PROBLEM BOOK IN PHONOLOGY:

A Workbook for Introductory Courses in Linguistics and in Modern Phonology

Morris Halle and G. N. Clements

A Bradford Book

The MIT Press
Cambridge, Massachusetts
and London, England

Contents

Introduction

1. The Linguistic Basis of Phonetics

To those who have not thought much about it, the human ability to produce and perceive speech may not seem particularly remarkable. Indeed, we speak our native language with little conscious effort and are normally unaware of the chain of events that must be set into motion when we engage in a conversation with another speaker of our language. It is only when we begin to try to understand in detail how this everyday human activity is possible that we begin to appreciate its complexity. Phonetics is the study of one aspect of human language: its physical realization. It is concerned, then, with the way in which speech is produced by the vocal mechanism (articulatory phonetics), the physical properties of the speech sounds produced (acoustic phonetics), and the way in which these sounds are perceived by the listener (psychoacoustics).

In studying speech sounds, the first thing to observe is that we are not studying physical events as such. Rather we are dealing with certain events—physiological, acoustic, psychological—just to the extent that we judge them to involve human speech. We are concerned, as phoneticians, not with sounds pure and simple, but with sounds produced and perceived by particular organisms (human beings) who have access to a certain type of knowledge, which we term "knowledge of language." It is for this reason that not all noises produced by human beings are of equal interest to the phonetician. For example, the "ah" we are asked to produce when the doctor examines our throat or the noise we make in blowing out a candle are of little phonetic interest, even though speech sounds which are physically identical to these may happen to occur in language.

The best way to convince ourselves of the central role played by linguistic knowledge in the production and perception of speech is to consider the way we perceive spoken words. Our reaction to a speech event depends crucially on our linguistic experience. If we, as speakers of English, hear another English speaker produce the sequence of sounds that the phonetician symbolizes as ['p^{h}æs ðə'ʃugər], we will not "hear" these sounds as such but will "hear" the three word sequence *pass the sugar*. And yet there is nothing in the speech event itself that signals the end of one word and the beginning of another. It is only by virtue of our knowledge of English that we are able to "parse" this sequence of noises into recognizable English words. When the identical acoustic signal is presented to someone who has no knowledge of English, he can neither identify the words nor even tell us how many words are contained in the utterance.

The perception of intelligible speech is thus determined only in part by the physical signal that strikes our ears. Of equal significance to perception is the contribution made by the perceiver's knowledge of the language in which the utterance is framed. Acts of perception that heavily depend on active contributions from the perceiver's mind are often described as illusions, and the perception of intelligible speech seems to us to qualify for this description. A central problem of phonetics and phonology is then to provide a scientific characterization of this illusion which is at the heart of all human existence.

Since knowledge of a language is crucial for the perception and production of speech, it is natural to inquire into the character of this knowledge. Because of the central role that is played by the knowledge of words, it might be thought that our knowledge of a language consists of nothing more than a memorized list of words. That this can't be the case, that our knowledge of a language goes much beyond a memorized list of words, is shown by the following simple considerations. In (1) below we have given a small list of pseudo-words:

(1) sprash, sdrut, strup, skrig, sflick, sblish, sknap, splim

If a group of English speakers were asked to rank these words on a scale ranging from those which could easily be adopted into English to those which most certainly could not, it is very likely that *sprash, strup, skrig,* and *splim* would rank high, while *sdrut, sflick, sblish,* and *sknap* would rank low. These responses on the part of English speakers show that our knowledge of the language cannot be limited to knowledge of a list of words. Since the pseudo-words in (1) were made up expressly for this "experiment," our subjects could not have memorized them. Their judgments must, therefore, be based on something else than a memorized list of words. We get some insight into the nature of this "something else" when we notice that the difference between the pseudo-words that are judged to be like English and those judged to be unlike English consists in the consonant sequences with which the words begin. Specifically, speakers judge that [spr, str, skr, spl, skw, sky] are admissible onsets of English words on the basis of their familiarity with words like the following:

(2) spring string scrimp splint squint skew

On the other hand, clusters such as [sdr, sbl, skn] etc. are inadmissible onsets in English words.

It is, of course, not plausible to suggest that when we were taught English we were made to memorize the list of all admissible onsets of English words. None of our subjects is likely to recall any childhood experience of this sort, nor are many parents likely to recall discussing with their children what consonant sequences may or may not begin English words. But from the fact that we have not been taught explicitly the list of admissible onsets of English, we may not conclude that we have also not learned them. Although teachers do not like to dwell on it for obvious reasons, there can be little doubt that most of what any normal person knows she or he has learned without being taught, and this is especially true with regard to knowledge of language, as any course in linguistics ought to make amply clear.

Having granted that some, or even most, of our knowledge of language is acquired without benefit of teaching, we still must explain why we learn the list of word onsets in the course of memorizing the words of the language. Is there any reason for us to focus on these onset lists rather than on a myriad of other facts like the last digit of the telephone number of our five best friends, or the middle initial of American Nobel Prize winners?

The only plausible answer that has ever been suggested is that our memory is so constructed that when we memorize words, we automatically also abstract their structural regularities. We suppose, to be specific, that human storage space for memorizing words is at a premium so that every word must be memorized in a maximally economical form in which redundant (predictable) properties are eliminated. Since the principles governing word-initial consonant clusters in English capture an important redundant property of English words, access to these principles allows speakers to store English words in their maximally economical, redundancy-free form. If we know that three-consonant onsets of English words all begin with [s], we do not need to store in our memory, when learning words such as those in (2), the information that allows us to distinguish [s] from all other consonants of the language. Since we can make this saving in memory storage for every word that begins with these consonants, it is to our obvious advantage to memorize the rule rather than to clutter up our memory with these redundant facts.

Different rules, of course, govern the words of different languages, but there is no language that lacks such principles altogether, that does not place severe constraints on permissible sequences of consonants and vowels in words. Hence it is always to the language learner's advantage to abstract the structural principles determining the shapes of the words because that allows him to store in a maximally efficient way the enormous vocabulary, amounting to many thousands of items, which all normal speakers carry in their heads.

The preceding remarks have an important implication for the form in which we must assume words are stored in our memory. The uninitiated observer might suppose, as a first hypothesis, that storage is in terms of unique acoustic images or Gestalts associated with each individual word, or perhaps in terms of certain habitual patterns of motor activity representing our pronunciation of each word. In this view, learning the sound shape of a given English word would be analogous to the way we learn the sound of the waves on the ocean or the movements involved in tying our shoelaces or putting on an overcoat. This form of memory, however, is incompatible with the results of the pseudo-word experiment described above. The results of that experiment imply that a crucial step in learning the words of a language consists of abstracting the principles that govern the permissible sequences of consonants and vowels in our language. If we memorized words as acoustic images or patterns of motor activity, it is totally mysterious how we could ever develop these principles. This mystery would be cleared up at once if we assumed that we perceive words unlike other physical events, that we perceive them not as acoustic images or as patterns of motor activity but as sequences of discrete speech sounds or *phonemes.*

It might be objected at this point that this claim entails the implausible presupposition that every normal human has a view of language that is identical with that of a person who has command of an alphabetic writing system, for there are obviously millions of perfectly normal speakers of ordinary human languages who are illiterate or who are literate in nonalphabetic writing systems. This objection fails to take into account the fact that we possess knowledge of a great many things without being consciously aware of it. One of the major pursuits of Socrates in Plato's dialogues was to make his interlocutors conscious of the huge body of knowledge that they possessed without being remotely aware of this fact. Whatever else one may think of Plato, he surely succeeded in establishing the existence of unconscious knowledge. In order to write in an alphabetic writing system, an individual must have conscious access to the phoneme sequences by means of which the words have been stored in his memory. The fact that all languages can be written in an alphabetic writing system shows that speakers of all languages can be made aware of this aspect of their linguistic knowledge. It does not mean that speakers must be consciously aware of this aspect of their knowledge in order to understand others or to speak themselves. It only means that we may possess knowledge without being aware of this fact. Indeed one may think of linguistics as the science that attempts to characterize in detail one type of such knowledge, namely that possessed by normal speakers of their mother tongue.

2. Distinctive Features

We have just seen that an essential component of the unconscious knowledge that speakers have of their mother tongue is constituted by representations of words as sequences of discrete phonemes. An important further discovery that has been made about the phonemes is that these are not further indivisible units, but are themselves complexes of attributes or features which recur in all human languages. We turn now, therefore, to an examination of the evidence that supports this view of phonemes as bundles of distinctive features.

The most obvious observation one might make about the process of speaking is that it involves a complicated set of maneuvers using the lips, tongue, and other structures at the upper end of our respiratory and digestive tracts. The production of natural speech requires the precise coordination of these various components in such a way as to attain a series of well-defined articulatory states or target configurations following upon each other in close succession in time. In order to see how these states are achieved, it is necessary to examine the activity of the components one by one.

