from a variety of languages (see Odden for further discussion and references). As Odden remarks, this type of assimilation cannot be accounted for within Sagey's model, in which the spreading features do not form a single constituent of their own.

3.4.5 The Aperture Node

Consider next constriction degree. Hyman (1988), Clements (1989b, 1991) and Odden (1991) have presented evidence from several languages that vowel height features may spread as a single unit, supporting the aperture (or vowel height) node proposed above. Here we consider a further illustration from Brazilian Portuguese, discussed by Quicoli (1990) and Wetzels (1993).

Brazilian Portuguese vowels form a four-height system, / i u e o ɛ ɔ a/. Underlying mid stem vowels undergo an interesting pattern of alternation in stressed prevocalic position, as is shown by a comparison between 2nd and 1st person forms of the present indicative. (The structure of these examples is: stem + theme vowel + person/number ending.)

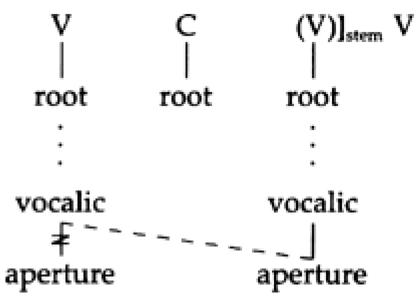
(47)

2nd person:			1st person:		
mɔr-a-s	[mɔ́ras]	"you reside"	mɔr-a-o	[mɔ́ro]	"I reside"
mɔv-e-s	[mɔ́ves]	"you move"	mɔv-e-o	[móvo]	"I move"
sɛrv-i-s	[sɛ́rves]	"you serve"	sɛrv-i-o	[sírvo]	"I serve"

In the 1st person forms, the mid stem vowels assimilate to the height of the following non-low "theme" vowel, becoming upper mid before [-e] and high before [-i]; the theme vowel is concomitantly deleted. No assimilation takes place in the 2nd person forms. This pattern is regular across the verb conjugation.

In his analysis of these forms, Wetzels proposes that theme vowels are deleted in hiatus before another vowel. However, by a "stability" effect similar to that found in many tone languages, their aperture node relinks to the stem vowel, replacing its original node. This analysis is illustrated below, where "(V)" represents the skeletal position of the deleted theme vowel. (As the stem consonant is not specified for vocalic and aperture nodes, it does not block the rule.)

(48)



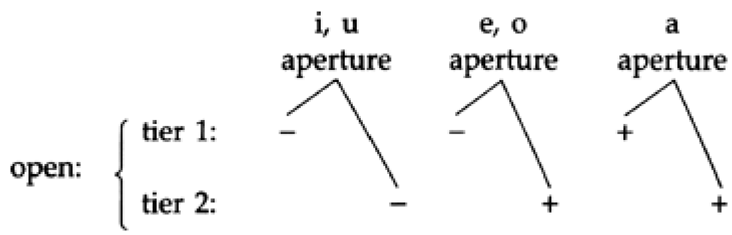
Unlinked nodes under the deleted vowel slot are subsequently deleted. As Wetzels points out, this rule (and many similar rules of height assimilation discussed in the literature) cannot be expressed as a unitary process if the aperture features are not grouped into a single unit, in the general framework assumed here.

We have not so far discussed the vowel height features as such. Traditionally, generative phonologists have used the binary features [high]and [low] to distinguish among high, mid, and low vowels, and have added a further feature such as [tense] or [ATR] to express a fourth height if necessary. These features are assumed in the geometries proposed by Sagey (1986), Hyman (1988), and Odden (1991), among others, and continue to represent the main trend in the field. However, vowel height has received a good deal of attention in recent years, and several alternative systems have been proposed. We discuss two here, both of which model vowel height in terms of aperture rather than tongue body height, consistent with the general assumptions of a constriction-based framework.

In one, vowel height (together with other vowel features) is treated as a privative feature called a particle or component, usually represented *a*, interpreted as vocal tract aperture (Schane 1984; Rennison 1986; Anderson and Ewen 1987). If it stands alone, this feature designates the low vowel [a], and if it is combined with other features, it designates vowels with some degree of openness (for example, when combined with the palatal component *i* it designates a relatively open palatal vowel such as [e] or [ɛ]). This model directly expresses the fact that when [a] coalesces with [i], the result is usually a non-high vowel such as [e]; this follows from the fact that [e] is just the combination of the particles *a* and *i*. A problem, however, results from its failure to provide a feature or particle corresponding to [+high] or [-low]: it is unable to express assimilatory vowel raising in terms of autosegmental spreading. Yet assimilatory raising is common across languages, and exhibits characteristics quite parallel to assimilatory lowering (see, e.g., Clements 1991; Kaze 1991).

A second alternative to the standard system, proposed by Clements (1989b, 1991), proposes a single feature [±open]. Unlike the particle *a*, [open] is a binary feature, either value of which may spread. To express various degrees of vowel height, the feature [open] is arrayed on several rank-ordered tiers. On the highest-ranked tier, [open] assigns vowels to one of two primary height registers, [-open] (relatively high) and [+open] (relatively low). Any height register can be subdivided by further assignments of [open] on the next lower-ranked tier. For example, the familiar three-height system /i u e o a/ can be represented as shown below (redundant feature values included), where the higher of the two primary registers, designated by the [-open] specifications on tier 1, is subdivided into higher and lower secondary registers on tier 2:

(49)



Natural classes are defined in terms of feature values on each tier. Thus low vowels are those which are [+open] on tier 1, high vowels are [-open] on tier 2, and so forth. In this system, assimilatory raising is stated as the spread of [-open] to [+open] on a designated tier. If no tier is specified, raising applies across all tiers, producing (if the rule is structure-preserving) the effect of stepwise or scalar raising.

One advantage of both of these approaches is that they allow us to eliminate the use of [ATR] as an ersatz vowel height feature, i.e., one motivated only by the need to describe a fourth height; this is because systems with four or more vowel heights can be analyzed in terms of additional *a*-particles or [open] tiers. Such analyses are strongly motivated in languages like Kimatuumbi in which "[ATR]" spreads with other features of vowel height (Odden 1991), since if [ATR] were really involved, one would expect it to spread with place features, not height features. [27].

3.4.6 The Vocalic Node

Let us now consider the status of the vocalic constriction node itself, which we called *vocalic* in (40).[28] By grouping all place and aperture features of vocoids under the vocalic node, we predict that all these features should be able to spread freely across intervening consonants, even if they are specified for place features of their own. This is because consonants (at least those with no secondary articulations; see below) have no vocalic node that would block them.

There is considerable evidence that this prediction is correct. An example can be cited from the Servigliano dialect of Italian, as described by Camilli (1929).[29] The vowel system of Servigliano is / i u e o ɛ ɔ a/, which is reduced to [i u e o a] in unstressed positions. The following examples illustrate a regular pattern of alternation involving post-tonic stem vowels (note that all final vowels in these examples are suffixes):

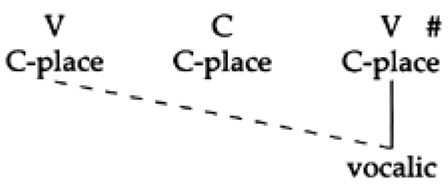
(50)

birikɔ́kan-a	"apricot tree"	birikókun-u	"apricot"
pétten-e	"comb"	péttin-i	"combs"
álam-a	"soul"	álem-e	"souls"
prɛ́dok-o	"I preach"	prédik-i	"you preach"
stómmuk-u	"stomach"	stómmik-i	"stomachs"

Strikingly, the final stem vowel is identical to the suffix vowel in all cases. Related forms such as predik-á "to preach", with stem vowel [i], and stomme-k-ósa "nauseating" (fem. sg.), with stem vowel [e], show that this vowel may have a different, unpredictable form in pretonic position, and must thus be specified for at least some features in underlying representation. It appears then, that we must postulate a total vowel assimilation rule which spreads the features of the suffix vowel to a post-tonic stem vowel. Crucial to the point at issue, all consonants are transparent, whatever their places and manners of articulation.

