Autosegmental Phonology and Underspecification Theory

1. Autosegmental phonology: history and essentials
2. Best studied autosegmental system
3. Feature geometry
4. Underspecification theory: history and development
5. The theoretical basis of underspecification theory
6. Conclusion

1. Autosegmental phonology: history and essentials

Autosegmental phonology is an approach to certain problems in formal phonological theory, especially in the area of prosody, which seeks explanations for traditional phonological problems in the ways that relatively small units -- smaller than traditional phonemes -- combine geometrically to form the larger, observable components of a spoken utterance. These techniques have been used in the analysis of such areas as tonal systems, vowel and nasal harmonies, vowel and consonant length, morphological control of vowel/consonant sequencing, constraints on possible assimilations, and certain aspects of syllabification.

In certain respects, autosegmental phonology continues a line of exploration begun by some American phonologists in the 1940s and 1950s, including Bernard Bloch, Zellig Harris, Charles Hockett, and Kenneth Pike. This tradition within American structuralism was concerned with the fact that a number of basic problems in phonological analysis require treating the sound stream not as a blend of a single sequence of phones, but rather as the production of an orchestra, composed of the independent articulators of the vocal apparatus, each producing a separate melody, but each not always matching the other note for note, nor organized together rhythmically but for beat. At the same time, phonologists in Great Britain, under the influence of J. R. Firth's original line of exploration, developed certain parallel ideas exploiting Firth's specialized notion of prosody, and this approach gave rise to a number of influential publications which analyzed problems of prosody in a number of disparate languages of the world.

Autosegmental phonology was developed within the tradition of classical generative phonology, that of The Sound Pattern of English (Chomsky and Halle 1968, hereinafter SPE), first in Goldsmith (1980 [1974], 1976a,b) and then quickly by other researchers, such as Clements (1976a), McCarthy (1979, 1981), and Harris (1980). In the subsequent period most of the concepts and tools of autosegmental phonology have become standard equipment in generative phonology as well as in other frameworks that interact with phonology in the anglophone world.

At the heart of autosegmental phonology lie three notions:

First, that phonological representations have a geometry more complex than that of a single sequence of phones; minimally they consist of a set of parallel rows (or tiers) of segments, where the segments on each tier are specified for a set of features specific and unique to that tier, and where the segments on each tier are associated with segments on other tiers by association lines. Association lines

in the derived or phonetic representation indicate a relationship of simultaneity, while at deeper levels of representation they specify a more abstract relationship among the segments on separate tiers (also known as autosegments). A typical example is given in (1), which illustrates with an example from Tonga how a problem of tone is addressed in an autosegmental approach.

(1) placed approximately here

Second, the proposal to study phonological problems through the use of tiers and association lines is illustrative of a more general principle, the notion that many problems in phonology could be reduced to the study of the ways in which units, small enough to be considered subphonemic, combine locally to form larger units, much as electrons and protons combine to form atoms, and atoms combine to form molecules. This notion contributed to the development of underspecification theory (see below), as well as other accounts of segmental systems not involving underspecification (e.g., Goldsmith 1985).

Third, the principles of autosegmental combination emphasized the possibility of (a) reduced dependence on language-particular rules (especially in conjunction with the Wellformedness Condition as an active principle; see below) and hence (b) decreased abstractness in phonological representations.

Six basic arguments were offered during this initial period for autosegmental representations:

a. Stability. It is often the case that the deletion of a major segment, such as a vowel or a consonant, leaves behind a phonological trace of its presence, such as the tone or nasality of the segment, or the increased quantity of the segments that remain. Autosegmental representation accounts for this stability by effecting the deletion on one tier, leaving the other phonological material on another tier unaffected by the deletion; the remaining material is free then to reassociate to the other segments.

b. Contour tones (many to one and many to many associations): In certain domains, notably that of tone but also in the case of diphthongs, certain nasal sequences, and some other complex segments, a single major segment, such as a vowel or a consonant, appears to be specified twice, in sequence, for a specific feature. For example, a vowel with a falling tone can often be shown to be specified in its early period as a high tone, and specified later in its period as a low tone. This cannot be done coherently if segments are treated as being atomic, but autosegmental phonology proposes that the tonal features, for example, are placed on a distinct autosegmental tier, and hence two or more tonal autosegments are permitted to associate with a single vowel position.

c. Floating tones: Students of tonal systems have long recognized the existence of "floating tones", phonological units that have a clear status in the grammar of a language, and yet which contain no major segments (vowels or consonants). Autosegmental representation provides a means of dealing with the presence and the behavior of such floating elements, which consist of autosegments existing purely on such tiers as the tonal tier, the nasality tier, and so forth.

d. Action at a distance: Much work on autosegmental phonology, especially the work discussed below on feature geometry, has focused on the possibility of establishing the generalization that all phonological action is local: that is, that when one phonological segment affects another (whether it is a case of deletion, assimilation, dissimilation, or any other sort), the two segments must be adjacent. Even

from a traditional, purely linear perspective, this generalization often holds, as in cases of regressive place of articulation assimilation, or palatalization before high front vowels. Other cases of natural assimilation are not so obviously local (for example, vowel harmony, or assimilation of the tone of one vowel to that of the next vowel), but become purely local if the features in question are analyzed as residing on a separate autosegmental tier. Given a particular account of feature geometry, the locality condition becomes a very useful diagnostic tool in phonological theory.

