

The Sound Pattern of English
N. Chomsky & M. Halle
Harper-Row, New York, 1968

SETTING

1. Grammar

The goal of the descriptive study of a language is the construction of a grammar. We may think of a language as a set of sentences, each with an ideal phonetic form and an associated intrinsic semantic interpretation. The grammar of the language is the system of rules that specifies this sound-meaning correspondence.

The speaker produces a signal with a certain intended meaning; the hearer receives a signal and attempts to determine what was said and what was intended. The performance of the speaker or hearer is a complex matter that involves many factors. One fundamental factor involved in the speaker-hearer's performance is his knowledge of the grammar that determines an intrinsic connection of sound and meaning for each sentence. We refer to this knowledge—for the most part, obviously, unconscious knowledge—as the speaker-hearer's "competence." Competence, in this sense, is not to be confused with performance. Performance, that is, what the speaker-hearer actually does, is based not only on his knowledge of the language, but on many other factors as well—factors such as memory restrictions, inattention, distraction, nonlinguistic knowledge and beliefs, and so on. We may, if we like, think of the study of competence as the study of the potential performance of an idealized speaker-hearer who is unaffected by such grammatically irrelevant factors.

We use the term "grammar" with a systematic ambiguity. On the one hand, the term refers to [the explicit theory constructed by the linguist and proposed as a description of the speaker's competence.] On the other hand, we use the term to [refer to this competence itself.] The former usage is familiar; the latter, though perhaps less familiar, is equally appropriate. The person who has acquired knowledge of a language has internalized a system of rules that determines sound-meaning connections for indefinitely many sentences. Of course, the person who knows a language perfectly has little or no conscious knowledge of the rules that he uses constantly in speaking or hearing, writing or reading, or internal monologue. It is this system of rules that enables him to produce and interpret sentences that he has never before encountered. It is an important fact, too often overlooked, that in normal, everyday discourse one understands and produces new utterances with no awareness of novelty or innovation, although these normal utterances are similar to those previously produced or encountered only in that they are formed and interpreted by the same grammar, the same internalized system of rules. It is important to emphasize that

there is no significant sense of "generalization" in which these new utterances can be described as generalizations from earlier experience, and no sense of the term "habit" in which the normal use of language can be described as some kind of "habit system" or as "habitual behavior." We cannot, in other words, characterize the internalized, mentally represented system of rules that we call the "grammar" in terms of any other significant concept of psychology.

To summarize, then, we use the term "grammar" to refer both to the system of rules represented in the mind of the speaker-hearer, a system which is normally acquired in early childhood and used in the production and interpretation of utterances, and to the theory that the linguist constructs as a hypothesis concerning the actual internalized grammar of the speaker-hearer. No confusion should result from this standard usage if the distinction is kept in mind.

2. Linguistic universals

General linguistics attempts to develop a theory of natural language as such, a system of hypotheses concerning the essential properties of any human language. These properties determine the class of possible natural languages and the class of potential grammars for some human language. The essential properties of natural language are often referred to as "linguistic universals." Certain apparent linguistic universals may be the result merely of historical accident. For example, if only inhabitants of Tasmania survive a future war, it might be a property of all then existing languages that pitch is not used to differentiate lexical items. Accidental universals of this sort are of no importance for general linguistics, which attempts rather to characterize the range of possible human languages. The significant linguistic universals are those that must be assumed to be available to the child learning a language as an a priori, innate endowment. That there must be a rich system of a priori properties—of essential linguistic universals—is fairly obvious from the following empirical observations: Every normal child acquires an extremely intricate and abstract grammar, the properties of which are much underdetermined by the available data. This takes place with great speed, under conditions that are far from ideal, and there is little significant variation among children who may differ greatly in intelligence and experience. The search for essential linguistic universals is, in effect, the study of the a priori *faculté de langage* that makes language acquisition possible under the given conditions of time and access to data.

It is useful to divide linguistic universals roughly into two categories. There are, first of all, certain "formal universals" that determine the structure of grammars and the form and organization of rules. In addition, there are "substantive universals" that define the sets of elements that may figure in particular grammars. For example, the theory of transformational generative grammar proposes certain formal universals regarding the kinds of rules that can appear in a grammar, the kinds of structures on which they may operate, and the ordering conditions under which these rules may apply. We shall study these questions in detail, in connection with the phonological component of a generative grammar. Similarly, general linguistic theory might propose, as substantive universals, that the lexical items of any language are assigned to fixed categories such as noun, verb, and adjective, and that phonetic transcriptions must make use of a particular, fixed set of phonetic features. The latter topic, once again, will occupy us in this book. We will be concerned with the theory of "universal phonetics," that part of general linguistics that specifies the class of "possible phonetic representations" of sentences by determining the universal set of pho-

netic features and the conditions on their possible combinations. The phonetic form of each sentence in each language is drawn from this class of possible phonetic representations.

3. *Phonetic representations*

What exactly is a phonetic representation? Suppose that universal phonetics establishes that utterances are sequences of discrete segments, that segments are complexes of a particular set of phonetic features, and that the simultaneous and sequential combinations of these features are subject to a set of specific constraints. For example, universal phonetics may provide us with the feature "consonantal," which distinguishes [+consonantal] phonetic segments such as [p], [t], [θ], [s], [ʃ] from [-consonantal] phonetic segments such as [u], [i], [a]; and the feature "strident," which distinguishes [+strident] segments such as [s] and [ʃ] from [-strident] segments such as [p], [t], and [θ]. Among the "simultaneous constraints" of universal phonetics would be the condition that no phonetic segment can be both [-consonantal] and [+strident]; the feature "strident" does not provide a further classification of the category of [-consonantal] segments. Among the "sequential constraints" might be certain conditions that assign a maximal length to a sequence of [+consonantal] phonetic segments, that is, to a consonant cluster. There will be many other constraints of both sorts, and they must be met by each phonetic representation in each language.