Let us consider how this rule can be expressed in terms of the feature hierarchy. Since the vocalic node plays a role similar (if not identical) to that of Sagey's dorsal node in the analysis of vocoids, we assume that it is linked to the same position, that is, under the place (i.e., C-place) node. The rule then applies as follows:

(51)



The Sageyian model cannot express the alternations in (50) straightforwardly, since the spreading of more than one vowel feature at a time can only be expressed as the spreading of the dorsal node, or a higher node (see(39)). This model predicts that velar consonants, which are [dorsal], should be opaque, but as the last examples show, this prediction is incorrect.[30]

Other ways of linking the vocalic node are equally consistent with the evidence from total vowel assimilation rules; for example, it could be linked directly to the root node.[31] However, the linkage given above is supported by further evidence, which we discuss in the following sections.

3.5 Major and Minor Articulations

A further source of evidence for feature organization comes from the study of so-called secondary articulations. Phoneticians define a secondary articulation as "an articulation with a lesser degree of closure occurring at the same time as another (primary) articulation" (Ladefoged 1982, p. 210), and usually reserve the term for inherent, as opposed to contextually-determined, articulations. The four most commonly-occurring types are *labialization*, typically realized as the addition of lip-rounding to the primary articulation; *palatalization*, typically involving the raising and fronting of the tongue body in the direction of the hard palate; *velarization*, tyically realized as tongue backing; and *pharyngealization*, involving the retraction of the tongue root. Following the arguments in Chomsky and Halle (1968), it is widely accepted that secondary articulations involve the same features as the articulatorily similar vowels; thus palatalization involves (some or all of) the features of [i], labialization the features of [u], and so forth.

The definition given above is not adequate as a phonological definition, since it is based on phonetic criteria. For this reason, Sagey proposes to redefine primary and secondary articulations in terms of a purely phonological distinction between major and minor articulations (1986, 1989). She observes that in most types of complex consonants, only one degree of closure is distinctive; the other is fully predictable, and its degree of closure need not be specified in the representation. The articulator whose stricture is predictable is termed the *minor* articulator, and the other the *major* articulator. For example, in languages with secondary labialization, the degree of labial stricture in a segment is always predictable from its other features, and so labialization constitutes the minor articulation. The stricture of the other, primary articulation may be distinctive, as in languages that contrast a labialized velar stop [k^w] and fricative [x^w], and this articulation is accordingly the major one.

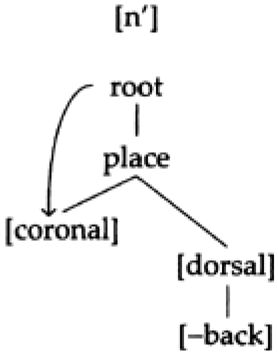
This definition is a purely phonological one, and does not, unlike the phonetic definition, entail that a minor articulation has a wider degree of closure than a major one. Indeed, this is not necessarily the case. In Ubykh, for example, the minor articulation of labialization is realized as lip rounding in velars (e.g., [k^w]) but as simultaneous closure in alveolars (e.g., [t^p]) (Comrie 1981). Anderson (1976) adduces evidence that the labiovelar stops [kp gb] found in many African languages consist of one primary and one secondary component, a distinction which can be reinterpreted in terms of major and minor articulations in Sagey's sense. Clicks, involving two simultaneous closures, can be analyzed into a major dorsal and minor coronal (or labial) articulation (Sagey 1986, 1989). However, when two simultaneous constrictions actually differ in degree of closure, the (phonological) major articulation always appears to coincide with the (phonetic) primary articulation, and the minor articulation with the secondary articulation.[32]

Sagey's proposal places the study of multiple articulations on a solid phonological footing, and has been widely accepted. Given the distinction between major and minor articulators, however, several fundamental questions emerge: How are major and minor articulators organized in feature representations? How is the major articulator distinguished from the minor articulator?

3.5.1 The Organization of Multiple Articulations

As before, we will consider two alternative models. In Sagey's model (1986), all oral articulator features, major and minor, link directly to the place node as sisters. Thus in complex segments, major and minor articulator features are not formally distinguished in terms of node organization as such. To distinguish them, a device called a "pointer" is introduced, which links the root node (and the stricture features it dominates) to the major articulator feature. This conception is illustrated below, where we give the representation of a palatalized coronal consonant such as [n′]. Notice that palatalization is characterized in terms of a [−back] dorsal node, just as is a front vowel.

(52)



This model predicts that if the place node spreads, both the major and minor articulator features should spread with it.

Data from Irish confirm this prediction. In Irish, according to Ní Chiosáin (1991), a nasal consonant optionally assimilates to the place of articulation of a following consonant. Just in case this assimilation takes place, the nasal adopts the secondary articulation of the following consonant, becoming palatalized before a palatalized consonant and plain before a plain consonant (see(53)). In other words, when the major articulator features spread, the minor articulator features spread too.

(53)

k′ɑːn g′ɑr	→	k′ɑːŋ′g′aːr	"a short one"
gan x′il	→	gaŋ′x′iːl	"with no sense"
kiːrɑːn b′eg	→	kiːrɑːm′ b′eg	"a small moor"
nə k′iːn′ + ɣuːwə	→	nə k′iːŋ ɣuːwə	"the black ones"

We may explain this pattern on the assumption that major and minor articulation features both link under the place node, consistently with (52).

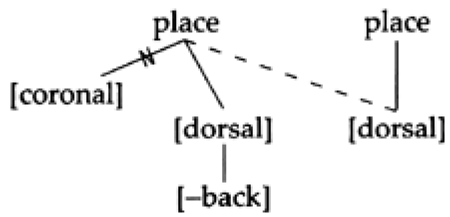
However, Ní Chiosáin notes further data that raise a problem for the Sageyian model. Before a velar consonant, the nasal may assimilate only the major dorsal articulation, as shown in (54), illustrating a palatalized nasal before a plain velar:

(54)

nə k′iːn′ + xorkrə → nə k′iːŋ′ xorkrə "the purple ones"

Indeed, this realization is the preferred one.[33] It motivates a further rule of dorsal assimilation, also optional, which spreads the dorsal node alone. However, this result is unexpected in the Sageyian model. If the palatalized coronal [n′] assimilates to the dorsality of a (nonpalatalized) velar sound, it will acquire a second dorsal node, while losing its coronality:

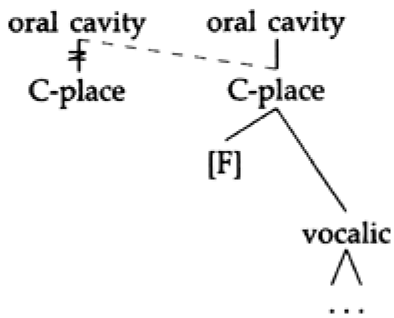
(55)



But since a single segment may not be doubly specified for dorsality, we expect the nasal's original dorsal node, representing palatalization, to be automatically delinked, to preserve well-formedness.[34]

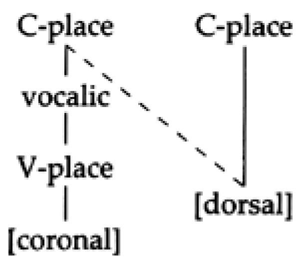
Let us now consider how complex segments can be represented in the constriction-based model. The commonest secondary articulation types – labialization, palatalization, velarization, and pharyngealization – can be very naturally characterized as minor articulations involving the features [labial], [coronal], [dorsal], [pharyngeal], respectively, supplemented by appropriate vowel height features as necessary (Clements 1990b, 1993; Herzallah 1990; Hume 1990, 1992). If we assume that these features (or at least the first three; see section 3.3.4 for a discussion of [pharyngeal]) are members of the vocalic constituent, linked under the C-place node, then the spreading of the latter in rules of place assimilation will automatically entail the spreading of minor articulations. We represent this analysis in (56) ([F] = any major articulator feature):