When we speak, air comes from our lungs and excites the cavities in our nose, mouth, and throat, causing them to vibrate. When the cavities are excited in this way, they emit an audible sound. The nature of the sound that emerges is determined primarily by two factors: the precise manner in which the air in the cavities is excited, and the internal geometry of the cavity itself.

It is very easy to separate the contribution of these two components because we can readily keep one constant and vary the other. For example, when we sing or pronounce the vowel [a] on various pitches, we keep the geometry of our nose, mouth, and throat cavities constant, as we can verify by looking at a mirror. What changes in producing these various pitches is the way we excite these cavities. One way of exciting them in speaking is by forcing a rapid succession of little puffs of air through the vocal cords at the bottom end of our vocal tract. The mechanism by which we do this is quite similar to the mechanism by which we force air rapidly through our lips in order to express the fact that we feel cold. Although we cannot directly see the operation of the vocal cords as the air is forced through them, we can feel its effect by placing our fingers over the Adam's apple as we pronounce the words *I sigh* or *so-so*; we will then feel a slight tingle in our fingers as we pronounce the vowels, but none as we pronounce the consonants. The tingle indicates that English vowels are produced with vocal cord excitation, while its absence in the consonant [s] shows that the vocal cords are at rest in this consonant. Sounds produced with vocal cord vibrations are called *voiced*; sounds produced without vocal cord vibrations are called *voiceless.*

Returning now to the different pitches of the vowel [a], it can be shown that as the rate of the puffs of air passing through the vocal cords increases, the pitch of the vowel goes up, while as the rate decreases, the pitch goes down. Note, however, that as long as we keep the cavity geometry the same we go on producing the same vowel.

It is also possible to keep the manner of excitation constant and vary the cavity geometry. In this case, we produce a sequence of different vowels, all on the same pitch. This can once again be verified by watching the rapid movements of the jaw and lips in a mirror as we pronounce the vowels [i,e,a,o,u] in succession.

We have seen here that in producing a simple speech sound we must control at least two independent factors: the way we excite the cavity and the shape that we choose to give it. It is obvious that these two factors are quite unconnected. The excitation is controlled by the expansion and contraction of the lungs and by various adjustments that we make at the vocal cords. The geometry of the cavity, on the other hand, is controlled by the muscles that move the tongue, the lips, and other structures near them. This *componential* structure of speech sounds is their most striking property.

In addition to vocal cord excitation, there are two other types of excitation in speech. The first of these is turbulence or fricative noise produced by forcing air through a narrow constriction, as for example in producing the initial consonants in *so*, *foe*, and *show*. The second is called plosion and is produced by a sudden switching off and on of the air stream as in the medial consonants of *appear*, *attack*, and *acute*. It is easy to see that fricative noise and plosion are mutually exclusive types of excitation. Since fricative noise is produced only when air flows through the mouth, it is incompatible with plosion, which requires the flow of air to be totally interrupted. By contrast, both fricative noise and plosion can be produced either with or without vocal cord vibration. We thus have another example of the composite character of speech sounds: both fricatives and plosives appear in two varieties, one produced with vocal cord vibrations and the other without:

(3)

	fricatives	plosives
voiced:	v z ž	b d g
voiceless:	f s š	p t k

(Here [ž] indicates the last sound in *rouge* and [š] the first sound in *shoe.*)

The composite structure of speech sounds is not limited to the way we set an air stream into motion but extends equally to the way we control the geometry of our vocal tract. Consider, for example, the way we produce the final consonants in the words: *rub, Rudd, rug*; *rum, run, rung*. The configuration of the tongue, lips, and larynx is exactly the same for each of the pairs *rub* and *rum*, *Rudd* and *run*, and *rug* and *rung*. The only difference between the members of these pairs is that in producing the last sound of the second set of words (*rum, run, rung*), we lower the velum—the mobile wall visible at the back of the mouth when we say "ah," which terminates in the fleshy appendage known as the uvula—so that air from the lungs passes behind it and on up through the nose. This maneuver has the effect of exciting the air in the nasal cavity, a fact that we can quickly establish by placing a finger on the side of the nose while saying these words and prolonging the final sound: in doing so we discover a kind of vibration at the end of each of the words of the second set, but none at the end of the words of the first set. Sounds produced with a lowered velum are called *nasal*; those produced with a raised velum are *nonnasal* or oral.

As ordinary speakers we have no need, of course, to place a finger on the nose of the person with whom we are speaking in order to determine whether or not his nasal cavities are excited, for the obvious reason that when the nasal cavity is excited the acoustic output is modified in a specific way that our auditory system can readily detect. In much the same way we can determine whether or not someone's vocal cords are vibrating while producing a given sound without putting our fingers on the speaker's throat. Our auditory system is so constructed that it tells us whether a sound is voiced or voiceless. The same is also true of the other types of excitation, fricative noise and plosion; their presence or absence in a sound is readily perceived by our auditory system. Thus, the machinery we have for producing speech and for perceiving it operate in tandem. Both the muscles controlling the vocal tract, and the auditory system that analyzes the signal, treat speech sounds not as atomic, further unanalyzable entities but as simultaneous complexes of properties or features. Moreover, to a significant degree the set of properties in the two domains—articulation and perception—overlap. This close match between articulation and perception is quite surprising, for the two systems subserve radically different vital functions. The articulatory muscles are part of the alimentary and respiratory systems which functionally are quite unrelated to the auditory system.

We can easily go further with the decomposition of speech sounds. Consider, for example, the final sounds of the words of the following three sets:

(4) a. lip, rub, leaf, leave, rim

b. lit, lid, tooth, lathe, rice, rise, wren, rich, ridge, rush, rouge

c. back, bag

The final sounds of the first set, which we may symbolize as [p,b,f,v,m], respectively, are produced with a constriction that is formed by the lower lip. We call these sounds *labial*. The final sounds of the second set, symbolized by [t,d,θ,ð,s,z,n,č,ǰ,š,ž], are produced with a constriction formed by raising the blade (extreme front) of the tongue. These sounds are called *coronal*. The final sounds of the third set, symbolized by [k,g], are produced with a constriction formed by raising the body of the tongue. Additional sounds produced with this part of the tongue include the final sounds of *bang*, symbolized by [ŋ], and of the German name *Bach*, symbolized by [x]. These sounds are termed *high* or *velar*.

Furthermore, for the purposes of the present discussion, we must identify the property of *stridency* as illustrated below:

(5) strident sounds: [f,v,s,z,š,ž,]

nonstrident sounds: [θ,ð,x]

As will be noticed, strident sounds are produced by directing the airstream against a secondary obstruction—either the sharp edges of the upper teeth, as in the case of [f,v,s,z], or the alveolar ridge (the hard structure in which the upper teeth are embedded), as in the case of [š,ž].

The last property to be considered here is that which distinguishes the initial consonants in the words in (6a) from those in (6b):

(6) a.	mat	nat	let	rat	wet	yet
b.	bat	tat	get	cat	vet	fat

In producing the set of initial consonants in (6b), the vocal tract is narrowed or totally closed so that air is trapped inside of it and, as a result, the air pressure inside the cavity is greater than that outside. By contrast, when the sounds in (6a) are produced, there is no obstruction to the air flow and no pressure build-up inside the cavity. It is customary to refer to the sounds in (6a) as *sonorants* and to those in (6b) as *nonsonorants* or *obstruents.*

We conclude this section with a list of distinctive features (7). This set of features is sufficient to define and distinguish, one from another, the great majority of the speech sounds used in the languages of the world. Many of the features on this list have been discussed above, but are included here for reference.

(7) **The Articulatory Correlates of the Distinctive Features**

1. **syllabic/nonsyllabic:** [±syl]. Syllabic sounds are those that constitute syllable peaks, nonsyllabic sounds are those that do not. Syllabic sounds are typically more prominent than contiguous nonsyllabic sounds. (Vowels, syllabic consonants vs. glides, nonsyllabic consonants.)

2. **consonantal/nonconsonantal:** [±cons]. Consonantal sounds are produced with a sustained vocal tract constriction at least equal to that required in the production of fricatives; nonconsonantal sounds are produced without such a constriction. (Obstruents, nasals, liquids vs. vowels and glides.)

3. **sonorant/obstruent:** [±son]. Sonorant sounds are produced with a vocal tract configuration sufficiently open that the air pressure inside and outside the mouth is approximately equal. Obstruent sounds are produced with a vocal tract constriction sufficient to increase the air pressure inside the mouth significantly over that of the ambient air. (Vowels, glides, liquids, nasals vs. stops and fricatives.)

4. **coronal/noncoronal:** [±cor]. Coronal sounds are produced by raising the tongue blade toward the teeth or the hard palate; noncoronal sounds are produced without such a gesture. (Dentals, alveolars, palato-alveolars, palatals vs. labials, velars, uvulars, pharyngeals.)

5. **anterior/posterior:** [±ant]. Anterior sounds are produced with a primary constriction at or in front of the alveolar ridge, while posterior sounds are produced with a primary constriction behind the alveolar ridge. (Labials, dentals, alveolars vs. palato-alveolars, palatals, velars, uvulars, pharyngeals.)