More specifically, a phonetic representation has the form of a two-dimensional matrix in which the rows stand for particular phonetic features; the columns stand for the consecutive segments of the utterance generated; and the entries in the matrix determine the status of each segment with respect to the features. In a full phonetic representation, an entry might represent the degree of intensity with which a given feature is present in a particular segment; thus, instead of simply subdividing segments into [+strident] and [-strident], as in the example just given, the entries in the row corresponding to the feature "strident" might indicate degrees along a differentiated scale of "stridency." The phonetic symbols [p], [t], [θ], [i], [u], etc., are simply informal abbreviations for certain feature complexes; each such symbol, then, stands for a column of a matrix of the sort just described.

To recapitulate, the phonetic representation of an utterance in a given language is a matrix with rows labeled by features of universal phonetics. The grammar of the language assigns to this phonetic representation a "structural description" that indicates how it is to be interpreted, ideally, in this language. More generally, we may say that the grammar of each language assigns a structural description to each member of the universal class of possible phonetic representations. For example, the grammar of every language will assign structural descriptions to phonetic representations such as (1) and (2):¹

- (1) ilvyēdradēmē ("il viendra demain")
- (2) hiylkλm+ tōmarō ("he'll come tomorrow")

¹ We omit much phonetic detail that should be specified in universal representations but that is irrelevant to the exposition here. This is the course we will generally follow in discussing particular examples. In the representation (2), and in other representations in this chapter, we include the "boundary symbol" +, which can be taken as specifying a certain type of transition between phonetic elements. Actually, however, we will suggest later that boundary symbols do not appear in phonetic representations.

The grammar of English will assign to (1) a structural description indicating that it is not a sentence of English at all, and to (2) a structural description that specifies the elements of which it is composed on the various linguistic levels, the manner of their organization, the interrelations of these abstract representations, and so on. The grammar of French will supply this information for (1), and will designate (2) as a nonsentence. Many elements of the class of possible phonetic representations will be designated as "semi-grammatical sentences," not well-formed but nevertheless interpretable by analogy to well-formed sentences in ways that are, for the moment, not well understood.²

4. Components of a grammar

The class of possible phonetic representations is of course infinite. Similarly, the class of phonetic representations designated as well-formed sentences in each human language is infinite. No human language has a limit on the number of sentences that are properly formed and that receive a semantic interpretation in accordance with the rules of this language. However, the grammar of each language must obviously be a finite object, realized physically in a finite human brain. Therefore, one component of the grammar must have a recursive property; it must contain certain rules that can be applied indefinitely often, in new arrangements and combinations, in the generation (specification) of structural descriptions of sentences. Every language, in particular, contains processes that permit a sentence to be embedded within another sentence, as the English sentence *John left* is embedded in the sentence *I was surprised that John left*. These processes can apply indefinitely often to form sentences of arbitrary complexity. For example, the sentence *I was surprised that John left* can itself be embedded in the context *Bill expected —*, giving, finally, *Bill expected me to be surprised that John left*, after various obligatory modifications have taken place. There is no limit to the number of applications of such processes; with each further application, we derive a well-formed sentence with a definite phonetic and semantic interpretation.

The part of a grammar which has this recursive property is the "syntactic component," the exact form of which will not concern us here.³ We will, however, make certain assumptions about the abstract objects generated by the syntactic component, that is, about the "syntactic descriptions" that can be formed by the application of its rules.

The syntactic component of a grammar assigns to each sentence a "surface structure" that fully determines the phonetic form of the sentence. It also assigns a far more abstract "deep structure" which underlies and partially determines the surface structure but is otherwise irrelevant to phonetic interpretation, though it is of fundamental significance for semantic interpretation. It is important to bear in mind that deep structures are very different from the surface structures to which we will restrict our attention and that they provide a great deal of information not represented in surface structures.

To recapitulate, a grammar contains a syntactic component which is a finite system of rules generating an infinite number of syntactic descriptions of sentences. Each such syntactic description contains a deep structure and a surface structure that is partially determined by the deep structure that underlies it. The semantic component of the grammar

² For discussion of this matter, which we will exclude from consideration henceforth, see Section IV of Fodor and Katz (1964), and pages 148 ff. of Chomsky (1965), as well as many other references.

³ For recent discussion, see Katz and Postal (1964) and Chomsky (1965).

is a system of rules that assigns a semantic interpretation to each syntactic description, making essential reference to the deep structure and possibly taking into account certain aspects of surface structure as well. The phonological component of the grammar assigns a phonetic interpretation to the syntactic description, making reference only to properties of the surface structure, so far as we know. The structural description assigned to a sentence by the grammar consists of its full syntactic description, as well as the associated semantic and phonetic representations. Thus the grammar generates an infinite number of sentences, each of which has a phonetic and semantic representation; it defines an infinite sound-meaning correspondence, this correspondence being mediated by the abstract syntactic component and the structures it generates.