(56)



As Ní Chiosáin points out, this type of constituent structure is supported by the Irish data. First, it allows us to express the spreading of all C-place features as a unit, accounting for the data in (53); if the features of minor articulation were linked to a higher node, such as the oral cavity or root node, they would not be affected by place assimilation. [35] In addition, it allows the independent spreading of a single major articulator feature [F], directly accounting for Irish examples like (54). Note, in particular, that the spreading of the velar's [dorsal] node to the C-place node of the nasal will not trigger the delinking of the nasal's vocalic node, since the combination of a major [dorsal] node and a minor [coronal] node under the vocalic node is well-formed, and indeed constitutes the canonic representation of a palatalized velar:

(57)



The two models also differ in the means they use to distinguish major and minor articulator features. In the Sageyian model, as we have seen, this distinction is not made by node organization and requires the pointer. Aside from this function, the pointer plays no role in the theory. In the constriction based model, major and minor articulator features are distinguished by node organization alone, since the major feature is always the superior node in the hierarchy. The major feature is always the superior node in the hierarchy. The major articulation in any complex consonant is interpreted with the values of the stricture feaures [continuant, approximant, sonorant] present in the higher structure, and the minor articulation is assigned its noncontrastive degree of closure by independent phonetic rules and principles. In this theory, the pointer is not necessary.[36]

3.5.2 The Node Structure of Vocoids

Given this account of minor articulations, we may return to an earlier question concerning the internal structure of vocoids. We have seen that the vocalic node characterizes the functional unity of vocalic features, and expresses minor articulations in consonants. In the latter, as we have seen, the vocalic node links crucially under the C-place node.

There is reason to believe that the same structure holds in vocoids. It is a striking crosslinguistic generalization that consonantal place features do not appear to be able to spread as a unit from one consonant to another across vowels (Clements 1990b, 1993).[37] For example, although we commonly find rules in which a nasal assimilates to an adjacent consonant in all its place features, we never find rules in which a nasal assimilates to all place features of a consonant *across a vowel*. Thus while rules having the effect of (58a) are common, rules like (58b) appear to be unattested:

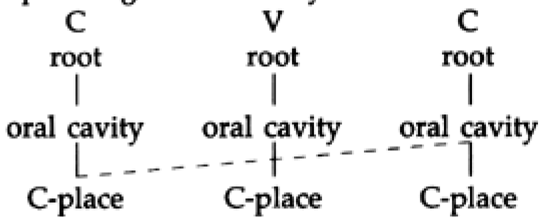
(58)

This fact cannot, apparently, be explained in terms of any general prohibition against the spreading of place features to a nonadjacent consonant, since single articulator features are not constrained in this way. For example, many languages have rules of coronal assimilation in which the coronal node spreads from consonant to consonant across vowels and certain consonants.[38] The rule of n-retroflexion in Sanskrit is instructive (Whitney 1989). By this rule, the first /n/following retroflex [s] or [r] is retroflexed to [n], provided no coronal consonant intervenes, and a sonorant or vowel follows (Schein and Steriade 1986, after Whitney 1989). Consider, for instance, the base form/brahman-/ “brahman”, from which a number of inflected forms are derived. We find, for example, that [n] is assimilated to [r] in the locative singular [brahman-i], though not in the vocative singular [brahman], where no sonorant or vowel follows. Following the analysis of Schein and Steriade (1986), the rule in question spreads the coronal node of the [r] rightward across the intervening vocoids and noncoronal consonants to the following [n]; since the coronal node dominates the features of retroflexion, these features travel with it. Thus, Sanskrit shows that single articulator features, such as [coronal], may spread across vowels and consonants alike.[39]

Both of these patterns follow directly from the structure of the model. The assimilation of all consonantal place features as a unit can only be expressed as the spreading of the C-place node. If vowels also bear a C-place node, the C-place node of consonants cannot spread across them without violating the NCC (27), as shown below:

(59)

Spreading is blocked by the NCC:



(For the same reason, a consonant's oral cavity or root node cannot spread across a vowel.) In contrast, vowels are not opaque to the spread of a *single* articulator feature. For example, a front (i.e., [coronal]) vowel does not block the spreading of [coronal] in Sanskrit, since the front vowel's [coronal] node links to the V-place tier, while the consonant's [coronal] node links to the C-place tier, by our assumptions. (Recall that the NCC (27) applies only to association lines linking elements on the *same* tiers.)[40]

3.6 Are Articulator Features Binary in Vocoids?

At this point, an obvious question arises: Since the articulator features [labial], [coronal], and [dorsal] are one-valued in consonants, shouldn't they be treated as one-valued in vowels as well? But wouldn't such treatment be empirically wrong, given that standard feature theory treats [round] and [back] as binary?

As it turns out, however, the evidence in favor of the binary nature of [round] and [back] is far from overwhelming. Already, Steriade (1987a) has noted that it is difficult to find genuine cases in which [-round] spreads. Although both values of [back] appear to spread, in the constriction-based framework rules spreading [-back] can be reinterpreted as rules spreading [coronal], and rules spreading [+back] as rules spreading [dorsal]. The real problem cases for a fully one-valued interpretation of articulator features in vocoids involve rules which have traditionally been defined on [α back]. These are of two main types: (i) assimilatory rules in which both values of [back] must spread, and (ii) dissimilatory rules which assign some vowel the value [-α back] in the presence of an adjacent [α back] vowel. We briefly review one example of each type below.

In the system of palatal vowel harmony in Turkish, as described by Clements and Sezer (1982), harmonic suffixes acquire the value [αback] from the first preceding vowel. Most consonants are transparent to harmony, as shown in (60a). However, the underlying palatalized consonants /l̦ r̦ k̦/ and the back velar /K/ are opaque, blocking harmony from the preceding vowel and instituting new harmony domains of their own, as shown in (60b, c).

(60)

(a) Regular vowel harmony (nom. sg./acc. sg.)

ip	ip-i	"rope"
kɨz	kɨz-ɨ	"girl"
ek	ek-i	"joint"
tak	tak-ɨ	"arch"

(b) Opaque [–back] consonants /l̦ r̦ k̦ /

sual̦	suāl̦-i	"question"
har̦f	har̦f-i	"letter"
/idrak̦/	idrak̦-i	"perception"

(c) Opaque [+back] consonant /K/

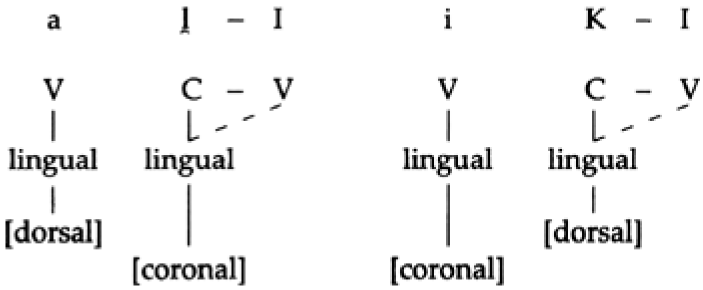
tasdiK	tasdiK-ɨ	"confirmation"

In Clements and Sezer's analysis, opaque consonants are assigned the phonetically appropriate value of [±back] as a feature of secondary articulation. Since all instances of [±back] occur on the same tier, the opacity of the consonants in (60b, c) follows from the No-Crossing Constraint (27).