**e. Melody levels:** in many languages, a subset of the phonological features function as separate morphemes. Again, tone is the clearest case, where it is common for tone to be determined not by lexical elements (especially in verbal systems) but to form a separate morpheme marking, for example, tense, mood, or affirmation. This possibility arises in autosegmental phonology because the segments on each tier represent autonomous parts of the representation, and therefore can be independent lexical entries. This observation was critical in the development of the analysis of Arabic consonantism and vocalism.

**f. Geometrization of quantity** The phonological length (not the phonetic duration) of long vowels and long (or "geminate") consonants has a dual character. In some respects, such long segments act like single segments, while in other respects they behave like sequences of two segments. This apparent contradiction can be well handled within an autosegmental framework.

The focus on properties of the geometry of phonological representations, and on the automatic implementation of automatic processes to insure some of that geometry's basic consistency, was encapsulated in a three part principle called the Wellformedness Condition (WFC) (Goldsmith 1980 [1974], 1976a,b), whose form is illustrated in (x) in the case of the treatment of tonal systems (in which tones are on one autosegmental tier, while vowels are on a separate tier).

(2) around here

Subsequent work (especially by D. Pulleyblank) established that cases (a) and (b) are not cross-linguistically universal. Many languages place a restriction on tones, permitting them to associate to no more than a single syllable, and in such cases in the absence of sufficient tones, syllables will remain toneless. Condition (c) has remained an important principle for analysis.

Another principle, the Obligatory Contour Principle (OCP), expresses the restriction frequently found against adjacent identical autosegments (the name coming from the fact that two non-identical adjacent tones typically form a contour, or dynamic, tone). The OCP was proposed in Goldsmith (1980 [1974], 1976b), reflecting an observation in Leben (1973), but Goldsmith (1976b) argued that it was an epiphenomenon due to principles of language acquisition. Later work, especially by Odden (1986), McCarthy (1986) and Yip (1988), has continued the debate regarding the scope and interpretation of the OCP.

2. Best studied autosegmental system

We will survey some of the best studied cases utilizing autosegmental phonology: tonal systems, skeletal tier, harmony systems, and melodic templates.

2.1 Tonal Systems. A large proportion of the world's languages are tone languages, in that lexical or grammatical contrasts are expressed by the tones realized on the syllables of the word. The tonal systems of Africa, Asia, and of Central America had been the subject of considerable scholarly effort, but their distinctive characteristics continued to elude a satisfactory formal analysis using the tools of

formal phonological theory available to generative phonology and other developed phonological theories that did not employ the geometrical resources of autosegmental phonology.

2.2 Skeletal tier: Work by McCarthy (1979, 1981) on the autosegmental treatment of vocalism and consonantism of Arabic led quickly to a number of studies of the treatment of vowels and consonants from an autosegmental perspective, notably Clements and Keyser (1983), which also developed further Kahn's formal treatment of syllabification (1976).

These proposals encouraged work on the analysis of phonological quantity. The treatment of vowel and consonant length was already beset by a number of problems in the purely linear system of representation espoused both by many structuralists and by classical generative phonology. The most fundamental problem was how to resolve the paradoxical character of long vowels and consonants, for in some respects these elements behave like single units, while in other respects they behave like sequences of two units. (They act like single units in the respect that many languages which prohibit obstruent clusters from appearing intervocally, like Japanese or Hausa, nonetheless permit long [i.e., geminate] obstruents intervocally; they act like sequences in terms of their contribution to syllable weight: in a quantity-sensitive accentual system in which a closed syllable (of the form CVC) attracts accent, a long or geminate consonant appearing intervocally will cause the syllable to its left to be treated as a closed syllable.)

Other properties of long vowels and consonants were studied by Kenstowicz and Pyle (1973), and later within an autosegmental framework by Schein and Steriade (1986) and Hayes (1986). The most important of these additional properties were the strong tendencies for geminate consonants to be immune to separation by epenthetic vowels (dubbed integrity) and their tendency not to undergo other mutation processes, most notably consonant weakening processes (inalterability).