6. **labial/nonlabial:** [±lab]. As the term implies, labial sounds are formed with a constriction at the lips, while nonlabial sounds are formed without such a constriction. (Labial consonants, rounded vowels vs. all other sounds.)

7. **distributed/nondistributed:** [±distr]. Distributed sounds are produced with a constriction that extends for a considerable distance along the midsaggital axis of the oral tract; nondistributed sounds are produced with a constriction that extends for only a short distance in this direction. (Sounds produced with the blade or front of the tongue vs. sounds produced with the tip of the tongue. This feature may also distinguish bilabial sounds from labiodental sounds.)

8. **high/nonhigh:** [±high]. High sounds are produced by raising the body of the tongue toward the palate; nonhigh sounds are produced without such a gesture. (Palatals, velars, palatalized and velarized consonants, high vowels and glides vs. all other sounds.)

9. **back/nonback:** [±back]. Back sounds are produced with the tongue body relatively retracted; nonback or front sounds are produced with the tongue body relatively advanced. (Velars, uvulars, pharyngeals, velarized and pharyngealized consonants, central vowels and glides, back vowels and glides vs. all others.)

10. **low/nonlow:** [±low]. Low sounds are produced by drawing the body of the tongue down away from the roof of the mouth; nonlow sounds are produced without such a gesture. (Pharyngeal and pharyngealized consonants, low vowels vs. all others.)

11. **rounded/unrounded:** [±round]. Rounded sounds are produced with protrusion of the lips; unrounded sounds are produced without such protrusion. (Rounded consonants and vowels vs. unrounded consonants and vowels.)

12. **continuant/stop:** [±cont]. Continuants are formed with a vocal tract configuration allowing the airstream to flow through the midsaggital region of the oral tract; stops are produced with a sustained occlusion in this region. (Vowels, glides, *r*-sounds, fricatives vs. nasal and oral stops, laterals.)

13. **lateral/central:** [±lat]. Lateral sounds, the most familiar of which is [l], are produced with the tongue placed in such a way as to prevent the airstream from flowing outward through the center of the mouth, while allowing it to pass over one or both sides of the tongue; central sounds do not involve such a constriction. (Lateral sonorants, fricatives and affricates vs. all other sounds.)

14. **nasal/oral:** [±nas]. Nasal sounds are produced by lowering the velum and allowing the air to pass outward through the nose; oral sounds are produced with the velum raised to prevent the passage of air through the nose. (Nasal stops, nasalized consonants, vowels and glides vs. all other sounds.)

15. **advanced/unadvanced tongue root:** [±ATR]. As its name implies, this feature is implemented by drawing the root of the tongue forward, enlarging the pharyngeal cavity and often raising the tongue body as well; [−ATR] sounds do not involve this gesture. ([+ATR] vowels such as [i,u,e,o] vs. [−ATR] vowels such as [ɪ,ʊ,ɛ,ʌ,a].)

16. **tense/lax:** [±tense]. Tense vowels are produced with a tongue body or tongue root configuration involving a greater degree of constriction than that found in their lax counterparts; this greater degree of constriction is frequently accompanied by greater length. (Tense vowels vs. lax vowels.) We note that this feature and the last (ATR) are not known to cooccur distinctively in any language and may be variant implementations of a single feature category.

17. **strident/nonstrident:** [±strid]. Strident sounds are produced with a complex constriction forcing the airstream to strike two surfaces, producing high-intensity fricative noise; nonstrident sounds are produced without such a constriction. (Sibilants, labiodentals, uvulars vs. all other sounds.) The feature [+strid] is found only in fricatives and affricates.

18. **spread/nonspread glottis:** [±spread]. Spread or aspirated sounds are produced with the vocal cords drawn apart, producing a nonperiodic (noise) component in the acoustic signal; nonspread

or unaspirated sounds are produced without this gesture. (Aspirated consonants, breathy voiced or murmured consonants, voiceless vowels and glides vs. all others.)

19. **constricted/nonconstricted glottis:** [±constr]. Constricted or glottalized sounds are produced with the vocal cords drawn together, preventing normal vocal cord vibration; nonconstricted (nonglottalized) sounds are produced without such a gesture. (Ejectives, implosives, glottalized or laryngealized consonants, vowels and glides vs. all others.)

20. **voiced/voiceless:** [±voiced]. Voiced sounds are produced with a laryngeal configuration permitting periodic vibration of the vocal cords; voiceless sounds lack such periodic vibration. (Voiced vs. voiceless consonants.)

3. Natural Classes of Sounds

In (2) we gave examples of the admissible three-consonant onsets in English words, and we argued at length that English speakers have knowledge of the admissible onsets of their language. We have as yet not stated the principle governing these onsets by virtue of which speakers can distinguish "possible" from "impossible" pseudo-words in a list such as (1). As a first approximation we may say that English words are subject to the limitation that in a three-consonant onset cluster, C_1 must be [s], C_2 must be one of the set [p,t,k], and C_3 must be one of the set [r,l,w,y].

An important aspect of this limitation is that it involves not just random sets of phonemes but sets that share some features in common. Thus, for example, the set [p,t,k] shares the features [– voiced, – continuant], whereas the set [r,l,w,y] shares the features [+sonorant, – nasal]. And we find shared features in all sorts of phonological regularities in all sorts of languages: they generally involve natural classes of phonemes.

To cite one other example where the set of sounds [p,t,k] plays a role in English phonology, we observe that these sounds are pronounced with a special puff of breath or aspiration when they occur word-initially before stress as, for example, in such words as *pill, till, kill.* When these phonemes occur in other environments, they are not aspirated, as, for example, in *spill, still, skill.* In order to see more clearly what is involved in these cases, it is useful to examine the natural classes of phonemes as defined in the feature list (7). Let us assume, as proposed by Roman Jakobson, that phonemes are nothing but bundles or complexes of features and that therefore the only way that we can refer to a phoneme is by listing the features that compose it. When we attempt to follow this procedure, we discover that we need not list for each phoneme all features given in (7), but rather a smaller set of "defining" features. Thus, for example, we may uniquely designate the phoneme [p] by the four features:

(8) [– sonorant, + labial, – voiced, – continuant]

It is not possible to omit any of these features, for if we did we would be identifying not a single sound but a set of sounds. For example, if we omitted the specification of the feature [continuant], we would be identifying the set [p,f]; if we omitted the specification of [voiced], we would identify the set [p,b], and if we omitted the specification of [labial], we would identify the set [p,t,k], i.e., the set that can occupy the middle position in English three-consonant onsets and also is subject to aspiration word initially before stress.

Unlike the sets just reviewed, a set like [p,r] can be identified only by specifying more features than are required for each of the two sounds individually. In order to identify the set [p,r], we would have to specify the disjunction (9):

(9) [− sonorant, + labial, − voiced, − continuant]

or

[+ sonorant, − labial, − nasal, + coronal, − high, + continuant]

which mentions all the distinctive features required to identify [p] and [r] separately. We can now define the difference between "natural" and "unnatural" classes of sounds in the following way: "natural" classes can be specified by a single conjunction of features as in (8); "unnatural" classes require a disjunction for their specification as in (9).

We have already noted that the languages of the world appear overwhelmingly to favor natural sets of sounds in their rules. Translated in terms of feature specifications of the sort just illustrated, this means that the languages of the world prefer to deal with sets of sounds that require few specified features for their identification rather than sets that require many. If we now postulate that the rules and regularities that represent a speaker's knowledge of the phonology of his language are represented in the speaker's memory in terms of distinctive feature specifications, then this observed preference on the part of the languages of the world becomes readily comprehensible: it is but another facet of the need to conserve space in the speaker's memory, which we have already had occasion to invoke in accounting for the results of our "experiment" with the pseudo-words in (1).

4. Phonological Rules and their Interactions

Phonology—as opposed to phonetics—is not primarily concerned with the physical or sensory properties of speech sounds as such, but with the systems of rules that determine their possibilities of combination and their phonetic realization in each language. Students of language have long realized that the sounds of language are governed by rules of various sorts. Some of the greatest achievements in 19th century linguistics, for example, involved the discovery of the effects of such phonological rules or "sound laws" on the course of historical change. Thus, for example, part 1 of Grimm's Law replaces [p,t,k] by [f,θ,x], respectively, while part 2 replaces [b,d,g] by [p,t,k]. However, although the discovery of sound laws was extensively pursued by linguists for well over a century, little attention was paid to the exact nature and psychological status of these laws.

Largely as a result of work carried out since the end of the Second World War, it has become clear that a central component of every speaker's knowledge of language consists of rules of grammar, including phonological rules. As already noted, such rules are not directly accessible to conscious reflection; and most of us become aware of them only under very special circumstances. For example, one of the common difficulties that we encounter in speaking a foreign language is to remember to suppress the phonological rules of our native language. Thus, English-speaking tourists in France will normally betray their origins unmistakably in their pronunciation of words like *Paris* by their non-French aspiration of the initial sound (following, inappropriately in this case, the English aspiration rule), and their insistence on pronouncing the English retroflex [ɽ] instead of the (French) uvular [r] for the medial sound.