To interpret these data in terms of one-valued features [coronal] and [dorsal], we must find a way of spreading these features to the exclusion of all others, while accounting for the opacity effects. Consider a possible analysis along the following lines. Let us assume that Turkish vowels fall into two classes, palatal and velar, defined as [coronal] and [dorsal] respectively. Let us fruther suppose, following a suggestion by Browman and Goldstein (1989), that these two features form a single constituent, termed *lingual* in view of the fact that both involve the tongue. On these assumptions, vowel harmony can be expressed as the spreading of the lingual node. The palatalized consonants /l r k/ are now underlyingly specified for the feature [coronal], and /K/ for [dorsal],

both of which constitute minor articulations under the vocalic node. Since these features link to the lingual node, which lies on the same tier in consonants and vowels, they will block the propagation of the lingual node from the preceding vowel, and will themselves spread onto the suffix vowel /I/, as follows:

(61)



This analysis predicts that consonants specified for [coronal] or [dorsal] as secondary articulation will always block the spreading of the lingual node.[41]

An example of the apparent binary nature of [back] of the second type can be drawn from Ainu, as discussed by Itô (1984). In this language, whose vowels are /i u e o a/, vocalic suffixes are added to CVC roots to form CVC + V stems. After many stems, the suffix vowel is simply a copy of the root vowel. However, after a lexically marked set of roots, it is realized as the high vowel [i] or [u] which has the opposite value of [back] from the root vowel. Examples include *ket-u* "to rub" and *pok-i* "to lower".

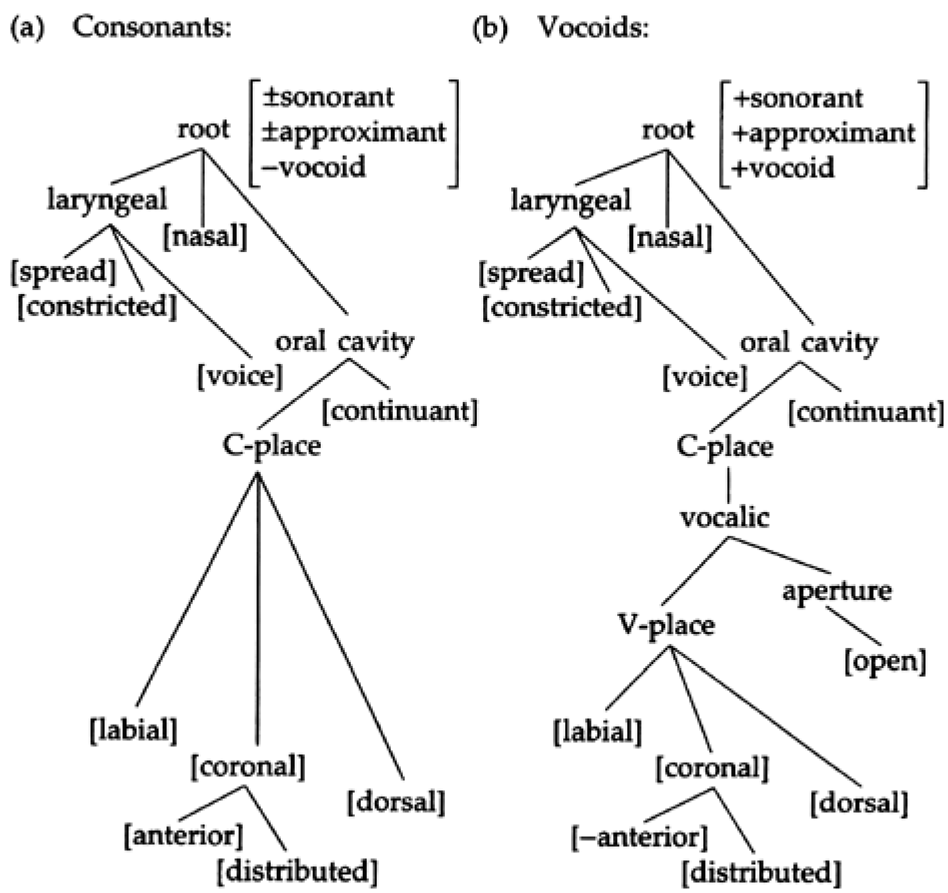
We may account for this pattern without recourse to a binary feature [±back] on the assumption that an OCP-driven constraint applies to stems, disallowing two adjacent identical lingual nodes. (This analysis presupposes that the lingual node may not be multilinked.) Since by principles of contrastive feature specification, every non-low vowel must have at least one lingual feature, the only way the suffix vowel can be realized consistently with the OCP is by selecting the alternative lingual feature from the root vowel. Thus it must be [dorsal] if the root vowel is [coronal], and [coronal] if it is [dorsal].

There is some evidence that the lingual node may be needed in the description of consonants as well as vocoids. Note that the class of lingual consonants is coextensive, in the buccal cavity, with the class of nonlabial consonants. Thus rules that appear to require reference to the class of [-labial] sounds can be reformulated as rules defined on lingual sounds. Examples are not hard to find. In Mandarin Chinese, for instance, lingual obstruents (velar, uvular, retroflex, and dental, except for the dental nonstridents [t t^h]) are replaced by laminal palato-alveolars before the high front vowels [i ü], while labials occur freely in this position (Clements 1976). In Slovak, [æ] is backed to [a] after lingual, but not labial consonants (see note 43). Thus there is at least suggestive evidence that the lingual node may be needed for consonants as well as vowels. In sum, if this somewhat speculative analysis is on the right tract, it would appear unnecessary to retain binary place features in vocoids.

3.7 Summary and Discussion

We summarize the discussion up to this point in the form of figure (62), illustrating some of the better-established class nodes and their form of organization in consonants and vocoids (as noted above, consonants with secondary articulations include a vocalic node under the C-place node, not illustrated here):

(62)



Any particular segment is represented with an appropriate selection of these features (among perhaps others) in its fully specified form. For instance, [k] has a dorsal node under the place node, but labial and coronal nodes are absent. A labiovelar consonant such as the Yoruba [kp] has both labial and dorsal nodes. Some features, such as [−voice], [+cont], and [−anterior], are universally noncontrastive in vocoids. Any speech sound can be represented in this general form. Following the universality principle (5) discussed above, we suggest that this mode of organization holds for all segment types in all languages.

A few further comments are in order. First, (62) differs from the earliest proposals in not including a supralaryngeal node. McCarthy (1988) has shown that alternative explanations are available for most of the phenomena (especially those involving debuccalization) that were originally cited in its favor. However, Dell (1993) offers new arguments for this node based on assimilation rules in two East Asian languages. In a Chinese dialect spoken in the Yongding prefecture, Fujian province, syllable-initial /h/ assimilates all supralaryngeal features from a following syllabic nasal, retaining only its aspiration. Thus the form /hm/ is realized as [Mṃ], /hṇ/ as [Nṇ], and /hn/ as [Nŋ] (upper-case letters designate voiceless aspirates). Here, apparently, the supralaryngeal node of the nasal spreads onto /h/, whose inherent laryngeal features are preserved. In Yi (a Tibeto-Burman language), in certain syllables whose onset is a (voiced or voiceless) sonorant and whose peak is a high vowel, the supralaryngeal features of the onset consonant spread onto the peak: thus, /M**ɨ** / is realized as [Mṃ], /L**ɨ**/ as [Lḷ], etc. Again, an analysis in terms of supralaryngeal node spreading readily suggests itself; Dell shows that a number of alternative analyses can be rejected. As examples of this sort are still rare, we have not included the supralaryngeal node in (62), but further cases would support its reconsideration.

Second, the discussion so far has not touched on two features whose affiliation is still unclear, [lateral] and [strident]. In the case of [lateral], the two competing hypotheses are attachment under the coronal node or the root node. The major argument for coronal attachment comes from the node implication criterion, as discussed in section 3.1; if we attach [lateral] under the coronal node, we directly account for the fact that all segments bearing it are phonologically [coronal], without the need for further stipulation.[42] However, there are at least four problems for this view: (a) when a nasal assimilates in place to a lateral sound, it normally does not become lateral (see the Chukchi form *ten-leut* in (35) as well as similar forms in, e.g., Catalan and Yoruba, though Levin (1987) also cites several exceptions to this generalization); (b) when a lateral assimilates in place to a nonlateral, it normally retains its laterality (e.g., Spanish, Tamil); (c) when the oral cavity node spreads from [l] to [s] in intrusive stop formation (e.g., *false* [...l^{t}s]), the resulting intrusive stop is central, not lateral (see section 3.3.3); (d) lateral obstruents may be fully transparent to rules of long-distance assimilation involving coronal obstruents (for the case of Tahltan, see Shaw 1991). These facts strongly argue that [lateral] occurs above place in the feature hierarchy. If so, it may be that [lateral] sounds are universally coronal just by virtue of the way this feature is defined.