We are not concerned here with deep structures and the rules that generate them, the rules that relate them to surface structures, or the rules that assign semantic interpretations to syntactic descriptions. We are limiting our attention to surface structures, phonetic representations, and the rules that assign a phonetic representation (possibly several phonetic representations, in the case of free variation) to each surface structure.

5. Surface structures

The surface structures generated by the syntactic component have the following characteristics. Each consists of a string of minimal elements that we will call "formatives." Each formative is assigned to various categories that determine its abstract underlying form, the syntactic functions it can fulfill, and its semantic properties. For example, the formative *boy* will belong to the category of elements with initial voiced stops,⁴ to the category "noun," to the category "animate," to the category "male," etc. This information about formatives will be presented in a "lexicon," which forms part of the syntactic component of the grammar. The organization of the lexicon will not concern us here; we simply assume that the full categorization of each formative is represented in the surface structure. In fact, we may think of the lexical entry of a formative as nothing other than a list of the categories to which it belongs. The categories are sometimes called "features." We will refer, as we proceed, to phonological, syntactic, and semantic features.

The surface structure must indicate how the string of formatives it contains is subdivided into "phrases," each phrase being a certain continuous substring of the string of formatives. The analysis of strings into phrases is a "proper bracketing" in the sense that phrases can overlap only if one is contained in the other. Thus, if *A*, *B*, *C* are formatives, the surface structure of the string *ABC* cannot specify *AB* as a phrase and *BC* as a phrase, for the string may be bracketed either as ((*AB*)*C*) or as (*A*(*BC*)) but not in both ways simultaneously.

The phrases furthermore are assigned to certain categories, and this information may be represented by putting labels on the brackets. Take, for example, the sentence (3):

(3) *we established telegraphic communication*

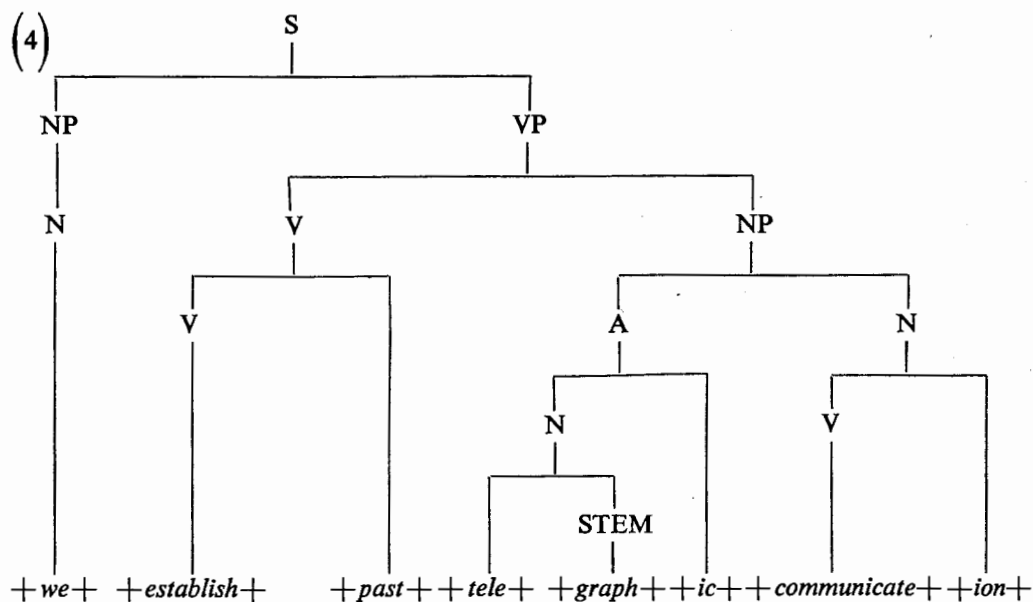
In (3), the string underlying *we* is assigned to the same category as the string underlying

⁴ This underlying representation will be abstract in a sense that we will later describe in detail. For example, although the formative *boy* is always represented phonetically with a back vowel, we will present evidence showing that it should be represented in surface structure—that is, before the phonological rules apply—with a front vowel.

telegraphic communication, namely, to the category "noun phrase." Similarly, the other phrases are assigned to certain universal categories.

We will make the empirical assumption that the surface structure of a sentence is precisely a proper bracketing of a string of formatives, with the bracketed substrings (the phrases) assigned to categories selected from a certain fixed universal set of categories. The complete string is assigned to the category "sentence" (S); the other phrases are also assigned to categories that are provided by general linguistic theory, such as the categories "noun phrase" (NP) and "verb phrase" (VP). These universal categories are on a par with the phonetic categories (bilabial closure, frontness, etc.) provided by universal phonetic theory. As we noted earlier, the categories of universal phonetic theory determine a certain infinite class of possible phonetic representations from which the phonetic forms of sentences of any human language are drawn. Similarly, the universal set of phrase categories (NP, VP, etc.), together with the universal lexical categories (noun, verb, adjective) and the universal lexical features that define the class of "possible formatives," provides us with an infinite class of possible surface structures, from which the surface structures of sentences of any particular language are drawn. In other words, general linguistics should provide definitions, in terms independent of any particular language, for the notions "possible phonetic representation" and "possible surface structure." The grammar of each language relates phonetic representations to surface structures in a specific way; and, furthermore, it relates surface structures to deep structures, and, indirectly, to semantic interpretations, in ways that are beyond the scope of our present study.