2.3 Vowel and Nasal Harmony. Vowel harmony has played an important role in the development of phonological theory. In the development of distinctive features, the importance of vowel harmony systems has arisen from the perception that the contrasts governed by vowel harmony normally involves one single phonological feature, not a random amalgam of phonological differences; hence even a single vowel harmony system can be a sharp diagnostic tool for a general theory of vowel features. In the classical period of generative phonology following the publication of The Sound Pattern of English (Chomsky and Halle 1968), some important criticisms of that model arose from the treatment of vowel harmony systems, notably in Kiparsky (1968). Questions such as the following were raised: (a) in a vowel harmony system where all the vowels of a stem are typically in agreement for a the harmonizing feature, is it possible to specify in an non-arbitrary fashion a single vowel among those in the stem as the underlying bearer of that specification? If the vowel harmony class of a stem is treated as a property of the morpheme as a whole, rather than as a property of an individual segment, does it not follow that vowel harmony is being treated as a morphological rather than a phonological phenomenon; and does it not follow that if we treat vowel harmony as a property of the morpheme we deny the phonological character of one of phonology's most important sources of insight? (b) if the vowels in the stem are fully specified, is it in any way possible to treat harmony within the stem and harmony between the stem and its affixes in a unified fashion? (c) do neutral vowels require the postulation of abstract vowels that get normalized late in the derivation?

The application of autosegmental theory to vowel harmony systems resolved some of these issues. By treating the harmonic feature on a separate autosegmental tier, and by treating it as underlyingly unassociated, it would be possible to resolve the unfortunate asymmetries perceived in the analysis of
vowel harmony in classical generative phonology: no vowel of the stem was privileged, association to stem and affixes was formally parallel, and the treatment was thoroughly phonological rather than morphological. Nonharmonic stems, in which irregularly the vowels did not all agree with respect to the expected vowel harmony pattern, were analyzed as being stems which irregularly were endowed with two vowel harmony autosegments underlyingly, one of them preassociated.

Abstractness. One of the major problems faced by the classical theory of generative phonology was the lack of limits on abstractness in phonological analysis. While abstractness is by no means an easy notion to define, a consensus arose that abstractness could be characterized as the positing of segments or segment types in underlying representation that do not appear on the surface. Two cases were central: first, those where a segment S was posited underlyingly in a morpheme M which never surfaced with all its underlying specifications in any word in which the morpheme M appeared (though the segment S might occur on the surface in other words with other morphemes); and second, those where an underlying segment T was posited underlyingly which never surfaced as such anywhere in the language. While never designed to deal with the issues circling around these cases, autosegmental representations changed many of the assumptions that went into them. For example, in the case of a vowel harmony system containing a stem whose vowels all are drawn from one class of harmonic vowels, but whose suffixes choose vowels from the other harmonic set, generative phonological accounts needed to posit either an abstract vowel appearing in the stem belonging to the suffix's harmony class, or else deal with the facts by means of rule exceptions. An autosegmental account, however, would posit a suffixal harmony autosegment which would associate to any suffixal material, causing those suffixal vowels to be from the appropriate vowel harmony class, distinct from the stem's harmony class. This suffixal autosegment is not abstract, however, by any technical definition: it appears directly in the surface phonological representation.

2.4 Melodic Templates. McCarthy (1979, 1981) argued that the long-studied patterns of vowels and consonants in Arabic, Hebrew, and other Semitic languages could be better understood if their formal properties were directly compared to the treatment of autonomous tones in African languages. Comparing Semitic consonantism, for example, to the way in which Tiv, a Niger-Congo language of Nigeria, utilizes tones to specify tense and aspect, McCarthy proposed that the Arabic verb stem should be thought of as being constructed out of three autonomous autosegmental tiers. The central tier specified an abstract pattern of vowels and consonants, serving as the backbone of the representation (and hence coming to be known as the skeletal tier). The three (or occasionally two) consonants that determine lexical semantics are placed on a tier for consonants, while the pattern of vowels which express voice and aspect are placed on their own vocalic tier. The vowels on their tier associate to the vocalic position on the skeletal tier, just as the consonants on their tier associate to the consonantal positions on the skeletal tier, creating a representation as in (3). The skeletal position thus plays a double role: it determines the relative order of the vowels and the consonants (though not the order of the consonants with respect to each other, or the vowels with respect to each other), and it determines the patterns of single and double positioning for both consonants and vowels (in the sense that certain skeletal patterns require the double association of the second consonant, and also that lexical stems composed of two or four consonants, rather than the canonical three, will have special but largely predictable patterns of association to the skeletal pattern).

(3) placed approximately here

Similar patterns of verbal structure have been adduced in several languages, and this general style of interaction of morphology and phonology led directly to the growth of prosodic morphology.

3. Feature Geometry

Shortly after McCarthy's work, discussed above, which proposed to treat certain cases of consonantism and vocalism within an autosegmental context, two proposals were made to extend this treatment to subsegmental phenomena (Goldsmith 1979, 1981; Thrainsson 1978), both considering first cases where laryngeal features act autonomously, but also moving on to consider point of articulation as a complex specification acting autonomously. These considerations also reflect suggestions just a few years earlier by a number of linguists, including notably Stephen R. Anderson (1974) and Roger Lass (1976), focusing on the featural autonomy of these and other sets of phonological features.