As a further example of a phonological rule of English consider the following pairs of words:

(10)	seating	seeding
	writing	riding
	beating	beading
	coating	coding

The majority of speakers of American English, in normal conversational speech, pronounce the words of the first column identically to the words of the second column, producing the orthographic *t* and *d* as a lightly articulated, voiced sound formed with a constriction produced by the tongue tip that we will symbolize as [D] ("tap"). What is interesting is that these words are not pronounced identically if the ending *-ing* is removed: *seat* is distinct from *seed*, *write* is distinct from *ride*, and so forth. We are clearly dealing here with a further rule of English phonology, which causes the phonemes [t] and [d] after a stressed vowel to be realized as a tap whenever another vowel follows in the same word.

This rule of English is particularly instructive, for two reasons. The first is that it shows us that underlying distinctions between phonemes can be merged or neutralized by the operation of regular phonological rules. As a consequence it is not always possible to determine the underlying (phonemic) representation of a word on the basis of its surface (phonetic) representation, for in dialects where *riding* and *writing* are pronounced indistinguishably, it will not be possible to decide for any given utterance of these words whether the last stem consonant is underlyingly [d] or [t].

The second point derives from the observation that not all dialects completely suppress the difference between words with underlying [t], such as *seating*, and words with underlying [d], such as *seeding*. An important phonetic difference between *seat* and *seed* lies in the fact that the vowel of the second word is produced with perceptibly greater length than that of the first. This is due to a regular rule of English according to which a vowel has greater duration before a voiced consonant than before a voiceless consonant. Further examples of this rule include *lap* versus *lab*, *lock* vs. *log*, *bus* vs. *buzz*. For the speakers of the dialect in question, this length distinction is preserved in the words *seating* and *seeding*, even though the difference between the orthographic *t* and *d* has itself been completely neutralized.

The reader who normally uses the tap sound in words like *seating/seeding* may easily ascertain whether he or she belongs to this latter dialect group by performing the following test. Write down the words *seating* and *seeding* on a sheet of paper, in random order with ten examples of each. Ask an English-speaking acquaintance who shares your dialect as nearly as possible to serve as subject. Read the list word by word to your subject, leaving enough time between each word so that the subject can write down the word she or he hears. If you make a consistent difference in length between the vowels of *seating* and *seeding* (and if, moreover, your subject has detected this difference), the subject's test score should be close to 100% correct. If, however, you do not make a distinction (or if your subject has not noticed your distinction), the expected score will be about 50% correct (the result that could be obtained by random guessing.)

How may we account for the difference between speakers of the dialect that maintains a vowel length distinction between *seating* and *seeding* (let us call this dialect A) and speakers of the dialect which preserves no difference between these words (let us call this dialect B)? We know that both dialects share the two phonological rules in question: the tap rule, and the rule which lengthens stressed vowels before voiced consonants. The difference between the dialects must, therefore, be attributed to the way the two rules interact. Whereas in dialect A the rules interact so as to maintain the two forms as phonetically distinct, in dialect B the same two rules interact so as to eliminate the phonetic contrast.

This difference in rule interaction can be captured formally if it is assumed that the rules of the phonology are applied in a specific order, that when a rule applies to a word it modifies the representation of the word, and that the rule ordered *n*th in the sequence of rules applies not to the underlying representation of the word, but rather to the representation created by the rule ordered before it. This type of computation of the surface form or phonetic output has been called a *derivation*. In (11) we illustrate the derivations of the words *seating* and *seeding* in the two dialects A and B. Notice that in dialect A the lengthening rule is ordered before the tap rule, while in dialect B the tap rule is ordered before the lengthening rule (vowel length is indicated by a colon).

(11)

Dialect A

	seating	seeding
Lengthening rule	—	i:
Tap rule	D	D
Output	s[iyD]ing	s[i:yD]ing

Dialect B

	seating	seeding
Tap rule	D	D
Lengthening rule	i:	i:
Output	s[i:yD]ing	s[i:yD]ing

The principle of rule ordering makes it possible to account for a great variety of fascinating phonological phenomena in terms of the interaction of a few simple rules. Rule ordering is, therefore, extensively represented in the problems collected in this book.

5. The Nature of Phonological Representations

We have so far tacitly assumed that phonological representations have the form of linear concatenations of phonemes, in much the same sense that written English consists of concatenations of letters. Notice, however, that written English does not provide a fully adequate analogy for phonological representation, since there is much information of relevance to spoken English that is not preserved in the standard writing system. One such type of information is stress. For example, when the word *convert* is used as a verb, main stress falls on the second syllable (*they will convert to the metrical system*), but when this word is used as a noun, main stress falls on the first syllable (*he is a convert to the metrical system*). This distinction is maintained even when these words are pronounced out of context.

A further aspect of English that is systematically omitted in the linear representations of written language is intonation: the "melody" with which a word, phrase, or sentence is pronounced. For example, if we produce the sentence *the Red Sox lost again* with a "flat" intonation, falling at the end, this sentence is understood as a statement of fact. If, on the other hand, this sentence is produced with a rise in pitch at the end, it is understood as a question, or if the rise is very great, as an expression of incredulity. Nearly all known languages make systematic use of stress and intonation in giving phonetic form to their sentences. Of special interest in this connection are the so-called tone languages where differences in pitch are used to distinguish one word from another. In the Ewe language of Africa, for example, [to] spoken with a high pitch may mean either "mountain" or "ear"; spoken with a rising pitch, it means "mortar," and with a low pitch, "buffalo." To represent such properties of language, writing systems have traditionally been supplemented with a set of diacritic marks. Thus, the Ewe words just discussed are commonly represented in studies of this language as shown in (12):

(12) tó 'mountain', 'ear'; tǒ 'mortar'; tò 'buffalo'

It is possible to think of representations such as those in (12) in two distinct ways. On the one hand, one may view the diacritic marks as a means of extending the letter stock of our alphabet. From this point of view the difference between the symbols *ò* and *ó* is parallel to that between *o* and *e* or *m* and *n*. Alternatively, it is possible to view the diacritic tone marks as representing phonetic entities that are separate but equal to those represented by the consonant and vowel letters. The appropriate analog here is the musical score of a song for a single voice unaccompanied. In the score, the text is represented by a sequence of letters, and the melody by a sequence of notes, the two sequences of symbols running in parallel on two separate lines or tiers.

For reasons to be detailed below, we adopt the latter view. Following Goldsmith (1979), we propose that the Ewe words will be represented as in (13), where tones and phonemes are represented on separate *autosegmental* tiers and where association lines link the tones to the phonemes that "bear" or manifest them:

```
(13)  phoneme tier:       to        to        to
                          |         /\        |
      tonal tier:         H         LH        L
```

If tones are represented as a sequence of units on a tier separate and equal to that of the phonemes, then it is not necessary that the units on one tier should coincide one:one with units on the other tier. The fact that the autosegmental notation permits a one:many relationship between units on two tiers is a strong point in its favor, for it is this relationship that we find in actual languages. As illustrated in the second example of (13), a given vowel may be pronounced with an entire sequence of tones or melody. Moreover, a given tone may be held over any number of vowels or fractions of vowels, as we shall now see.

In the Ngizim language, verbs in the perfective are characterized by the tone melody LH, whereas in the subjunctive, verbs have the melody L if they begin with a syllable of the shape Cə, and the melody H otherwise. The following examples illustrate:

(14)	perfective:	jà kə̀rú	'we stole'
		jà kàasú	'we swept'
	subjunctive:	jà kə̀rì	'that we steal'
		jà káaší	'that we sweep'

In this language, we see that verb tenses are characterized by different tone melodies, depending in part on the character of the first syllable. Using our multitiered notation, we may represent these forms as follows:

```
(15)   kə ru    kaa su     kə ri    kaa ši
       |  |     |   |       \/        \/
       L  H     L   H       L         H
```

We observe that the subjunctive morpheme, consisting of a single H tone or a single L tone, might be regarded as "discontinuous" in the sense that it is realized on discontinuous or nonadjacent parts of the word.

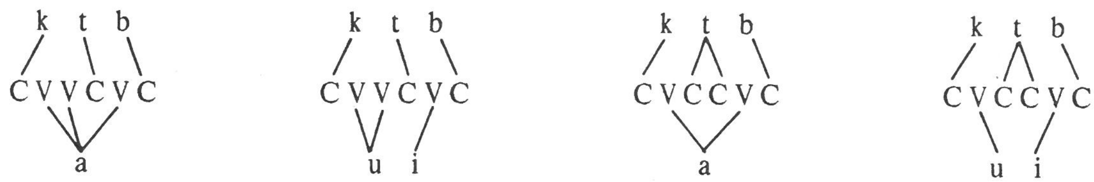
Tone is not the only phonological feature that may appear on a separate tier. Other features that characterize domains smaller or larger than a single segment may also be treated in this way. Indeed, in many languages we find strong evidence that the "skeleton" or canonic shape of a word should be abstracted as an entity distinct from the particular consonants or vowels that characterize it. The evidence in these cases is much the same as that which led us to recognize tonal "melodies" in languages like Ngizim.