To give a concrete example, the grammar of English might assign to the sentence (3) a surface structure which can be represented in the equivalent forms (4) and (5):⁵



(5) [S [NP [N +we+]]NP [VP [V [V +establish+]]V +past+]]NP [A [N +tele+ [STEM +graph+]]STEM]N +ic+]]A [N [V +communicate+]]V +ion+]]N]NP]VP]S

⁵ Once again (see note 1), we omit details which are irrelevant here. We assume, for the purposes of this example, that the formatives are *we*, *establish*, *past*, *tele*, *graph*, *ic*, *communicate*, *ion*. The node labeled A represents the lexical category "adjective"; the other labels have been mentioned previously.

The interpretation of the notational devices used in (4) and (5) should be obvious. We intend these representations to indicate that the formative *we* is both an N and an NP, the formative *establish* a V, the formative string *tele graph* an N, the formative string *tele graph ic communicate ion* an NP, the full string an S, etc.⁶ Furthermore, each formative has an analysis as a set of intersecting categories, in a way that we shall specify in more detail below. The + symbols represent formative boundaries which, by convention, automatically mark the beginning and end of each formative.

5.1. LEXICAL AND PHONOLOGICAL REPRESENTATIONS

To recapitulate, we presuppose, for our description of English sound patterns, a grammar with a syntactic component that assigns to each sentence a surface structure such as (4)-(5), that is, a proper labeled bracketing of a string of formatives. Our main concern here will be the "phonological component," that is, the system of rules that applies to a surface structure and assigns to it a certain phonetic representation drawn from the universal class provided by general linguistic theory. In particular, the phonological rules of English must assign to the surface structure (4)-(5) a phonetic representation much like (6):

(6) wiyəstəbliʃt + tɛləgræfɪk + kəmyʊnəkɛjʃən

The phonetic representation (6), corresponding to the underlying surface structure (4)-(5), is a feature matrix of the sort described earlier. In the surface structure, the individual formatives (for example, the lexical formatives *we*, *establish*, *tele*, *graph*, *communicate*, and the grammatical formatives *past*, *ic*, *ion*) will themselves be represented as feature matrices of an abstract sort, and we must now say a few words about this kind of representation. We shall distinguish between "lexical representations" and "phonological representations." We shall use the term "lexical representation" in reference to formatives which are provided directly by the lexicon, i.e., the lexical formatives as well as certain grammatical formatives which happen to appear in lexical entries. There may be other grammatical formatives introduced directly by the syntactic rules themselves. Thus the syntactic rules and the lexicon, applied in a manner that does not concern us here, provide for each utterance a representation as a string of formatives with surface structure.

Notice, however, that the surface structure must meet two independent conditions: first, it must be appropriate for the rules of phonological interpretation; second, it must be "syntactically motivated," that is, it must result from the application of independently motivated syntactic rules. Thus we have two concepts of surface structure: input to the phonological component and output of the syntactic component. It is an empirical question whether these two concepts coincide. In fact, they do coincide to a very significant degree, but there are also certain discrepancies. These discrepancies, some of which we discuss as we proceed, indicate that the grammar must contain certain rules converting the surface structures generated by the syntactic component into a form appropriate for use by the phonological component. In particular, if a linguistic expression reaches a certain level of complexity, it will be divided into successive parts that we will call "phonological phrases," each of which is a maximal domain for phonological processes. In simple cases the whole sentence is a single phonological phrase; in more complex cases the sentence may be re-analyzed as a sequence of phonological phrases. The analysis into phonological phrases

⁶ Since in representations such as (4) the category labels are placed above the elements in the string that belong to these categories, one frequently speaks of the category as "dominating" a string or a part of a string. Thus, with respect to (4), we will say both that *we* "is an" N and that *we* "is dominated by" N.

depends in part on syntactic structure, but it is not always syntactically motivated in the sense just mentioned. If the syntactic component were to be connected to an orthographic rather than a phonetic output system, the reanalysis into phonological phrases would be unnecessary. Writers, unlike speakers, do not run out of breath, and are not subject to other physiological constraints on output that require an analysis into phonological phrases.

In addition to a reanalysis into phonological phrases in complex cases, the "readjustment rules" relating syntax to phonology make various other modifications in surface structures. It seems that in general these modifications involve elimination of structure, that is, deletion of nodes in representations such as (4) or of paired brackets in representations such as (5). One can easily imagine why this should be so. Reasoning along lines suggested in Miller and Chomsky (1963, Part 2), let us suppose that perception involves a two-stage memory. The first stage is a short-term system quite limited in capacity and operating in real time in the sense that it must remain available for receiving the incoming signal, and the second stage is a very large system that operates on information supplied to it by the short-term real-time system. The short-term first stage must provide an initial analysis of the signal that is just sufficient in detail to permit the second-stage system to derive the deep structure and semantic interpretation. We might expect a language to be so designed that a very superficial analysis into phrases can be performed by a system with limited memory and heavy restrictions on access. To relate this speculation to the discussion of surface structure, it appears that the syntactic component of the grammar generates a surface structure Σ which is converted, by readjustment rules that mark phonological phrases and delete structure, to a still more superficial structure Σ' . The latter then enters the phonological component of the grammar. We might speculate, then, that a first stage of perceptual processing involves the recovery of Σ' from the signal using only the restricted short-term memory, and that a second stage provides the analysis into Σ and the deep structure that underlies it. From this point of view, it would be natural to suppose that the readjustment rules that form Σ' from Σ will have the effect of reducing structure. It is, incidentally, worthy of note that the transformations that form surface structures from deep structures also characteristically have the effect of reducing structure, in a sense which can be made precise.⁷