Traditionally, [strident] has been used to distinguish the "noisy" fricatives and affricates (labiodentals, sibilants, uvulars) from the "mellow" ones (bilabials, dentals, palatals, velars); see, e.g., Chomsky and Halle (1968). More recently some linguists have suggested that this feature, like [lateral], should be restricted to coronal sounds; if this proposal is correct, it reopens the question of how bilabial and labiodental fricatives can be distinguished in languages like Ewe, in which they form minimal contrasts. Since place assimilation does not usually affect stridency, we maintain the conservative position that [strident] links under the root node, while hoping that future work will clarify the status of this feature.

4 The Expression of Assimilation Rules

We are now in a position to take up the formulation of rules of place assimilation between consonants and vocoids in more detail. Consider, as an example, the rule of palatalization and coronalization in Acadian French, which causes the velar consonants /k g/ to shift to palatalized velars [k^{j} g^{j}] or palatoalveolar affricates [tʃ d3] before front vowels. This rule is optional, the choice between the various realizations being determined in part by sociolinguistic considerations (see Hume 1992, after the descriptions by Lucci (1972) and Flikeid (1988)).

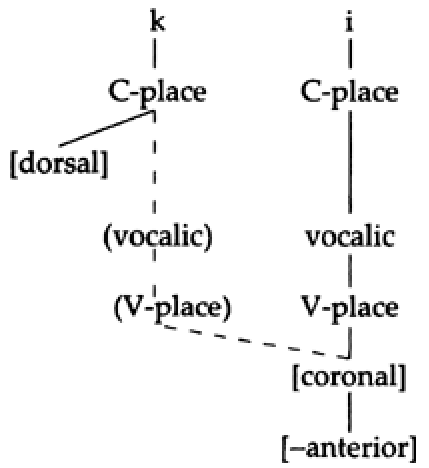
(63)

/kø/	kø	~	k^{j}ø	~	tʃø	"tail"
/gɛte/	gɛte	~	g^{j}ɛte	~	dʒɛte	"to watch for"

The rule must be phonological rather than phonetic, since it has lexical exceptions such as [pike] "to sting" and [mokø] "teasing", which are always pronounced with a plain velar consonant. Furthermore, as Hume notes, it feeds other phonological rules.

The palatalized variants [k^{j} g^{j}] must result from the spreading of the [coronal] feature of the front vowel onto the velar. Specifically, since the velar becomes a palatalized velar, not a coronal, [coronal] must link under its V-place node as a minor articulation. Thus the rule must spread [coronal] from the V-place node of the vowel onto the consonant, with interpolation of new V-place and vocalic nodes as is required to preserve well-formedness. Thus it applies as follows:

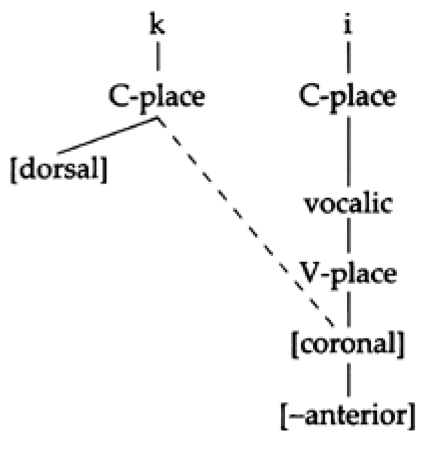
(64)



The fact that palatalization and coronalization have the same set of exceptions suggests that coronalization applies only to forms which have first been palatalized. Words like [pike] "to sting" are marked as exceptions to palatalization, and therefore cannot be coronalized. In this analysis, if palatalization applies to a form, its minor [coronal] articulation may optionally be reassigned major articulator status by a process of *promotion* (Clements 1989a), according to which a consonant's minor articulation is delinked and copied under its C-place node, where it replaces its original major articulation. If the minor [cornal] articulation already bears a redundant [-anterior] specification in the palatalized form, it accompanies the [coronal] node when it is copied, creating a nonanterior coronal, such as the palato-alveolar sounds [tʃ dʒ].

In many other languages, however, there is no direct evidence for an intermediate palatalized stage in the coronalization process. For example, in Slovak the velars /k g x ɣ/ are realized as [tʃdʒ ʃʒ] respectively, when followed by a front vocoid, /i e æ j/, e.g., [vnuk] "grandson", / vnúk+ik/ [vnutʃik] (dim.),/vnúk+æ/ [vnútʃa][43] (dim.). Unlike in Acadian French, velars are never palatalized in Slovak (Rubach, forthcoming). To account for such cases, Hume (1992) characterizes coronalization as an elementary rule type in which the [coronal] feature of front vocoids spreads directly to the C-place node of the velar, replacing its original [dorsal] feature (presumably, again, as the unmarked mode of application). In this analysis, coronalization is expressed as follows:

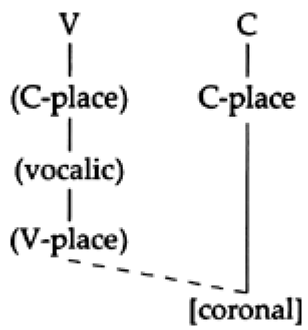
65



The expression of coronalization as an elementary rule type is not possible in a framework in which front vowels are characterized as [dorsal, -back] (Sagey 1986). In such a framework it is inexplicable that the assimilation of a velar ([dorsal]) consonant to a front ([dorsal]) vowel should give rise to a [coronal] consonant. To account for this change, one could, of course, posit some sort of restructuring convention having the effect of trading in the [dorsal] node for a [coronal] one in the context of the feature [-back]. This type of approach is not without problems, however, as is noted by Broselow and Niyondagara (1989), Lahiri and Evers (1991), and Hume (1992). For example, the relationship between [coronal] and [-back] is an arbitrary one. No formal property of the theory predicts that a velar consonant should become [coronal] in the context of a [-back] vowel, as opposed to, e.g., a [+back], or a [+rounded] one. Most important, perhaps, is that such an analysis requires a restructuring rule to account for a common process such as coronalization. Restructuring rules are powerful and highly arbitrary devices. By incorporating them into the theory, we seriously weaken one of our fundamental goals, which is to seek a formalism capable of expressing common processes in terms of simple descriptive parameters.

Consonant-to-vowel assimilation receives an equally simple account in the constriction-based model. We illustrate with an example from Maltese Arabic (Hume 1992). As discussed earlier (see (43)), the vowel of the imperfective prefix is always realized as [i] before a stem-initial coronal obstruent. Assuming that the prefix vowel is underlyingly unspecified, this realization can be accounted for by a feature-filling rule according to which the [coronal] node of the consonant spreads leftward to the vowel, as in (66) (showing interpolated node structure). Vowel height is later assigned by an independently-motivated default rule.[44]

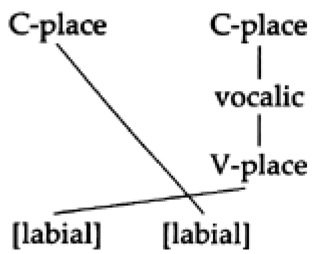
(66)



4.1 The No-Crossing Constraint Revisited

Before leaving the discussion of assimilation, we must consider a further interesting property of the constriction-based model. As we have just seen, this model allows the oral articulator features to link to different tiers: C-place and V-place. As a result, it potentially allows configurations of the following type:

(67)

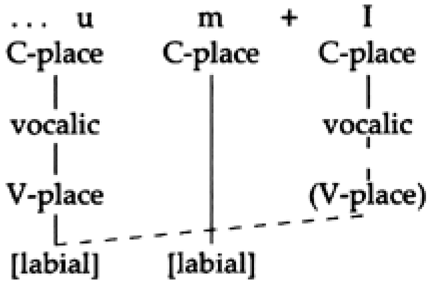


Although the lines linking the two instances of [labial] to higher nodes (C-place and V-place) “cross,” they do not violate the NCC (27), since the higher nodes are not on the same tier. Without some further constraint, then, such configurations are theoretically possible.