Work by Mohanan (1983), Mascaro (1983), Clements (1985), and Sagey (1986) led to a model of feature geometry, which modified the autosegmental model to include two new formal devices, class nodes and a hierarchical relationship among features and class nodes. In this model all phonological features occur on separate autosegmental tiers (called feature tiers), in a fashion that is fixed across languages (that is, not available for variation from language to language). Features are organized in a universal, hierarchical fashion into groups of features, each such grouping being represented as an autosegment called a class node (for example, the two or three features defining tone, or the features that define consonantal point of articulation). The class node dominates the autosegments on the feature tiers, and may be thought of as inheriting its specification from them. Class nodes in turn are dominated by higher class nodes, all the way up to the highest class nodes, which form a tier called the root node tier. This structure is illustrated in (4). A (feature or class) node will normally be dominated by a geometrically higher node within its original segment, so to speak, but under the influence of a phonological rule, it may be dominated by a class node in a neighboring segment. In this way, the lines which indicate a pure dominance relationship within a segment play the role also of association lines when a (feature or class) node is multiply associated. One of the original guiding intuitions of this model is the principle that all assimilations will be modeled by adding exactly one association line to a representation. This will always result in an autosegment becoming multiply associated (or, of course, in becoming more multiply associated if it is already multiply associated). If the autosegment is a single feature (hence a terminal node of a feature geometry tree), the assimilation involves that single feature; if the autosegment is the highest node, the root node, the assimilation is a total assimilation; if the autosegment is a class node in-between these two, the assimilation is a partial assimilation involving just those features specified on that class node.

(4) located approximately here.

A great deal of research in the following years has gone into developing a single, consistent organization of phonological features which can deal with problems of assimilation, weakening, etc., from this point of view. The most important constraints placed on this research program from its earliest stage have been (a) that the feature set and hierarchical organization is universal, not language particular, and (b) that assimilation rules add exactly one association lines, so that the class of possible assimilation rules forms an easily specified and universal lattice. (Each of these assumptions have been experimentally dropped by certain researchers, but their importance in maintaining feature geometry as an empirical enterprise has made this move appear suspect.) One important result that has emerged is the treatment of oral point of articulation using three privative features (Labial, Coronal, Dorsal), each corresponding to the involvement of one of the major articulators of the mouth.

4. Underspecification Theory: History and development

Most theories of phonology agree that the mental representations that underlie a phonetic signal contain less information than the signal itself. Arguments for this position are easy to come by, and derive almost exclusively from the fact that many properties of a phonetic signal are predictable. Consider, for example, aspiration in English stop consonants. Voiceless consonants are generally aspirated when they occur in the syllable onset, while they are unaspirated in the coda, making aspiration predictable in this class of segments. Presuming, as generative phonology does, that predictable information is not included in a speaker's mental lexicon, words such as [tsæb] 'tab' and [phln] 'pin' are stored without any information about the aspiration of their initial consonants. The generative position is based on the assumption that the mental storage space available to humans is limited and must therefore be kept free from redundant information, an assumption that Steriade (1995) calls "lexical minimality". This hypothesis is supported by second language acquisition and slips of the tongue, which suggest that many aspects of the realization of words are indeed rule-governed.

The most important question that arises in conjunction with the assumption of underspecification regards the nature of mental representations. One major tradition within generative grammar has held that morphemes, and not words, are the smallest meaningful units that are stored in the lexicon. It would be thus the lexical representations of morphemes and not of words which would be the smallest meaningful units stored in the lexicon. It would thus be the lexical representations of morphemes that would be redundancy-free. It is a well-known fact, however, that there are many regularities with respect to how segments are strung together to form morphemes. In English, for example, there are no morphemes that begin with a cluster of more than three consonants. Halle (1959) and Chomsky and Halle (1968) propose to capture such regularities through lexical redundancy rules, which, as their name suggests, apply in the lexicon.

In Halle (1959), sequence structure rules are placed at the beginning of the phonological component, while segment structure rules and phonological rules apply over the course of a derivation, the former to fill in missing feature values, and the latter to change existing feature specifications. This position is changed in the latter part of The Sound Pattern of English (Chomsky and Halle 1968), where all redundancy rules are relegated to the lexicon, supplying missing feature specifications before any phonological rule has a chance to apply. This change was considered necessary in light of Lightner's (1963) and Stanley's (1967) objection that the inclusion of blanks in underlying representations leads to the ternary use of binary features. Most notably Stanley (1967) argued that if underlying phonological representations contain zero entries in addition to plus and minus values of features and if phonological rules are allowed to change feature specifications, then they can derive three distinct matrices in violation of binarity. Stanley therefore proposed that all morpheme structure rules apply before the phonological rules of a language, so that the input to the phonological component is fully specified.