Let us return now to our discussion of lexical and phonological representations. We have used the term "lexical representation" to refer to the representation of formatives provided by the lexicon. As we have stated, however, the structures generated through the interaction of syntactic and lexical rules are not quite appropriate, in certain cases, for the application of the rules of the phonological component. They must be modified by certain readjustment rules (of a sort to which we will return in Chapter Eight, Section 6.5, noting, however, that our investigation of the effects of surface structure on phonetic representation has not yet reached a level of depth and complexity that requires a detailed, formal analysis of these processes).⁸ These readjustment rules may somewhat modify the labeled bracketing of surface structure. They may also construct new feature matrices for certain strings of lexical and grammatical formatives. To take an obvious example, the verb *sing* will appear in the lexicon as a certain feature matrix, as will the verb *mend*. Using letters of the alphabet as informal abbreviations for certain complexes of features, i.e., certain columns of a feature matrix, we can represent the syntactically generated surface structure underlying the

⁷ See Miller and Chomsky (1963). See also Ross (1967) for further relevant observations of a different sort on reduction of structure under transformations.

⁸ See Bierwisch (1966) for a very interesting study of readjustment rules of the sort mentioned here.

forms *sang* and *mended* as $[_v \text{ [v]sing}]_v \text{ past}]_v$ and $[_v \text{ [v]mend}]_v \text{ past}]_v$, respectively, where *past* is a formative with an abstract feature structure introduced by syntactic rules. The readjustment rules would replace *past* by *d*, as a general rule; but, in the case of *sang*, would delete the item *past* with the associated labeled brackets, and would add to the *i* of *sing* a feature specification indicating that it is subject to a later phonological rule which, among other things, happens to convert *i* to *æ*. Designating this new column as *, the readjustment rules would therefore give the forms $[_v \text{ s}^* \text{ng}]_v$ and $[_v \text{ [v]mend}]_v \text{ d}]_v$, respectively. We shall refer to this representation—and in general to the representation given by the application of all readjustment rules—as the “phonological representation.”

Other terms that might have been used in place of the terms just proposed are “morphophonemic representation” or “systematic phonemic representation.” We have avoided these terms, however, because of the technical meaning they have been given in various theories of sound structure developed in modern linguistics. The term “morphophonemic representation” seems to us appropriate only if there is another linguistically significant level of representation, intermediate in “abstractness” between lexical (phonological) and phonetic and meeting the conditions placed on “phonemic representation” in modern structural linguistics. We feel, however, that the existence of such a level has not been demonstrated and that there are strong reasons to doubt its existence.⁹ We will make no further mention of “phonemic analysis” or “phonemes” in this study and will also avoid terms such as “morphophonemic” which imply the existence of a phonemic level. Notice that the issue in this case is not terminological but rather substantive; the issue is whether the rules of a grammar must be so constrained as to provide, at a certain stage of generation, a system of representation meeting various proposed conditions. The references in note 9 explain our position, and we will say no more about the matter here.

5.2. ON THE ABSTRACTNESS OF LEXICAL REPRESENTATIONS

We have said that the underlying representations, lexical as well as phonological, are abstract as compared with phonetic representations, although both are given in terms of phonetic features. The meaning of this remark will become clearer as we proceed. There is, however, one very obvious sense in which the underlying representations are more abstract than the phonetic representations. Consider, for example, the word *telegraph*. This has several different variants in actual phonetic representations:¹⁰

- (7) $\overset{1}{t} \overset{3}{e} \text{ l} \overset{3}{\text{æ}} \text{ gr} \overset{3}{\text{æ}} \text{ f}^{11}$ (in isolation)
- (8) $\overset{3}{t} \overset{1}{e} \text{ l} \overset{3}{\text{æ}} \text{ gr} \overset{1}{\text{æ}} \text{ f}$ (in the context — *ic*; i.e., *telegraphic*)
- (9) $\text{t} \overset{1}{\text{æ}} \text{ l} \overset{1}{\text{e}} \text{ gr} \overset{1}{\text{æ}} \text{ f}$ (in the context — *y*; i.e., *telegraphy*)

It is quite obvious, however, that this phonetic variation is not fortuitous—it is not of the

⁹ We have presented our reasons for doubting the existence of a phonemic level, in the sense of modern linguistics, in various places. See Halle (1959), Chomsky (1964, 1966b), and Chomsky and Halle (1965), as well as Postal (1962, 1968), for arguments that seem to us fully convincing.

¹⁰ Notice that in the sentence (6) it has still another representation because of the stress modifications that take place in that context.

¹¹ Stress levels are indicated here and throughout by numerals, with “1” representing primary stress, “2” representing secondary stress, etc. (See also note 3 in Chapter Two on this subject.)

same type as the variation between *I* and *we*, which depends on specific assignment of the latter to the category of plurality. Given the grammar of English, if we delete specific reference to the item *we*, there is no way to predict the phonetic form of the plural variant of *I*. On the other hand, the rules for English grammar certainly do suffice to determine the phonetic variation of *telegraph* without specific mention of this lexical item, just as they suffice to predict the regular variation between *cat* and *cats* without specifically mentioning the plural form. It is quite obvious that English grammar is complicated by the fortuitous variation between *I* and *we* but not by the totally predictable variation between *cat* and *cats*. Similarly, the grammar would be more complicated if *telegraph* did *not* undergo precisely the variation in (7)–(9): if, for example, it had one phonetic form in all contexts, or if it had the form (7) in the context — *ic*, (8) in the context — *y*, and (9) in isolation.