Let us consider an example from Semitic, drawn from McCarthy (1981). In most Semitic languages, words are commonly formed from triliteral roots consisting of three consonants, such as [ktb] which has the general notional meaning 'write'. This root serves as the basis for constructing words according to strict canonic patterns. Thus, for example, in Classical Arabic verb roots, the canonic shape CVVCVC is associated with the grammatical meaning 'reciprocal' while the canonic shape CVCCVC is associated with the grammatical meaning 'causative'. Aspect and mood are determined by vowel "melodies" consisting of one to several vowels which extend across the word much as do the tones of Ngizim. Any verb in Classical Arabic, then, can be analyzed into three simultaneous components: the consonantal root, the CV-skeleton, and the vowel melody. From the root [ktb] we can form the following stems, among others:

(16)	CVVCVC:	kaatab	'correspond'	(perfective active)
		kuutib	'correspond'	(perfective passive)
	CVCCVC:	kattab	'cause to write'	(perfective active)
		kuttib	'cause to write'	(perfective passive)

We see that the C-positions in the CV-skeleta are "filled" by the consonants of the particular root (here, the root [ktb]), while the V-positions are occupied by the vowels of the vowel melody, [a] in the case of the perfective active, [ui] in the case of the perfective passive. What emerges from these examples is that whenever there are more C-positions than consonants, or more V-positions than vowels, one of the consonants or vowels "spreads" to occupy the extra position. This is apparent in representations like the following, where each morpheme is assigned to its own tier:

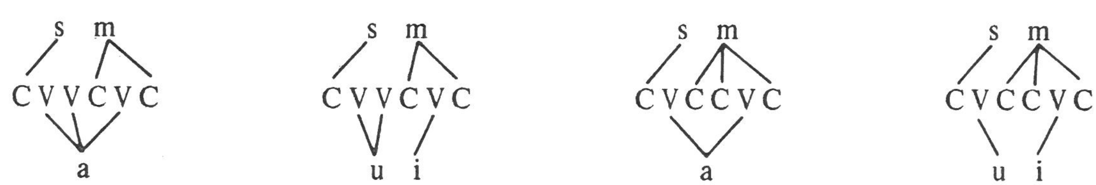
(17)



(kaatab) (kuutib) (kattab) (kuttib)

We see in the case of the perfective active forms that a single vowel may spread over nonadjacent positions in the CV-skeleton, just as in Ngizim a L or H tone melody spreads to all vowels within the word. That the same is true of consonants can be seen most strikingly in the case of roots having only two consonants in their melody, such as [sm] with the general meaning 'poison'. For this root, the verb stems that correspond to those given above for 'write' are *saamam*, *suumim*, *sammam*, and *summim*, respectively. In these words we see that the second consonant of the root melody spreads over all noninitial positions of the CV-tier:

(18)



(saamam) (suumim) (sammam) (summim)

The CV-skeleton, which plays a morphological role in Semitic languages, may play a phonological role in others, accounting for such properties as length, syllabicity, and timing even when these are not associated with specific grammatical meanings (see Clements and Keyser (1981 and forthcoming)). Consider, as an example, the following nouns in Turkish:

(19)

	nom	dat.	'his'	'our'
'room'	oda	odaya	odasɨ	odamɨz
'stalk'	sap	sapa	sapɨ	sapɨmɨz
'mountain'	da:	daa	daɨ	daɨmɨz
'la' (note)	la:	la:ya	la:sɨ	la:mɨz

These forms illustrate various suffixes used in the nominal inflection. We see from examining the first and second examples that the suffixes [-ya] and [-sɨ] are used after nouns ending in vowels, while [-a] and [-ɨ] are used after nouns ending in consonants. Similarly, the choice between [-mɨz] and [-ɨmɨz] depends upon whether the noun ends in a vowel or a consonant in the uninflected form. The third and fourth examples illustrate the fact that there are two kinds of long vowels in Turkish: those that (like the long vowel of *da:*) pattern like consonants in noun inflection, and those that (like the long vowel of *la:*) pattern like vowels. Notice that this patterning is perfectly regular: if a final long vowel patterns like a consonant in one inflected form, it patterns like a consonant in all other inflected forms.

These differences can easily be accounted for if we make two assumptions: first, that length is represented in Turkish just as it is in Arabic, by the association of a single phoneme with two adjacent positions on the CV-skeleton, and second, that a long vowel may be associated with either of the sequences VV or VC. Given these assumptions, we may represent the crucial difference between 'mountain' and 'la' in the following way:

```
(20)    C V C          C V V
        | \/           | \/
        d  a           l  a
```

From these representations it is evident that the rules accounting for the correct form of suffixes need only be sensitive to whether a noun stem ends in a C or a V on the CV-tier.

We see, then, that consonants and vowels, like tones, may "spread" over more than one position in a word. This parallelism between tones, on the one hand, and consonants and vowels, on the other, naturally raises the question of whether more than one nontonal feature may be linked with a single element of the CV-skeleton, just as more than one tone may be linked to a single vowel (see (13) above). In fact, there is good reason to suppose that this is the case. As many phoneticians have pointed out, the initial sounds of English words like *chip* and *job* consist of two components: a stop followed by a fricative, each of which is similar to phonemes that occur independently in English (compare the initial sound of *chip* with the initial sounds of *trip* and *ship*). Nevertheless, we cannot consider these sounds to consist simply of the two independent phonemes [t] and [š], since elsewhere in English no word may begin with a sequence consisting of a stop followed by a fricative (*tfip, ksip, dzip*). One way of expressing the fact that these sounds behave as single phonemes with regard to the distributional rules of English even though they are phonetically complex is to represent the feature "continuant," which distinguishes stops from fricatives, on a separate tier. Given this assumption, the words *ship* and *chip* differ only in respect to their representation on the continuant-tier, not on the remaining tiers:

```
(21)  continuant-tier:   +c +c -c        -c +c +c -c
                          |  |  |          \/   |  |
      CV-tier:            C  V  C          C    V  C
                          |  |  |          |    |  |
      other features:     Š  I  P          Š    I  P
```

(here we use lower-case *c* on the top tier to represent the feature "continuant" and upper case letters on the bottom tier to represent bundles of features not including the feature "continuant"). Other types of complex segments found in the world's languages, such as prenasalized stops, appear to be susceptible to this type of treatment as well.

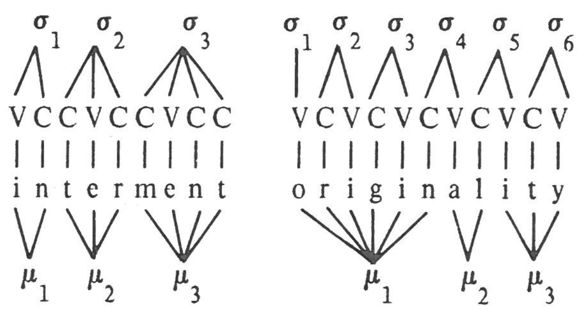
6. The Constituents of the Word

In the preceding section, we have developed the view that a phonological representation is a three-dimensional object, whose core is constituted by the C and V slots of the skeleton around which are disposed tiers consisting of distinctive feature bundles (consonants and vowels) which are linked to different slots in the skeleton. A full phonological representation may thus consist of a CV core surrounded by several autosegmental tiers containing information about such features as tone, nasality, point of articulation, and glottal states; i.e., about the phonetic properties of the C and V slots of the core.

The three-dimensional character of phonological representations makes it possible to solve another long-standing problem, in this case involving the fact that words are organized simultaneously into sequences of morphemes, on the one hand, and sequences of syllables, on the other. Until quite recently linguists assumed that the only way of delimiting sequences within the word was by means of junctures or boundary markers of various kinds. One problem with this device is that it introduces all sorts of additional symbols into the representation that, if taken seriously, make the statement of phonological rules very cumbersome. This problem is illustrated in (22), where "+" represents boundaries between morphemes and "/" represents boundaries between syllables:

(22) +/in+/ter+/ment+/ +/o/ri/gi/n+a/l+i/ty+/

In the word *interment*, as this representation shows, syllable boundaries and morpheme boundaries coincide, while in *originality* the syllable organization is quite unrelated to the morpheme organization. The opacity of these representations is a direct consequence of the fact that they are assumed to consist of feature complexes and juncture marks strung together in a single line. Once we view phonological representations as three-dimensional objects, as explained above, we have at our disposal a much more perspicuous way of dealing with the problem. Instead of delimiting the constituents by means of boundary symbols, we can represent each of the constituents on a separate tier as shown below (σ = syllable, μ = morpheme).