Unlike phonological rules which can alter existing structures, the function of lexical redundancy rules in The Sound Pattern of English is to spell out what the expected or unmarked combination of segments and features in morphemes is. Assuming, as dictated by the hypothesis of lexical minimality, that expected or predictable features are absent from lexical representations, it is the task of lexical redundancy rules to supply these features.

There are, however, a number of generalizations about the phonological structure of a language that hold not only for morphemes, but for entire words. For example, in Turkish, vowels agree for the feature [back] (and high vowels also for the feature [round]) with a preceding vowel, which can be

formalized linearly as $V \rightarrow [\alpha \text{ back}] / V[\alpha \text{ back}]$, applying iteratively. Agreement for the feature [back] holds within morphemes as well as in morphologically complex words and is thus both a morpheme structure condition and a condition on the phonological structure of words in Turkish. Within generative phonology of the SPE-tradition this generalization is captured in two components of the grammar: Once as a lexical redundancy rule which supplies every vowel in a morpheme with a specification for the feature [back], and then again as a phonological rule that applies in the phonological component of the grammar, changing the backness specification of suffixes, for example, in agreement with the last root vowel. Root and suffix vowels thus receive their value for the feature [back] through the application of two separate rules. Ideally, however, the distribution of back and front vowels in Turkish should follow from the application of just one rule which applies both within morphemes and across morpheme boundaries. The morpheme internal application of the harmony rule is, however, excluded in the post-SPE literature because of the Alternation Condition (Kiparsky 1968) which prohibits the application of obligatory neutralization rules in non-derived environments. Since on the classical generative view, root vowels in Turkish are fully specified for backness in the input to the phonology, the harmony rule could neutralize the underlying distinction between [-back] and [+back] vowels, if it applied root internally.

As a solution to this notorious duplication problem, Ringen (1975) suggested that feature-filling rules are not obligatory neutralization rules. On this modified view of obligatory neutralization rules, the harmony rule of Turkish could apply root-internally without violating the Alternation Condition, provided root vowels are unspecified for the feature [back] in the input to the phonology. Ringen thus proposed that Turkish root vowels (with the exception of the first harmonic vowel) are unspecified underlyingly and that the function of lexical redundancy rules be limited to capturing generalizations which are truly morpheme-specific; i.e., they should not fill in feature specifications that are predictable through the larger phonemeological context. She thus presents the first argument for segmental underspecification in the post-SPE literature.

Independently of Ringen, Clements (1976a,b) began to use underspecified representations in his analysis of vowel harmony systems. His purpose, as we have noted, was different from Ringen's, however: the starting point of Clements' analysis was that harmony features have a span or domain that is longer than a single segment, an assumption that is best expressed by placing the feature in question on an autosegmental tier of its own. A consequence of the autosegmental approach to harmony is that segments are unspecified for the harmony features before they associate.

Ringen's (1975) dissertation is also noteworthy for another idea. In her analysis of tongue root harmony in the West African languages Igbo and Diola Fogny, Ringen argues that only the feature value [+ATR] is specified underlyingly, while the complementary value [-ATR] is absent and filled in by a default rule. Although she does not draw a connection between markedness generalizations, on the one hand, and the absence of the value [-ATR] from underlying representations, on the other, her analysis foreshadows an approach that has gained much attention in phonological theory over the last fifteen years, the theory of radical underspecification. In a nutshell, radical underspecification proposes that only one value of every distinctive feature is specified at the deepest level of the phonology.

Radical underspecification forms an integral part of the theory of lexical phonology and morphology (Kiparsky 1982, 1985; Mohanan 1982). Lexical phonology addresses mainly two issues, namely (i) how the lexicon is organized, and (ii) what the place and role of lexical redundancy rules in a grammar are. Of these, only the second is of interest in the present context. Kiparsky (1982) proposed that

phonological rules are subject to the Strict Cycle Condition (Chomsky 1973, Mascaro 1976) which, in informal terms, states that phonological rules apply in a feature-filling fashion in non-derived environments, while changing features freely in derived environments, typically across a morpheme boundary, although even a tautomorphemic string may count as "derived" as long as it has undergone an appropriate prior phonological rule. The Strict Cycle Condition allows phonological rules to serve as feature-filling redundancy rules morpheme-externally, but as feature-changing phonological rules across a morpheme-boundary. Kiparsky thus eliminates the distinction between lexical redundancy rules and phonological rules, and instead proposes a new kind of rule, the lexical phonological rule, which fills both functions. Lexical phonological rules can supply feature specifications in non-derived environments, because Kiparsky maintains that only one value of every feature ([a F], let us say) can be specified in a given context at the deepest level of the phonology, while its complementary value ([-'a F]) is filled in by the phonological rules of a language.