In short, the phonetic variation of *telegraph* in certain contexts is not an idiosyncratic property of this particular lexical item but is rather a matter of general rule, applying to many other lexical items as well. Regular variations such as this are not matters for the lexicon, which should contain only idiosyncratic properties of items, properties not predictable by general rule. The lexical entry for *telegraph* must contain just enough information for the rules of English phonology to determine its phonetic form in each context; since the variation is fully determined, the lexical entry must contain no indication of the effect of context on the phonetic form. In fact, as we shall see, the lexical representation for the word *telegraph* should be (10), where each of the symbols *t*, *e*, . . . is to be understood as an informal abbreviation for a certain set of phonological categories (distinctive features):¹²

(10) +tele+græf+

Thus the lexical representation is abstract in a very clear sense; it relates to the signal only indirectly, through the medium of the rules of phonological interpretation that apply to it as determined by its intrinsic abstract representation and the surface structures in which it appears.

An analogous argument can readily be constructed for the abstract nature of the phonological representations, i.e., those representations that are determined from lexical representations by application of certain readjustment rules (and which, for the most part, are in fact identical with lexical representations).

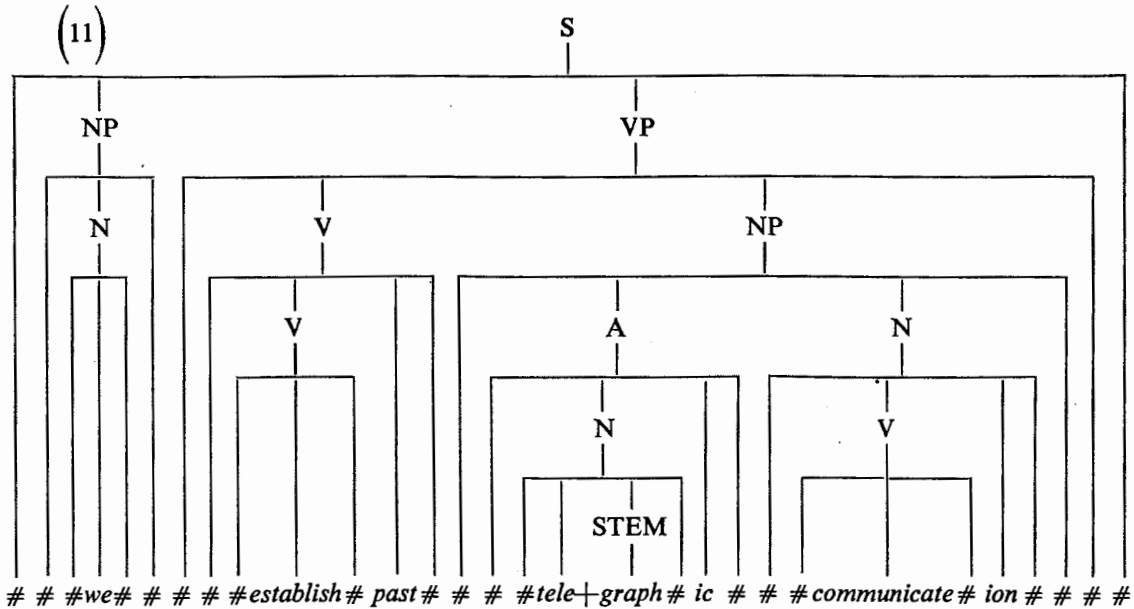
5.3. ANALYSIS INTO WORDS

One additional aspect of surface structure is crucial for our discussion. We will see that the phonological rules fall into two very different classes. Certain of these rules apply freely to phrases of any size, up to the level of the phonological phrase; others apply only to words. We must therefore assume that the surface structure of an utterance provides an analysis into a sequence of words. For example, the sentence (3), *we established telegraphic communication*, will be analyzed by its surface structure into the four successive words *we*, *establish+past*, *tele+graph+ic*, *communicate+ion*. The rules that form surface structure (or, perhaps, the readjustment rules discussed above) must provide this information, since it is required for the correct application of the rules of the phonological component of the grammar.

As a first approximation to the problem of analysis into words, let us assume that each lexical category (e.g., noun, verb, adjective) and each category that dominates a lexical

¹² In addition, the lexical entry will provide the other idiosyncratic syntactic information represented in (4)–(5), namely, the information that *graph* is a stem and *telegraph* is a noun.

category (e.g., sentence, noun phrase, verb phrase) automatically carries a boundary symbol # to the left and to the right of the string that belongs to it (i.e., that it dominates, in tree representations such as (4), or that it brackets, in bracket representations such as (5)). Under this assumption, we replace the representation (4) by (11) and modify (5) in a corresponding way:



Let us tentatively define a word as a string of formatives (one or more) contained in the context # # — # # and containing no occurrences of # #.¹³ Thus the words in (11) are *we*, *establish#past*, *tele+graph#ic*, and *communicate#ion*, as required. The principle just stated can be regarded, tentatively, as a universal principle for the interpretation of surface structures, and, as a first approximation, it works quite well. Among the readjustment rules discussed above, there will be some that modify the representation provided by this interpretive principle in ad hoc ways. For example, we shall see that although the boundary # is appropriate in *establish#ed*, as distinct from ordinary formative boundary (which we have been representing as +), it must be replaced by ordinary formative boundary in *tele+graph#ic* and *communicate#ion*, for reasons having to do with the applicability of certain phonetic rules.