(23)


7. Metrical Structure

All languages have restrictions on what phonemes can combine into sequences, and, as we have seen, speakers can readily tell whether or not a given sequence of phonemes would constitute a well-formed word in their language. As many linguists have noted, the domain over which a great many such constraints hold is the syllable. We have already discussed an example of this. By a rule of English, word-initial consonant clusters consisting of three members must have the form [s]+[p,t,k]+[l,r,w,y]. That this rule is in fact a constraint over syllables rather than over words is shown by the fact that it accounts for word-internal constraints on consonant sequences as well. Thus, while well-formed, existing words like *construct* and *astronomy* can be broken up into sequences of well-formed syllables (con-struct, a-stro-no-my), ill-formed "pseudo-words" like *consknuct* and *apftonomy* cannot. The phonological deviance of the latter can be explained in terms of this fact. Many other rules of English phonology can be shown to involve the syllable in their statement as well.

In addition to rules such as the one just illustrated, which involves restrictions on possible sequences of phonemes within the syllable, there are many rules that are sensitive to distinctions in what has been called *syllable weight*. Syllables that end in a single short vowel are termed *light syllables* and all other syllables are termed *heavy syllables*. We have already seen an example of a language in which syllable weight plays a phonological role: Ngizim. Many other languages are sensitive to this distinction as well. In this section we shall be concerned with examining an important class of rules which are frequently found to be sensitive to syllable weight: stress rules.

In languages making use of stress as a phonological property, each word is characterized by the fact that one syllable is singled out as the bearer of main stress. In the simplest case, this syllable is completely predictable, coinciding with a fixed position in the word counting from the beginning or end. Thus, for example, in Czech or Latvian main stress falls on the first syllable of the word, in Swahili main stress falls on the penultimate syllable, and in Turkish or Farsi main stress falls on the final syllable.

In other languages stress placement, while predictable, is subject to further principles which determine patterns of alternating stress across the word. In Maranungku, main stress falls on the initial syllable, and secondary stress falls on every second syllable thereafter. In Weri, main stress falls on the final syllable, with secondary stress falling on every second syllable before it. Examples are given below:

(24)	Maranungku:	lángkaràtetì	'prawn'	wélepènemànta	'kind of duck'
	Weri:	àkunètepál	'times'	ulǔamít	'mist'

A third factor that frequently enters into the determination of stress placement is, as mentioned above, syllable weight. A simple case is represented by Latin. In this language, main stress fell on the penultimate syllable, if this syllable was heavy, and otherwise on the antepenultimate:

(25)	Latin:	magíster	'teacher'	tábula	'board'
		legú:men	'vegetable'	aurícula	'ear'

The variety of stress systems that we have reviewed so far can be viewed as resulting from a small number of simple rule types which recur in one language after another in slightly different forms. For example, we have seen that there are rules that assign main stress to one end of the word or the other (Czech, Swahili, Turkish); rules that assign secondary stress in an alternating pattern from a fixed starting point (Maranungku, Weri); and rules that assign stress to heavy syllables only (Latin penultimate stress). More complex systems can be described as involving combinations of these rule types.

We shall here offer an informal account of the theory of stress that has come to be developed under the name of *metrical phonology*. While our account will be kept informal, our basic principles can be translated into the formal framework of metrical phonology in a straightforward way. The reader desiring a fuller account of this approach is directed to the work by Liberman, Prince, Hayes, Halle and Vergnaud, and others listed in the references at the end of this introduction.

It will be assumed that the theory of phonology recognizes two distinct ways of concatenating syllables (and other entities, as we shall see directly) into a sequence. One of these is the familiar joining together of entities like beads on a string, where the only relevant property is the position of one unit relative to the others. The second means of concatenating entities consists of setting up one unit as the *head* governing either its immediate neighbor or the entire substring of units on its right or on its left. Graphically we may picture this type of concatenation by means of *trees* such as those illustrated in (26), where the σ's on the bottom row represent the individual syllables in the sequence.

(26)

We have indicated the fact that a syllable is a head by dotting the branch of the tree dominating it. In the trees in (26a) the head governs its immediate neighbor on the right, while in (26b) the head governs the immediate neighbor on the left. Such trees will be called *bounded.* They contrast with the *unbounded* trees in (26c,d), where the head governs the entire sequence of units on its right in (26c), and on its left in (26d).

We shall assume that aligned with each sequence of syllables is a *metrical grid* which is composed of a sequence of slots, one slot for each syllable in the string. A slot may be empty or filled by an asterisk in accordance with the convention (27).

(27) Place an asterisk in the grid slot of a head.

The metrical grid permits us to read off the stresses on the different syllables of the word by the simple expedient of equating degree of stress with the number of asterisks aligned with a particular syllable. This convention accounts for the placement of asterisks in (26) and below.

The diagram in (26c) will be taken here as the formal way of representing stress contours of the words in a language like Czech or Latvian, where each word has only a single stress, which is located on the first syllable. Similarly, the diagram (26d) represents stress in a language like Turkish or Farsi, where the final syllable is the only stressed syllable in the word. Stress assignment in a language like Czech or Latvian will therefore be stated as in (28):

(28) a. Over the syllables of the word, construct a left-headed unbounded tree.

b. Construct the corresponding metrical grid, placing asterisks in conformity with (27).

Notice that (28b) is not an independent rule of the language but the automatic consequence of the decision to represent the stress contours of words by means of metrical grids. The rule for stress assignment in a language with word final stress will be identical to (28), except that the tree will be right-headed rather than left-headed.

We now inquire as to how the proposed framework would express the stress rules of Swahili, where stress falls on the penultimate syllable. The stress contours of Swahili words thus resemble those of Turkish or Farsi, except that the final syllable is systematically left out of consideration. To account for this type of stress system, which is quite common, we make use of the diacritic mark *extrametricality,* which can be assigned only to syllables that are final or initial in the word and which has the property (29):

(29) An extrametrical entity is excluded from metrical trees.

We can now state the stress rules of Swahili as in (30):

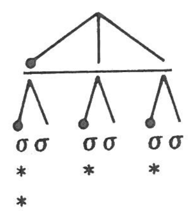
(30) a. The word final syllable is extrametrical.

b. Over the syllables of the word, construct a right-headed unbounded tree.

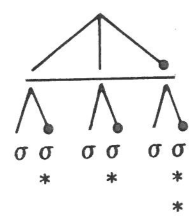
c. (28b)

The trees and grids in (26a) and (26b) represent words with stress on odd-numbered, respectively even-numbered syllables. Alternating stress contours of this type are found in such languages as Maranungku and Weri. They fail, however, to express the fact that two degrees of stress are distinguished in the words of these languages and that the initial, respectively final syllable of the word has greater stress than the rest. We capture this type of stress subordination by constructing a second layer of metrical trees whose bottom termini are not the syllables but the roots of the trees in (26a,b) as shown in (31):

(31) a.



b.



In other words for languages such as Maranungku and Weri we need stress rules such as those in (32) and (33) respectively.

(32) a. Over the syllables of the word, construct left-headed bounded trees.

b. Over the roots of the trees in (a), construct a left-headed unbounded tree.

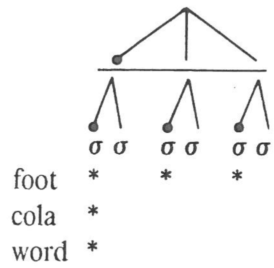
(33) a. Over the syllables of the word, construct right-headed bounded trees.

b. Over the roots of the trees in (a), construct a right-headed unbounded tree.

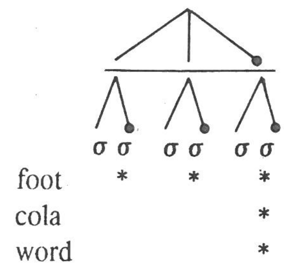
Notice that in conformity with (27) we have placed an additional asterisk in the grid slot corresponding to the head of each new tree.

It turns out that not only may languages distinguish between primary and secondary stress, as do Maranungku and Weri, but that some languages (for example, English; see below) distinguish three levels of stress. This fact is correlated with yet another fact, namely that in the languages of the world there are at most three layers of trees in the metrical structure of words, named respectively from bottom to top: *foot*-layer, *cola*-layer, *word*-layer. To reflect this formally, we shall assume that metrical grids always consist of three rows of slots. If there is only one layer of tree structure, the asterisks are inserted in the word-layer row; if there are two layers of structure, asterisks are placed in the foot-layer and word-layer row, as illustrated in (34) with respect to the trees in (31):

(34) a.



b.



In (34) an asterisk has been placed in the cola row in spite of the fact that in the tree diagram there are only two layers of structure. In the light of (27) a "hole" in the grid should have been expected in the cola row. To fill this "hole" we invoke the special convention (35):

(35) If an asterisk is placed in a particular slot in the metrical grid, the slot on each "inferior" row is automatically filled with an asterisk. [The foot row is "inferior" to the cola row which in turn is "inferior" to the word row.]