Radical underspecification has also been explored as a theory in its own right, initially by Archangeli (1984, 1988), Pulleyblank (1988a,b) and Archangeli and Pulleyblank (1986, 1989), and later by many others. The radical position contrasts with that taken by Steriade (1987) and Clements (1988), among others, who assume that both values of a feature are lexically specified if that feature is distinctive, a position that is known as contrastive underspecification.

In its most extreme form, the assumption that only one feature value is present underlyingly has given rise to the assumption that some (if not all) features are privative, a position defended most explicitly by Steriade (1994). A similar view is also taken by phonologists working within dependency phonology (Lass 1987, Anderson and Ewen 1987, and others) who propose a different set of features than traditionally assumed since SPE, however. It has also been influential in the development of feature geometry, which, as we noted in the discussion above, assumes that articulator nodes like labial, coronal, and dorsal are inherently privative or monovalent; i.e., the presence of an articulator node such as coronal contrasts with its absence at the surface.

5. The theoretical basis of underspecification theory

Current phonological theory recognizes two sources of underspecification: (a) lexical minimality and (ii) markedness.

Lexical minimality is rooted in the assumption that mental storage space is costly, so that lexical representations contain only those features that are necessary to distinguish the formatives of a language from one another. All feature specifications that are predictable either through the syntagmatic or paradigmatic context are absent from lexical representations and provided by redundancy rules, the morpheme structure rules, which convert "blanks" or 0 specifications into either a positive or a negative specifications for some feature F. A case of syntagmatic redundancy is found in Russian obstruent clusters, which always agree in voicing: only the second obstruent in a tautomorphemic cluster needs to be lexically specified for the feature [voice], while the specification of the first is predictable and can be filled in by a sequence structure rule. A typical example of paradigmatic redundancy is voicing in sonorants. Sonorants in Russian, as well as in many other languages, do not contrast in voicing; the feature [voice] can thus be left unspecified in sonorants and be filled in by a segment structure rule.

It must be noted that the predictability of [+voice] in sonorants renders predictable also the feature specification [-sonorant] in segments that are specified as [-voice]. Rather than leaving out the feature specification [-sonorant] in voiceless segments, it has become common practice in phonological theory.
to omit the [-voice] entry on obstruents. Even in those positions in which voicing is contrastive, and a generalization of this view has come to be known as radical underspecification (Kiparsky 1982, 1985, Archangeli 1984, Archangeli and Pulleyblank 1986). Radical underspecification theory assumes that only one value of every distinctive feature is part of the lexical representation, an assumption that is ultimately motivated by the desire to free the lexical level of as much information as possible. What motivates the choice of unspecified feature? The literature points in three general directions for an answer to this question: (a) the choice of unspecified feature values is determined by context-dependent universal markedness considerations; (b) the choice of unspecified feature values is determined by context-free markedness considerations; (c) the choice of unspecified feature values is determined by language-specific considerations.

With few exceptions, Chomsky and Halle (1968) consider markedness a context-dependent notion which reflects the cross-linguistic frequency with which certain feature values combine. For instance, the most frequently occurring vowel systems have unrounded non-low front vowels and rounded non-low back vowels, while the opposite arrangement is rare by comparison. The value [+round] is thus marked in non-low front vowels, while it is unmarked in non-low back vowels. It follows that [+round] cannot be considered either marked or unmarked without reference to the paradigmatic context in which it occurs. Similarly, while [+voice] must be considered the unmarked value in sonorants cross-linguistically, it is marked in obstruents. The feature [voice] therefore does not have a marked value across all segment types. The concept of markedness in SPE is hence reserved for combinations of features, but not for feature values in isolation.

This view contrasts with the assumption that some, if not all, features have a marked and an unmarked value without reference to any specific context. This view ultimately goes back to Trubetzkoy (1939), who introduces the notion of markedness to distinguish between the two members of a privative opposition. One member is characterized by the presence of a phonetic property, such as lip rounding, nasality or voicing, while the unmarked member lacks this property. On this view, the gesture of lip rounding is marked independent of the context, i.e., the segment type, in which it occurs.

While context-free and context-dependent markedness generalizations sometimes converge on the same segment type as being marked, as for example in the case of nasal segments, more often the two approaches to markedness cannot be reconciled. The feature [+round], for example, would be considered marked if only the physical gesture of lip rounding were taken into consideration, while no such generalization emerges if its cooccurrence with other features is taken into account.

The significance of markedness generalizations lies in their implications for a theory of "naturalness" and consequently the notion of simplicity in generative grammar. The assumption is frequently made that there is a set of default rules which correspond to the markedness generalizations of SPE. Presuming that these default rules form part of Universal Grammar, the unmarked value of every feature can be left out of the underlying representation; a default rule fills these values in free of any cost to the grammar. It must be remembered, however, that from a generative point of view a grammar consists of both a set of rules and the lexical representations of the morphemes of a language. In constructing the simplest grammar possible with a given set of data, language-specific factors must also be taken into account which can be under certain circumstances give rise to what is known as a markedness reversal.