To recapitulate, the rules of syntax will generate surface structures and a universal principle of interpretation will assign the boundary symbol # in certain places. The readjustment rules will modify the surface structure in various ad hoc ways, demarcating it into phonological phrases, eliminating some structure, and replacing some occurrences of # by +. The abstract object thus constructed (which we will also refer to as a "surface structure," or, if more explicitness is necessary, a "phonological surface structure," to contrast it with the syntactic surface structure generated by the syntactic component) enters the phonological component of the grammar and is converted by the phonological rules into a phonetic representation, in ways that we will specify in detail as we proceed. Certain of the phonological rules will apply only to words; others will apply freely to strings of formatives which may be words or subparts of words, or phrases that include words.

¹³ See Chapter Eight, Section 6.2, for a more careful analysis of the notion "word."

We will find it convenient to use labeled bracketing such as (5) rather than tree diagrams such as (4) and (11) for the representation of surface structure in the presentation of phonological rules. Since, by convention, every lexical category or category dominating a lexical category has # boundaries associated with it on the left and right, we will sometimes omit reference to these boundaries in the statement of rules. For example, a rule of the form (12) is to be understood as applying to the string (13):

$$(12) \quad A \rightarrow B / X \text{---} Y]_V$$

$$(13) \quad XAY\#]_V$$

Rule (12) states that an element of the type *A* is rewritten as a corresponding element of the type *B* when *A* appears in the context *X*—*Y* (that is, with *X* to its left and *Y* to its right) and when the item in question is a verb, i.e., is dominated by *V* or, equivalently, is bracketed by $]_V$. We will make these informal specifications more precise as we proceed.

6. Summary

The phonological component is a system of rules such as (12) that relates surface structures such as (11) to phonetic representations such as (6). As we proceed in our discussion, we will propose various specific hypotheses regarding the detailed form of representations such as (11) and (6), and we will also make specific proposals concerning the system of phonological rules that assign a phonetic interpretation to each surface structure.

We have already suggested that a phonetic representation such as (6) is actually a feature matrix in which the rows correspond to a restricted set of universal phonetic categories or features (voicing, nasality, etc.) and the columns to successive segments. We will propose further that such representations are mentally constructed by the speaker and the hearer and underlie their actual performance in speaking and "understanding." We will consider the question of the relation between such phonetic representations and actual speech signals, and the steps by which such representations might be constructed by the hearer on the occasion of reception of a speech signal. We have suggested, moreover, that each formative of the surface structure can also be represented as a feature matrix interpreted in a rather similar way, with rows corresponding to the universal phonetic and grammatical categories. The formative structure is much more abstract, however; its relation to the speech signal is not as direct as that of the phonetic representation.

We will propose that the rules of the phonological component have a fixed form and a specific organization, that they apply in a fixed manner determined by the labeled bracketing of the surface structure, and that they meet various additional conditions depending on their formal relations. These we propose as universal conditions, as aspects of general linguistic theory. We will try to show how, on the basis of these assumptions, many particular phenomena of English sound structure can be explained.

With these remarks on background assumptions, we can proceed to the analysis of English sound structure and of general phonological theory.

THE PHONETIC FRAMEWORK

1. *Phonetic representation*

1.1. *PHONETIC TRANSCRIPTION AND THE SPEECH SIGNAL*

The phonological component expresses the relationship between the surface structure of a sentence and its physical actualization insofar as this relationship is determined by grammatical rule, in the very general sense which we will elucidate below. The surface structure may be represented as a string of formatives, properly bracketed with labeled brackets (see Chapter One, Section 5). Given the surface structure of a sentence, the phonological rules of the language interact with certain universal phonetic constraints to derive all grammatically determined facts about the production and perception of this sentence. These facts are embodied in the "phonetic transcription." Therefore, as P. Postal has remarked, this transcription represents:

the derivative knowledge a speaker has about the pronunciation by virtue of his knowledge of the superficial syntactic structure of the sentence, the lexical items or formatives it contains and the rules of phonology . . . The phonetic transcription . . . is the most gross and superficial aspect of linguistic structure . . . It is the most important but far from the only parameter determining the actual acoustic shape of the tokens of the sentence.

Our conception thus differs from an alternative view that the phonetic transcription is primarily a device for recording facts observed in actual utterances. That the latter view is not tenable, in any very strict sense, has been known at least since mechanical and electrical recordings of utterances have revealed that even the most skillful transcriber is unable to note certain aspects of the signal, while commonly recording in his transcriptions items for which there seems to be no direct warrant in the physical record. But even if the phonetic transcription were as faithful a record of speech as one could desire, there is still some question whether such a record would be of much interest to linguists, who are primarily concerned with the structure of language rather than with the acoustics and physiology of speech. It is because of this question that many structural linguists have felt that phonetics has very little to offer them and have therefore assigned to it a secondary, peripheral role.¹

¹ As an illustration of this lack of interest in phonetics we may cite the numerous articles on phonological subjects that have appeared in the last thirty years in journals such as the *International Journal of American Linguistics* in which information concerning the phonetic properties of the phonemes of a language is often restricted to a simple listing of alphabetic symbols.