However, at present we know of no clear-cut evidence showing that configurations like (67) should be excluded. Indeed, Hume (1992) points out that they may be required in the constriction-based model, at least in the immediate output of rules. Consider, as an example, labial harmony in Turkish. In this system, the labiality of a stem vowel spreads to a high suffix vowel, even across labial consonants.

Thus, the form / mum-I/ “candle” (acc. sg.) is realized as [mum-u]. Labial harmony applies as follows:

(68)



The rightmost segments in this figure present an instance of (67).

Apparently, then, line-crossing must be allowed whenever it does not create violations of the NCC (27). Whether further constraints are required at a later level of description to exclude configurations of this type (for example, to satisfy the requirements of some particular model of phonetic interpretation) is an open question (see Hume 1992 for related discussion). A full examination of this issue, while interesting, would go beyond the scope of the present study.

5 The Phonetic Interpretation of the Feature Hierarchy

In the preceding sections we have reviewed phonological evidence motivating the feature hierarchy. The later discussion has introduced the idea that the basic organizing principle of the feature hierarchy is the *vocal tract constriction*. This view is based on two main considerations. First, articulator features, such as [labial], [coronal], and [dorsal], appear best defined in terms of the constrictions formed by the articulators, rather than using the vaguer notion of “articulator involvement.” Second, the phonological evidence shows that constrictions are represented by specific nodes in the feature heirarchy (oral cavity, vocalic), themselves defined in terms of dependent nodes representing the constriction's location (C- and V-place) and degree (continuance, aperture). We have suggested that this view allows for a new and more adequate treatment of such phenomena as vowel-consonant relations, the internal structure of vowels, and the representation of major and minor articulations.

It would be appropriate to offer some tentative remarks on the possible phonetic interpretation of this model, addressing such questions as, To what extent does feature organization reflect aspects of vocal tract structure? and Why should features be grouped together in terms of constrictions, as opposed to some other organizing principle? In fact, a constriction-based approach receives support from a variety of sources, including acoustic and articulatory theories of speech production, and for this reason, offers a plausible link between abstract phonological structure and phonetic interpretation.

The constriction-based model postulates that segment structure is organized in terms of oral tract constrictios which can combine with independent velic, pharyngeal, and laryngeal constrictions. Quite strikingly, this organization parallels the structure of the vocal tract, in that independently functioning articulations are assigned to independent tiers of the representation, and interdependent articulations are grouped together into constituents. This result, reached independently of phonetic considerations, provides a strong motivation for the model in the physical constraints on phonetic production. Yet at the same time, feature organization is not entirely reducible to physical or physiological considerations. In particular, we have seen evidence from common processes such as assimilation that oral tract constrictions are comprised of two types: consonantal and vocalic, with the latter embedded under the former. Even when produced simultaneously in consonants with minor articulations, these two types of constriction must be assigned to different tiers, and clearly this fact must reflect considerations other than strictly physiological ones. We suggest that this representational difference reflects a fundamental difference in the cognitive status assigned to vocoids and consonants as part of the competence of all speakers. The difference between consonants and vocoids is not merely a matter of their specification for [±vocoid], but involves a fundamental difference in their feature organization.

We must stress, furthermore, that this separation does not lead us to return to a "two-mouth" representation of segments in which consonants and vowels are defined by entirely different descriptive parametrs, as in some traditional approaches (see Ladefoged and Halle 1988 for a critique). On the contrary, our characterization of consonants and vowels is a unified one in the sense that largely the same set of features is used for both, with the organization of consonants and vowels uniformally oriented around the constriction as the basic unit.

That segment structure is indeed constriction-based is suggested by a variety of further observations. Consider, first the fact that most features can be defined directly in terms of the parameters of constriction location and degree. Thus, the place features (the articulator features and their dependents) define constriction location, and the articulator-free features define constriction degree. Note that if the basic unit of organization were articulator "involvement," as assumed in earlier work, we might expect to find features which characterize specific qualities of the articulator's movement (e.g., stiffness, velocity) rather than those relating to constriction shape and location. Insofar as dynamic features of this sort appear to be unmotivated phonologically, we derive further support for a constriction-based model of organization.

Other results in feature theory point in the same direction. As we have seen, McCarthy's studies of pharyngeal consonants (1989b, in press) also suggest that a strictly articulator-based approach to feature organization may be inadequate. This is because the natural class of [pharyngeal] consonants cannot be defined by the movement of any single articulator, but involve a constriction produced anywhere in the region between the oropharynx and the larynx.[45] Furthermore, Steriade's aperture theory, as we have seen, is based on constriction degree, ranging from full oral closure to maximum aperture. All these indicators suggest quite strongly that we are on the right track in viewing feature organization as constriction-based.

But at this point we may ask the question, Why should this be so? That the internal structure of segments is hierarchically-organized is not itself very surprising, given that linguistic structure is hierarchical at all other levels of representation (e.g., syntactic, semantic, morphological). What is less obvious is why phonological features should be organized in terms of the vocal tract constrictions they designate, instead of some other principle. In the remainder of this section we review recent research in speech production theory, which provides further support for the constriction-based organization of features.

Constrictions form the basis of many acoustically-based theories of speech production. These include, in particular, the source-filter theory as presented most completely in the work of Fant 1960 (see also Müller 1848, Chiba and Kajiyama 1941, and Stevens and House 1955), and the quantal theory of speech developed primarily by Stevens (1972, 1989). Fant showed that formant frequencies are determined by the shape of the supralaryngeal vocal tract, which acts as an acoustic filter. In the source-filter theory, the vocal tract is modeled as a tube closed at one end. Within the tube, constrictions typically form pairs of coupled resonators, such that the natural frequencies of any pair are approximately equal to the natural frequencies of the individual resonators, with some perturbation from these values resulting from the acoustic coupling between them.

Developing this model, Stevens finds that when a constriction is appropriately placed, the natural frequencies of the system are relatively insensitive to small modifications in its location; in other words, there are preferred regions within which moderate displacements of the constriction produce negligible effects in the signal. These regions form an important basis for establishing the acoustic and auditory correlates of distinctive features. As far as vowel production is concerned, Stevens (1972, p. 56) concludes that "vowels fall naturally into discrete categories instead of being identifiable as points on a continuum"; these categories, as well as those proposed by Wood (1982), are generally consistent with those that we have defined in terms of [labial], [coronal], [dorsal], and [pharyngeal]. In their further development of this approach, Mrayati, Carré and Guérin (1988) propose that the vocal tract can be divided into eight "distinctive regions" of nonequal length, defined by zero-crossings of the neutral tube sensitivity functions of the first three formants; these regions represent articulatory configurations that produce maximally stable and distinct acoustic targets in Stevens's sense, and again appear to be well correlated to the tongue and lip constriction locations defined by [labial], [coronal], [dorsal], [pharyngeal], and their dependent features.

Articulatory models of speech production also treat constrictions as central. In particular, the task-dynamic model of speech proposed by Browman and Goldstein (e.g., 1989, 1992) is based on the notion of gestures, defined as abstract characterizations of articulator movements whose "task," according to these writers, is the formation of specific vocal tract constrictions. The parallel between their model and feature-based phonological models is striking, and extends to rather subtle details, as they have themselves noted (Browman and Goldstein 1989). This is not to say that there are no important differences between the two models (see Clements 1992), but these differences are not irreconcilable in principle, and should not blind us to the significant parallels between the two approaches.