For every word we can determine what the unmarked or expected distribution of its segments is. In English, for example, a vowel is typically short if followed by a syllable with an unstressed non-final vowel. That is, there is a statistical preponderance of words like elephant and enemy over words like

vitamin (American pronunciation, with a long vowel in the first syllable) and bicycle. Kiparsky (1982) argues that the exceptionality of vitamin and bicycle should be encoded by representing their first vowel as [+long] underlingly, as opposed to the regular forms elephant and enemy, words whose first vowel is unspecified for length and which receive a specification for this feature by the lexical rule of Trisyllabic Laxing. Trisyllabic Laxing clearly expresses a language-specific generalization about the naturalness or unmarkedness of English words.

The question we have to consider is why the first vowel in elephant and enemy is unspecified for length underlingly. There are two answers sometimes given to this question: the first is that [-long] is the universally unmarked value in vowels which is thus unspecified underlingly; this interpretation is given by Kaisse and Shaw (1985), for example. The second answer is that [-long] is the unmarked value in vowels which occur in a position before an unstressed, non-wordfinal vowel in English, as evidence by the large number of English words which have a short vowel in this position. The fact that Kiparsky posits a language-specific rule of Trisyllabic Laxing to fill in the missing [-long] value suggests that underspecification of [-long] is motivated by language-particular and not universal considerations.

However, if language-specific generalizations about the syntagmatic distribution of features are captured by leaving the expected, unmarked value out underlingly and supplying it by rule, the sequence structure regularities of some language might make it necessary to specify the universally unmarked feature value in some context, and to supply the universally marked value by a lexical rule. Consider the following examples from German. In German, voiced and voiceless coronal fricatives occur freely in the onset of word-medial syllables, as in reis [Raeı̂z]n "to travel" and reissen [Raeı̂z@n] "to tear". In accordance with universal markedness generalizations we could decide to leave the value [-voice] unspecified underlingly in word-medial fricatives (as in fact in all obstruents) and to have it supplied by the universal redundancy rule [-son] --> [-voice]. In word-initial position, however, voiced coronal fricatives are far more frequent than their voiceless counterparts, which only occur in a few loan words, such as Sartre [SartR] and Satin [Sat@n@ng]. Voiceless fricatives are therefore the unexpected or "marked" segment type word-initially in German, while voiced fricatives are "unmarked." We could hence choose to leave the expected value [+voice] unspecified in coronal fricatives word-initially and posit a rule [+continuant, +coronal] --> [+voice] /#_ to supply the missing feature value. The initial fricatives in Sartre and Satin, by contrast, would be lexically specified for the unexpected value [-voice] and so escape the effect of the redundancy rule. In this particular instance we are dealing with a case of markedness reversal: [+voice] is the unmarked value in this class of obstruents word-initially, while [-voice] is the marked value.

Alternatively, we could choose to specify the universally marked value [+voice] in word-initial coronal fricatives and leave the universally unmarked value [-voice] out in Sartre and Satin. This approach would avoid the problem of syntagmatic markedness reversal. Since there is no evidence from alternation patterns that forces a decision between these two positions, we may try to evaluate these two analyses in terms of their relative complexity. While the first solution is clearly less complex in terms of the number of underlying feature specifications that it requires, we need to posit an additional rule of [+voice] insertion as part of the grammar of German, a step that might outweigh the savings in the lexicon. By contrast, if the second solution is chosen, a greater number of segments are specified underlingly, but no language-specific rule of [+voice] insertion is needed. The balance of the choice involves how important for our treatment of German the following generalization is: the overwhelming majority of German words that begin with a fricative begin with a voiced fricative. If we specified all

voiced fricatives as [+voice] underlyingly -- the second approach -- we miss this important generalization.

Language-particular considerations of feature distribution are also central to the model of underspecification developed by Archangeli (1984) and Archangeli and Pulleyblank (1986, 1989). These authors assume that only one value of every feature is specified underlyingly, regardless of the context in which it occurs, earning their theory the title of context-free radical underspecification, to use Mohanan's (1991) terminology. On their account, the underlyingly specified feature value can be the universally marked value [+F] in one language, but the universally unmarked value [-F] in another. That is, there is no connection between universal markedness generalization, on the one hand, and underspecification, on the other. In fact, these authors explicitly reject the assumption that only universally marked feature values are specified underlyingly, and maintain instead that universal markedness generalizations constitute nothing but options or "preferences" which can be ignored by individual languages. If a language chooses to make use of this possibility, a state of "markedness reversal" obtains.