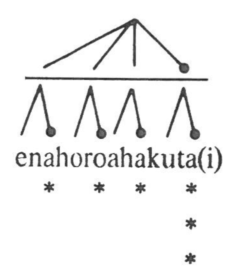
We have so far reviewed stress systems of several relatively simple types and seen how they are to be treated within our formal framework. More complex systems can be described in terms of the interaction of the principles developed above. Consider, for example, a stress system which combines the tree construction rules of (33) with the extrametricality rule of Swahili (30a). Such a system would place main stress on the penult syllable and secondary stress on alternating syllables preceding the penult. This situation is exemplified in Warao, as shown by words like the following:

(36) yàpurùkitàneháse 'verily to climb' enàhoròahàkutái 'the one who caused him to eat'

We derive these patterns by means of the constructions in (37) where the extrametrical syllable is enclosed in parentheses:

(37) a.

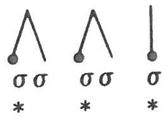
b.

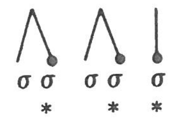


The assignment of stress by means of bounded (binary) feet to a string consisting of an even number of syllables is completely straightforward (see (37b) above). However, when such trees are to be constructed over a string containing an odd number of syllables, a question may arise as to how to deal with the extra odd syllable. Specifically, we may ask whether or not the extra syllable is to be treated as the head of a (single-branch) tree. The fact that in (37a) the initial syllable bears stress, shows that languages treat such leftover syllables as heads.

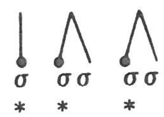
This fact has further consequences of some interest. Up to this point, in our examples left-headed bounded trees were constructed from left to right, and right-headed bounded trees were constructed from right to left. Notice, however, that left-headed bounded trees can also be constructed from right to left, or right-headed trees from left to right. As the example in (38) shows, different kinds of stress contours will emerge depending on the end of the word from which tree construction begins:

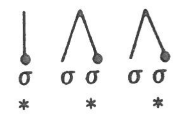
(38) a. left to right tree construction:


(left-headed)


(right-headed)

b. right to left tree construction :


(left-headed)


(right-headed)

We see, thus, that direction of tree construction may differentiate stress patterns of words. The question is, then, whether this option is available to natural languages.

An answer to this question is provided by a comparison of the stress contours of words in Maranungku (see (34a)) in which trees are constructed from left to right, with the stress contours of words in Garawa, in which trees must be constructed from right to left. Thus, we find Garawa examples like those in (39) which illustrate the principles in (40):

(39)

1 punjala	'white'
1 2 watjimpaŋu	'armpit'
1 3 3 2 nařiŋinmukunjinamiřa	'at your own many'

(40) a. Primary stress falls on the initial syllable.

b. Secondary stress falls on the penultimate syllable.

c. Tertiary stress falls on alternating syllables preceding the penultimate, except that

d. There is no stress on the second syllable.

We account for these facts with the help of the rules in (41):

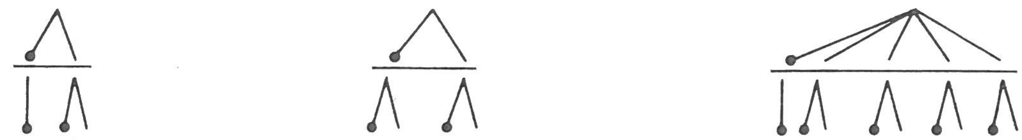
(41) a. Over the syllables of the word, construct bounded left-headed trees starting from the right end of the word.

b. Over the roots of the trees of (a), construct a left-headed unbounded tree.

c. Construct the corresponding metrical grid in conformity with convention (27).

In (42) we illustrate the effects of these rules on the examples quoted in (39):

(42)



```
punjala            watjimpaŋu          nařiŋinmukunjinamiřa
 *  *               *     *             * *    *    *    *
 *                  *                   *
 *                  *                   *
```

We note that the constructions in (42) fail to capture the facts in (40) in two respects. They do not indicate that the penultimate syllable has secondary rather than tertiary stress and they wrongly imply that there is stress on the second syllable. We cure these inadequacies by adding the two rules in (43), of which the first deletes an asterisk in position directly after primary stress, and the second enhances the stresses on the penultimate syllable from tertiary to secondary:

(43) a. Delete an asterisk in position directly after the primary stress.

b. Enhance the final foot-layer stress by adding an asterisk on the row directly below it.

These rules have the effect of changing (42) to (44), which represents the correct output (as the trees remain the same as in (42), they are omitted):

(44)

```
punjala            watjimpaŋu          nařiŋinmukunjinamiřa
 *                  *     *             *      *    *    *
 *                  *     *             *                *
 *                  *                   *
```

To summarize the discussion so far, we have seen that languages construct stress systems of great variety from an extremely small number of rule types. In particular, the following options in selecting rule systems have been identified:

(45) a. Extrametricality rule: present or absent.

b. Number of layers of tree construction: 1, 2, or 3.

c. Type of tree: bounded vs. unbounded; left-headed vs. right-headed.

d. Direction of tree construction (limited to bounded trees): left-to-right vs. right-to-left.

To conclude this discussion of metrical structure, we turn to an examination of the stress system of English. For reasons to become apparent below, we first consider stress in suffixed adjectives. As shown in the examples in (46), in adjectives formed with the suffix *-ic* primary stress falls on the presuffixal syllable:

(46)

Àsiátic	crỳptogámic	mònotónic	phòtográphic
Icelándic	sulphúric	climáctic	magnétic
fanátic	genéric	aquátic	dramátic

By contrast, as shown in (47), before the suffix *-al* primary stress falls on the presuffixal syllable only if it is heavy; if the presuffixal syllable is light, primary stress falls on the antepenultimate syllable:

(47) sùpernâtural | ànecdótal | àccidéntal
binómial | priméval | noctúrnal
evéntual | cerébral | larýngeal

Notice that the stress distribution in (47) follows the same regularities as that of the Latin examples given earlier in (25). We shall assume with Hayes (1981) that the suffixes in the words above (and in adjectives in general) are extrametrical. We assume further that in addition to extrametricality, languages have at their disposal a second diacritic feature, the *accent*, which is used to mark the location of phonologically unpredictable stresses. This feature will be indicated here by underlining the accented syllable. Rules of accent placement apply before tree construction, and trees produced by these rules must incorporate accented syllables as heads. In the case of the English examples in (46) and (47) we shall assume that both suffixes, *-ic* and *-al*, are extrametrical but that they differ in the way they assign accent to their stems.

The suffix *-ic* is special in that it is one of a small number of suffixes that trigger a special rule of accent placement which we state in (48):

(48) Assign accent to the last metrical syllable.

The suffix *-al*, on the other hand, represents the general case of stress in suffixed adjectives. Such adjectives undergo rule (49), which is identical to the Latin Stress Rule. In fact, there is evidence that this rule was incorporated into English as a consequence of the wholesale borrowing of words of Latin origin.

(49) Assign accent to the last metrical syllable if it is heavy.

In addition to accent assignment and extrametricality, English word stress is the result of the tree and grid construction rules (50):

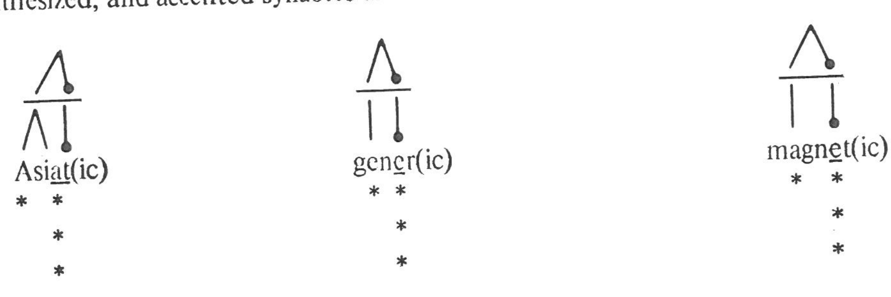
(50) a. Over the syllables of the word, construct left-headed bounded trees starting from the right end of the word.

b. Over the roots of the trees of (a), construct a right-headed unbounded tree.

c. Construct the corresponding metrical grid in conformity with conventions (27) and (35).

We illustrate application of (48—50) to examples with *-ic* below, where extrametrical syllables are parenthesized, and accented syllables are underlined:

(51)



We note that since we are constructing left-headed trees from right to left and the right-most syllable (being accented) must be a head, the right-most foot will always be a nonbranching (degenerate) foot. If the word contains two or more syllables to the left of the accented syllable, the nonbranching foot will be preceded by a branching foot, as in *Asiatic*. If the word has only one additional syllable to the left the prefinal foot will be nonbranching, as in *generic* and *magnetic*.