We see, then, that the notion "constriction" is central to many current theories of speech production, both acoustic and articulatory. It is therefore not surprising that phonological representations may be organized in terms of constrictions as well.

6 Conclusion

This study has attempted to summarize, and as far as possible to synthesize, some of the many recent contributions to the study of segment-internal structure. We have found considerable evidence for a hierarchical, multitiered model of feature organization along the lines presented above. Primary evidence for this model has been drawn from studies of phonological processes and segmental interactions in many languages. This evidence turns out to be surprisingly consistent from one language to another.

We have also seen that feature organization may reflect functional aspects of vocal tract organization in which independent (or partly independent) articulators, determining vocal tract constrictions, are assigned to independent, interacting tiers. In this sense, the model receives additional confirmation from an entirely independent source. While many interesting and important questions remain open and in need of further study, only some of which can be discussed in a general overview of this sort, a hierarchical approach to feature organization promises both to allow a substantially constrained account of phonological organization at the most abstract level, satisfying the requirements of formal linguistic theory, and to offer a bridge between phonological structure and phonetic interpretation which might be profitably explored in future work.

We gratefully acknowledge the helpful comments of Chris Barker, FranÇois Dell, John Goldsmith, John Kingston, and David Odden on an earlier version of this paper. This work was supported in part by a research grant to the second author from the social Science and Humanities Research Council of Canada.

1 For general discussion of features, see, e.g., Trubetzkoy (1939), Jakobson and Halle (1956), Chomsky and Halle (1968), Jakobson and Waugh (1979), Keating (1987), and Halle (1991).

2 The term "nonlinear" was first used in something like its current sense, to our knowledge, by Harris (1941), who distinguished between "successive (linear) phonemes" and "nonsuccessive" or "non-linear" phonemes such as stress. This term continued to be used by Harris, and especially Hockett, in some of their later writings, although it did not gain general currency as a designation for a general class of phonological theories until more recently.

3 It is also implicit in the IPA's graphic organization of consonant and vowel symbols into colums and rows with no further organization; see Ladefoged (1989) for criticism and an alternative proposal.

4 For suggestions that feature organization may be subject to a very limited degree of parameterization, see Mester (1986) and Cho (1990), among others.

5 Previous presentations of feature geometry have modeled phonological representations as sets of lines and planes (Clements 1985; Sagey 1986); however, planar structure is not crucial to the theory, and the following presentation adopts a purely two-dimensional approach.

6 If we prefer to consider that all phonological features are binary, the term "articulator node" (Sagey 1986) may be more appropriate; note, however, that whatever we choose to call them, *labial, coronal*, and *dorsal* have specific phonetic correlates just as other features do.

7 Rules applying in the context of nonlabial sounds have occasionally been cited in the literature. We suggest that these cases, where genuine, might be stated in terms of a node [lingual], to be discussed below. For a discussion of rules apparently referring to the class of noncoronal ("grave") sounds, see Christdas (1988), Yip (1989).

8 More accurately, in Sagey's model [±nasal] is immediately dominated by the soft palate node, which links to the root node.

9 According to Piggott's constraint, any node may immediately dominate at most one value of a given feature. This constraint, unlike (11), allows nodes to dominate sequences of nodes on a lower tier as long as they are not features.

10 We include [−approximant] in the definition of A_0 to exclude laterals, which are frequently analyzed as [−continuant] sounds.

11 Note that such segments cannot be described by linking a sequence like [+nasal] [−nasal] directly to their single root node, since this configuration is prohibited by the No-Branching Constraint.

12 The representation of length in terms of a moraic model such as that of Hayes (1989), in which long vowels occupy two positions and long consonants just one, is less straightforward; see Tranel (1991) and Sloan (1991) for discussion.

13 The explanation may be problematical, given much evidence that epenthetic vowels consist of empty skeletal slots whose content is filled in by later rules (see, e.g., Clements 1986; Archangeli 1988); the insertion of an empty slot would not itself give rise to crossing association lines. Alternatively, we might assume that in some languages only multilinked clusters (i.e., true geminates) are syllabified in the syllable coda, and that unsyllabifiable consonants trigger epenthesis; this would predict the same pattern. The important point for the purposes of the present discussion is that geminates created by assimilation show exactly the same properties as monomorphemic (and *ex hypothesis*, bipositional) geminates. See also chapter 8, this volume.

14 See note 13 for an alternative explanation. Note that in Kolami, place assimilation applies only before underlyingly voiced stops.

15 McCarthy convincingly argues that stems such as [samam] "poison" are derived from an underlying biliteral root / sm/ (McCarthy 1981, 1986).

16 Cho (1990, p. 94) notes that place assimilation is an optional rule which can be suppressed depending on the style and the rate of speech. Kim (1990) reports only the slow speech forms, and Martin (1951) only the fast speech forms.

17 Alternatively, it may delink the oral cavity node (see section 3.3.3 below), though not the place node, which does not dominate [continuant]. In a different analysis, Iverson and Kim (1987) delink *all* terminal features in the syllable coda, accounting for coronal and laryngeal neutralization at the same time. This very elegant rule, if correct, would require a relaxation of constraint (4).

18 See Kiparsky (1985) and Archangeli and Pulleyblank (1989) for discussion of a locality effect derived from marking conventions, according to which segments marked in the grammar as unable to bear a certain feature may neither receive this feature nor allow it to pass across them in the course of spreading.

19 Some linguists have suggested that the sonority degree of a given segment is determined not by features, but by node structure itself: roughly, the more class nodes in the structure, the greater (Rice 1992) or lesser (Dogil 1993) its sonority.

20 These forms, taken from Bogoras and Odden, exhibit the effects of vowel harmony. According to Bogoras (p. 653), [y] may also harden to [d] after /n/.

21 Other analyses of ISF have been propsed. Davis (1989) proposes to treat ISF in terms of two independent rules of [−cont] spread and place spread. Note, however, that since ISF is optional, this analysis predicts that each rule should be able to apply independently of the other in the same dialect, and that some dialects may have one rule and not the other. These predictions appear to be incorrect, since according to the literature on this subject, if ISF applies at all in a given dialect, it applies in toto. Iverson (1989) proposes to analyze ISF as the leftward spreading of [−sonorant]. However, as pointed out above, [sonorant] is not otherwise

known to spread independently of other features, and for this reason is usually represented as part of a feature matrix on the root node, as shown in (32).

22 For example, the Node Fission Convention proposed by Clements (1989b), which has the effect of splitting a single branching node into two nonbranching homologues.

23 (38b) reflects Halle's current terminology, in which “guttural” and “larynx” replace the earlier terms “laryngeal” and “glottis,” respectively.

24 “Front” in (41b) refers to the upper surface of the front of the tongue including the tip, the blade, and the forward part of the body of the tongue, which typically articulates under the hard palate. For the phonetic basis of this definition, see Hume (1992). Note that while the definitions in (41) apply to consonants and vocoids alike, consonants and vocoids typically implement them in somewhat different ways, consistently with their different articulatory requirements. Thus, [labial] consonants require a relatively narrow (and not necessarily protruded) lip constriction in order to acquire the radical vocal tract obstruction which, as noted earlier, is definitional of consonants, while [labial] vocoids require a relatively wide and protruded constriction in order to create a supplementary resonating cavity not sufficiently obstructed to produce consonantal frication. While these two types of lip configurations are somewhat different, both involve a labial constriction in the sense of (41a), and thus conform to the definition of [labial] sounds. Analogous remarks hold for [coronal] and [dorsal].