Universally unmarked values are provided by default rules, distinct from complement rules, which supply language-particular unmarked feature values and which are formed on the basis of the underlying feature specifications by a mechanism called "complement rule formation." According to Archangeli, these complement rules do not add to the cost of a grammar. If universal markedness generalization do not constitute absolute guidelines as to which feature values are present and which values are absent underlyingly, the question arises as to how two alternative analyses of a given set of data are to be compared and evaluated.

Consider, for example, the standard five-vowel system /i, e, a, o, u/: there are several possibilities as to what the underlying representation is. First, we could assume that only the universally marked feature values are specified underlyingly, as shown in (5); the default rules in (6) supply the missing feature values.

(5), (6) around here

Or we could make the opposite assumption, namely we could assume that the universally unmarked feature values [+high] and [-back] are specified underlyingly, as shown in (7), and that the missing values [-high] and [+back] are filled in by the complement rules in (8d) and (8e).

(7), (8) around here

The two analyses use exactly the same number of underlying feature specifications. And if complement rules do not add to the cost of the grammar, then the two systems are entirely equal in terms of their formal complexity, raising the question as to how a decision between these two analyses ought to be made.

Archangeli (1984) and Archangeli and Pulleyblank (1986) hold that the workings of a given phonological system are crucial for deciding between any two competing analyses. Context-free radical underspecification assumes that epenthetic segments are entirely unspecified underlyingly and that they receive their feature specifications on the basis of the redundancy rules that operate in a language. The result of this assumption is that the featural make-up of an epenthetic segment provides important clues as to which feature values are present and which values are absent underlyingly in a given language. In Yawelmani Yokuts, for instance, the epenthetic vowel is /i/, suggesting that [+high] is the unmarked

value of the feature [+high] in Yawelmani, inserted by a redundancy rule. In Spanish, by contrast, the epenthetic vowel is /e/; [-high] rather than [+high] must therefore be the unmarked value of the feature [high] in Spanish. Neither [+high] nor [-high] can therefore be considered to be the marked value of the feature [high] cross-linguistically.

In conclusion, neither context-sensitive radical underspecification (Kiparsky 1982) nor context-free radical underspecification (Archangeli (1984), Archangeli and Pulleyblank (1986)) assume that only universally marked feature values are specified underlyingly. Both theories instead allow for the possibility of markedness reversal, either in a particular phonological context, or independently of context. These two theories thus assume that there is no correlation between markedness generalizations, on the one hand, and underspecification, on the other.

6. Conclusion

Both autosegmental theory and underspecification theory pursue issues with a long tradition in phonological theory. Autosegmental phonology has played a major role in what has been called the geometricization of phonological theory during the past twenty years. Underspecification theory's relationship to autosegmental theory is an uncertain one; in certain respects, autosegmental representation lends a naturalness to underspecified representations, while in other respects underspecification's goals appear to be quite independent of those of autosegmental phonology. Both have been important in the treatment of harmony systems as well as local assimilation processes.

John Goldsmith
University of Chicago

Manuela Noske
József Attila University
Szeged, Hungary

Bibliography


Archangeli and Pulleyblank 1986. "The content and structure of phonological representations". University of Arizona and University of Southern California, ms.


---

(1)  \[ \text{ba la mu bon a} \]
\[ \text{H} \quad \checkmark \quad \text{L} \quad \text{H} \quad \text{L} \]

(2) *Wellformedness Condition*

(a) Each vowel is associated with at least one tone.
(b) Each tone is associated with at least one vowel.
(c) Association lines do not cross.

(3) Arabic
\[ \text{d r s} \]
\[ \text{C V C V C} \]
\[ \text{a} \]

(4) Feature geometry

artwork to be redrawn. This rendition is from an article by Douglas Pulleyblank, Feature Geometry and Underspecification, in *Frontiers of Phonology* (ed. Durand and Katamba), Longman's, 1995.
(5) \(i\quad e\quad a\quad o\quad u\)

\begin{align*}
\text{high} & \quad - \\
\text{low} & \quad + \\
\text{back} & \quad + \\
\end{align*}

(6) \begin{align*}
a. \ [+\text{low}] & \rightarrow [-\text{high}] \\
b. \ [+\text{low}] & \rightarrow [+\text{back}] \\
c. \ [ ] & \rightarrow [-\text{low}] \\
d. \ [ ] & \rightarrow [+\text{high}] \\
e. \ [ ] & \rightarrow [-\text{back}] \\
\end{align*}

(7) \(i\quad e\quad a\quad o\quad u\)

\begin{align*}
\text{high} & \quad + \\
\text{low} & \quad + \\
\text{back} & \quad - \\
\end{align*}

(8) \begin{align*}
a. \ [+\text{low}] & \rightarrow [-\text{high}] \\
b. \ [+\text{low}] & \rightarrow [+\text{back}] \\
c. \ [ ] & \rightarrow [-\text{low}] \\
d. \ [ ] & \rightarrow [-\text{high}] \\
e. \ [ ] & \rightarrow [+\text{back}] \\
\end{align*}