The grids produced in (51) do not correspond precisely to the phonetic facts as given in dictionaries such as J.S. Kenyon and T.A. Knott, *A Pronouncing Dictionary of American English* (G.C. Merriam Co., Springfield, 1944). Kenyon and Knott distinguish four degrees of stress. For example, in normal pronunciations of words like *hermaphroditic* and *ambassadorial* we find primary stress on the penultimate, respectively antepenultimate syllable, secondary stress on the second syllable, tertiary stress on the first syllable, and no stress on the remaining two syllables. In the Kenyon and Knott system of transcription, primary stress is marked by placing a raised vertical stroke to the left of the syllable in question and secondary stress by a lowered vertical stroke to the left of the syllable. Syllables with tertiary stress are distinguished from those without stress by the presence of a full vowel as opposed to a reduced vowel (schwa [ə] or sometimes [ɪ]). Thus, for example, we find transcriptions such as the following, in which we have added subscript numbers to indicate the degree of stress:

```
(52)  hermaphroditic        hɝ,mæfrə'dɪtɪk
                             3   2     1

      ambassadorial         æm,bæsə'dorɪəl
                            3    2    1
```

When we examine Kenyon and Knott's transcriptions of the three adjectives in (51), we find that the first syllable in *Asiatic* has a secondary stress, that in *generic* has a schwa, while *magnetic* is given with a full vowel without a stress mark (i.e., tertiary stress). Thus, our procedure yields the correct stress contour only in the last of the three words. To obtain the correct stress contours in the other two words, two supplementary rules are needed. First, noting that reduction affects only lax vowels that are unstressed, i.e., have no asterisk, we propose an asterisk deletion rule that is rather similar to the asterisk deletion rule of Garawa (43a):

(53) Delete an asterisk aligned with a light syllable in position directly before the primary stress.

Second, to obtain secondary stress on the first syllable of *Asiatic*, we require a stress enhancement rule resembling the rule proposed earlier for Garawa (43b):

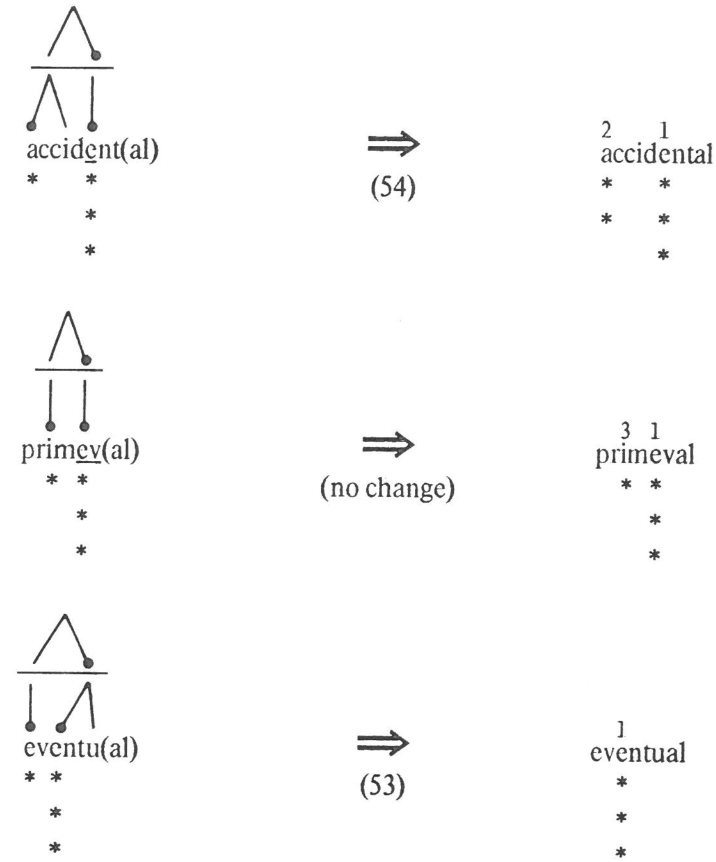
(54) Enhance a foot-layer stress by adding an asterisk on the cola row if the next syllable has no asterisk.

These rules apply to the metrical grids in (51) to produce the following surface representations:

```
(55)  2  1                  1                   3  1
      Asiat(ic)          gener(ic)           magnet(ic)
      *  *                  *                   *  *
      *  *                  *                      *
         *                  *                      *
```

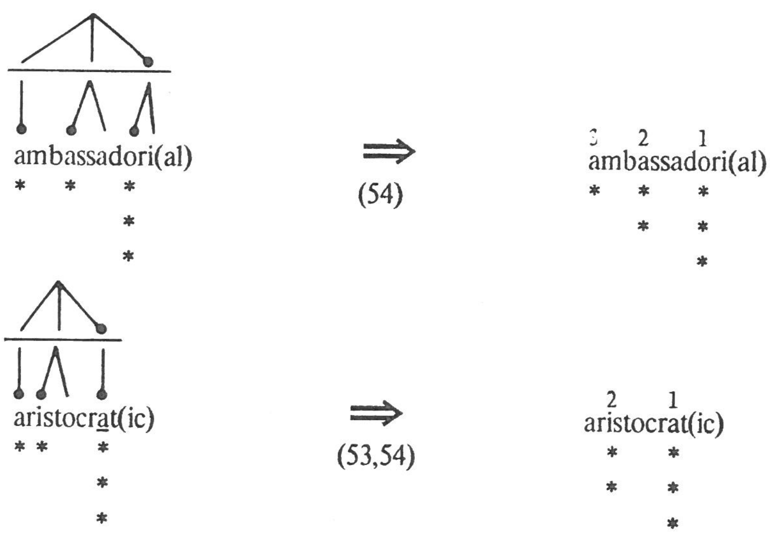
The stress contours of the adjectives in *-al* given in (47) are obtained in precisely the same fashion as those just discussed, except that these words do not trigger the special accent rule (48). Instead, accent is assigned to a heavy penultimate syllable by (49). Otherwise the general tree construction rules (50) apply. Stress deletion (53) and stress enhancement (54) function in exactly the same way as in the earlier examples. We illustrate this below:

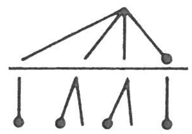
(56)



The set of rules developed to this point, without further modification, assigns the correct stress contours to longer words in *-ic*, and *-al* as well, as shown below:

(57)





```
syncategoremat(ic)          ⟹          3  2  2    1
 *   *   *   *                        syncategoremat(ic)
             *             (54)        *  *  *    *
             *                            *  *    *
                                                  *
```

The rules developed to this point are not limited to stress assignment in adjectives but (excepting rule (48), which is triggered only by a small number of suffixes such as *-ic*) account for the regular pattern of stress assignment in English nouns. We illustrate this result with the following examples. In column A we have placed nouns whose penultimate syllable is light, in columns B and C nouns whose penultimate syllable is heavy. (Note that here as in most English words the final syllable is extrametrical.)

(58)

A	B	C
America 1	Eliza 1	enigma 1
Andromeda 3 1	Pandora 3 1	hydrangea 3 1
incunabula 2 1	Gorgonzola 2 1	influenza 2 1
erotomania 2 1	Monongahela 2 1	extravaganza 2 1
intelligentsia 3 2 1	Ticonderoga 3 2 1	impedimenta 3 2 1
extrasyllabicity 2 2 1	onomatopoeia 2 2 1	counterpropaganda 2 2 1

Further Reading

Sections 1, 2, 3, 4:

Chomsky, Noam, and Morris Halle, *The Sound Pattern of English*, Harper and Row, N. Y., 1968.

Dell, François, *Generative Phonology*, Cambridge University Press, London, 1980.

Kenstowicz, Michael, and Charles Kisseberth, *Generative Phonology: Description and Theory*, Academic Press, N. Y., 1979.

Ladefoged, Peter, *A Course in Phonetics*, second edition, Harcourt Brace Jovanovich, N. Y., 1982.

Zue, Victor W., and M. Laferriere, "An Acoustic Study of Medial /T,D/ in American English," *Journal of the Acoustical Society of America* 66.4, 1039—50, 1979.

Sections 5, 6:

Clements, G. N., and S. J. Keyser, "A Three-tiered Theory of the Syllable," *Occasional Paper* no. 19, Center for Cognitive Science, M.I.T., 1981 (and *CV Phonology* forthcoming).

Goldsmith, John, *Autosegmental Phonology*, I.U.L.C and Garland Publishing, N.Y., 1979.

Halle, Morris, and Jean-Roger Vergnaud, "Three-Dimensional Phonology," *Journal of Linguistic Research* 1.1, 83—105, 1980 (and forthcoming).

Kahn, Daniel, *Syllable-based Generalizations in English Phonology*, I.U.L.C. and Garland Publishing, N.Y., 1980.

McCarthy, John, "A Prosodic Theory of Nonconcatenative Morphology," *Linguistic Inquiry* 12, 373—418, 1981.

Section 7:

Hayes, Bruce, *A Metrical Theory of Stress Rules*, I.U.L.C., 1981.

Halle, Morris, and Jean-Roger Vergnaud, *Three-Dimensional Phonology* (forthcoming).

Liberman, M. Y., and A. S. Prince, "On Stress and Linguistic Rhythm," *Linguistic Inquiry* 8, 249—336, 1977.

Prince, A. S., "Relating to the Grid," *Linguistic Inquiry* 14, 1983.

Selkirk, E. O., "The Role of Prosodic Contours in English Word Stress," *Linguistic Inquiry* 11, 563—605, 1980.