25 Alternatively, we could place [coronal], [labial], etc., on different tiers in consonants and vocoids, and extend our definition of “adjacency” in such a way that features on different tiers also count as adjacent if they are linked to adjacent root nodes (Selkirk 1988). All else being equal, however, we would prefer the simpler definition. Notice that the OCP applies less frequently to consonant + vocoid sequences such as *ty* than it does to consonant + consonant sequences (though see Clements 1990b, 1993 and Hume 1992 for further examples of “cross-category” dissimilations). This fact can be regarded as a special instance of the more general principle (noted by McCarthy, in press) that the OCP tends to apply in preference to sounds that share major class features.

26 The Sageyian model can express back vowel/velar consonant interactions by assuming that [+back] is redundantly present in (back) velars and uvulars. However, this assumption predicts that velars and uvulars should be opaque to the spreading of [±back] in vowel harmony and assimilation, which is not the case. A striking, if atypical example of a rule in which both front and back vocoids pattern with velars is the “ruki” rule of Sanskrit, in which / r u k i/ cause a following [s] to become retroflex (Whitney 1889). If [i] were [dorsal], as in Sagey's model, and retroflex sounds bear secondary dorsalization, all these sounds could be regarded as [dorsal]. However, it is mysterious why [s] should become retroflex in this context; if we were to spread Sagey's dorsal node rightward, [s] should palatalize to [s^j] after [i], and velarize to [x] after [k].

27 In vowel systems like that of Akan, in which [ATR] has been proposed as the basis of tongue-root based vowel harmony, it may be possible to replace [ATR] with [pharyngeal] or [radical]. Thus, there is increasing reason to believe that [ATR] can be dispensed with altogether.

28 The vocalic node was first proposed in unpublished work by Archangeli and Pulleyblank, who called it the S-place (i.e., secondary place) node.

29 Our discussion here is indebted to unpublished work by Nibert (1991), which first brought these phenomena to our attention.

30 To address problems of this sort, Steriade (1987b) suggested that velar consonants should be characterized by a new “velar” node, with [dorsal] reserved for vowels. This proposal correctly treats velar consonants as transparent to dorsal spreading, but raises other problems. For one, the velar node is an anomaly in articulator theory, since it does not designate an independent articulator. Moreover, as Mester and Itô point out (1989), this proposal makes it difficult to express the fact that velar consonants typically form a natural class with back vowels, not front vowels, as discussed above.

31 Or to the skeleton. Observe, however, that Servigliano does not satisfy the criteria proposed by McCarthy (1989a) for languages with template-based morphologies, in which vowels and consonants lie on entirely separate planes meeting at the skeleton. If we allowed such full segregation of vowels and consonants in all languages, we would predict, incorrectly, that rules of total consonant spreading across vowels, found in such templatemorpholog languages as Arabic and Hausa, would occur freely in languages with

concatenative and fusional morphologies.

32 See Maddieson (1990) for discussion (and rejection) of a proposed exception to this generalization in Shona. Note also that Sagey's definition allows for the possibility of segments with two major articulations. She exploits this possibility in her analysis of the surface contrast between [p^w] and [kp^w] in Nupe, treating the latter but not the former as having two major articulations, [dorsal] and [labial]. If this contrast can be reanalyzed in other terms, as seems possible (for instance, we might suppose that [p^w] does not have a [dorsal] component), then crucial cases of double major articulations, in Sagey's sense, appear to be rare and perhaps nonexistant, and a maximally constrained phonological theory would exclude them in principle, by appropriate constraints on representations.

33 Ní Chiosáin's description implies that either rule may apply in such examples; the choice of rule is not predictable from other phonological factors.

34 Instead, the second dorsal node must be delinked from the nasal, after triggering the loss of the coronal node; but no general principle predicts this delinking.

35 We cannot assume that the oral cavity (or root) node spreads, since [continuant] does not spread, as shown by our examples.

36 Like Sagey's model, the constriction-based model makes no formal claims regarding the phonetic degree of stricture of a minor articulation. It thus allows for the possibility of languages, like those discussed above, in which a minor articulation has the same degree of closure (or narrower) closure than a simultaneous major articulation. See Hume (1992) for further discussion of this point. Further evidence for the linking of minor articulations as a sister rather than daughter of the major articulation node can be cited from opacity patterns in Chilcotin (Clements 1990b, 1993). See also Goodman (1991) for comparison with the dependency-based model of Selkirk (1988), in which minor articulations are treated as daughters of major articulations.

37 It is not inconsistent to link vowels under the C-place node, since this node has no phonetic content. We may consider the C- and V-place nodes as in fact the same category of place, the terminological distinction between them being merely conventional.

38 The spreading of single C-place features (major articulations) to nonadjacent consonants appears to be restricted to [coronal], and in all known cases of [coronal] spreading, the target must also be [coronal]. We speculate that a more general constraint is at work, restricting long-distance C-place spreading to cases in which an OCP violation is involved. In effect, since spreading of [labial] or [dorsal] onto [dorsal] would be vacuous, since these features do not usually have dependents, such a constraint would limit long-distance spreading just to the observed cases. Such cases would then be motivated in a manner similar to rules of long-distance dissimilation which, as was discussed in section 2.2, are also OCP-driven.

39 As David Odden points out to us, if [n] can be regarded as [-distributed], an alternative analysis is possible in which only the coronal dependent feature [-anterior] spreads. For other, less controversial examples of long-distance coronal node spreading, see Poser (1982), Hualde (1988b), and Shaw (1991).

40 A further prediction of this model is that a vowel's vocalic node may not spread across a consonant bearing a minor articulation. This prediction is supported by the rule of vowel copy in Barra Isle Gaelic (Clements 1986) in which the epenthetic vowel is realized as a full copy of the preceding vowel across a palatalized or velarized consonant, except that the vowel is always front if the consonant is palatalized and back if it is nonpalatalized. In addition, the epenthetic vowel is always unrounded, even though rounding is distinctive. To account for these facts, we must assume that the vowel, but that it spreads the aperture node of the vowel and the V-place node of the consonant separately. If the vocalic node were not linked to the C-place node in vowels, we would expect the vocalic node of the vowel to be able to spread, since it would not violate the No-Crossing Constraint, incorrectly resulting in complete vowel copy.

41 Other examples of the spreading of both values of vocalic place features have been cited in Gaelic (Clements 1986) and Chilcotin (Clements 1993, p. 139), and can be treated in a similar way. Note that a further prediction of this approach is that languages may have harmony rules spreading just the [dorsal] or [coronal] node, instead of the lingual node. In such cases, it should be possible for [dorsal] to spread across [coronal] vowels, and vice versa.

42 While phonetic lateral velars have been reported in a number of languages, there is no evidence that any

of these sounds are both [dorsal] and [+lateral] at the phonological level; see Levin (1987) for careful discussion of this issue.

43 The low front vowel /æ/ is backed to [a] by an independent rule after nonlabial consonants. Thus, the diminutive suffix that surfaces as [a] in vnútf + a is the same suffix occuring in chláp + æ “man” (dim).

44 In this analysis, the major articulator feature [coronal] of the consonant links under the V-place node of the vowel, creating the unmarked vowel structure. We assume this is the normal mode of operation. Given our previous analysis of V-to-C place assimilation, however, it is natural to ask whether there are also two types of C-to-V spreading: one in which the consonant's major articulator feature links under the vowel's V-place node, as above, and another in which it links directly under the vowel's C-place node. These two analyses make subtly different predictions, as discussed by Hume (1992); we leave the question open here.

45 This conclusion does not of course follow from the alternative proposed by Halle (1989, 1992), in which [pharyngeal] is not an articulator feature but a class node, renamed “guttural.”

Cite this article

CLEMENTS, G. N. and ELIZABETH V. HUME. "The Internal Organization of Speech Sounds." *The Handbook of Phonological Theory*. Glodsmith, John A. Blackwell Publishing, 1996. Blackwell Reference Online. 31 December 2007 <http://www.blackwellreference.com/subscriber/tocnode?id=g9780631201267_chunk_g97806312012679>

Bibliographic Details

The Handbook of Phonological Theory

Edited by: John A. Glodsmith
eISBN: 9780631201267
Print publication date: 1996