See also comments in Chomsky (1964, page 69n and pages 76 f).

These problems do not arise when phonetic transcription is understood in the terms outlined above, that is, not as a direct record of the speech signal, but rather as a representation of what the speaker of a language takes to be the phonetic properties of an utterance, given his hypothesis as to its surface structure and his knowledge of the rules of the phonological component. Since in this view phonetics is concerned with grammatically determined aspects of the signal, there can be no question about the relevance of phonetics to the study of language. Moreover, since the phonetic transcription, in this sense, represents the speaker-hearer's interpretation rather than directly observable properties of the signal, the existence of certain discrepancies between the transcription and the signal can be understood. Thus it is no longer a problem that the transcription is composed of discrete symbols whereas the signal is quasi-continuous, or that the transcription provides information only about some properties of the signal and not about others, or, finally, that physically identical signals may have distinct phonetic transcriptions. Clearly, a person's interpretation of a particular speech event is not determined merely by the physical properties constituting the event. A person will normally not be aware of many properties manifest in the signal, and, at the same time, his interpretation may involve elements which have no direct physical correlates,² since what is perceived depends not only on the physical constitution of the signal but also on the hearer's knowledge of the language as well as on a host of extragrammatical factors.

Implicit in this approach is the view that speech perception is an active process, a process in which the physical stimulus that strikes the hearer's ear is utilized to form hypotheses about the deep structure of the sentence. Given the deep structure and the rules of the grammar, all other representations of the sentence can be derived, including in particular the phonetic transcription, which is the terminal representation generated by the grammar.³ These derived representations are used by the speaker to check his hypothesis against the external stimulus, which provides the data that stand in the most direct (though not necessarily a point-by-point) relationship with the phonetic transcription. Since the hypotheses made in speech perception are highly specific—that is, we understand our interlocutor to have said a particular sentence—they are highly improbable. Consequently even crude agreement between the external stimulus and the internally generated hypothesis suffices to confirm the latter. In other words the dependence of perception on properties physically present in the signal is less than total. What is more, there are many extragrammatical factors that determine how close a fit between data and hypothesis is required for confirmation.

// In the phonetic transcription an utterance is represented as a sequence of discrete units, each of which is a complex of phonetic features such as voicing, nasality, tongue height, etc. The phonetic transcription can therefore be taken to be a two-dimensional matrix in which the columns stand for consecutive units and the rows stand for different features. At this level of representation each feature is to be thought of as a scale. A particular entry in the matrix, then, indicates the position of the unit in question on the given scale. The total set of features is identical with the set of phonetic properties that can in principle

² In fact, we do not wish to exclude the possibility that under certain conditions distinctions that might be implied by the phonological rules of the language may not actually be realizable. This seems particularly to be true in the case of the different degrees of stress predicted by the stress subordination rules discussed in Chapter Three.

³ It is not necessarily the case that each deep structure determines a single phonetic representation; if the grammar contains optional rules or analyses, a given deep structure can underlie two or more phonetic transcriptions.

be controlled in speech; they represent the phonetic capabilities of man and, we would assume, are therefore the same for all languages.

As already noted, phonetic transcriptions consistently disregard many overt physical properties of speech. Among these are phonetic effects that are not locatable in particular segments but rather extend over entire utterances, such as the voice pitch and quality of the speaker and also such socially determined aspects of speech as the normal rate of utterance and what has been called by some writers the "articulation base":

the system of characteristic articulatory movements of a given language that confer upon it its general phonetic aspect; in French the mobility of the lips and forward position of the tongue (Marouzeau, 1943, p. 38).

In addition, phonetic transcriptions omit properties of the signal that are supplied by universal rules. These properties include, for example, the different articulatory gestures and various coarticulation effects—the transition between a vowel and an adjacent consonant, the adjustments in the vocal tract shape made in anticipation of subsequent motions, etc.

1.2. PHONETIC AND PHONOLOGICAL REPRESENTATION

As mentioned above, the phonetic transcription is related by the rules of the phonological component to a string of formatives with labeled bracketing which represents the surface syntactic structure of the sentence. We will now examine in some detail the manner in which these formatives are represented in a linguistic description. Many of the formatives are lexical items, the "roots" or "stems" of traditional grammar. A grammar must include a list of these items, for part of a speaker's knowledge of his language consists of knowing the lexical items of the language. It is by virtue of this knowledge that the native speaker is able to distinguish an utterance in normal English from an utterance such as Carnap's "Pirots karulized elatically" or from Carroll's jabberwocky, which conform to all rules of English but are made up of items that happen not to be included in the lexicon of the language.

The representations of the individual items in the lexicon must incorporate the knowledge which makes it possible for the speaker to utilize each lexical item in grammatically correct sentences. This includes certain syntactic information which the speaker must have. For example, he must know that a particular item is a noun and that it belongs to a large number of intersecting categories such as "animate" or "inanimate," "human" or "non-human," "feminine" or "masculine." Since the only question of interest here is whether or not a given item belongs to the category in question, it is natural to represent this information by means of a binary notation: *cow*, for example, would be specified as [+animate, -human, +feminine]. In addition to these syntactic features, each lexical entry must contain specified features which determine the phonetic form of the item in all contexts. We shall call these the "phonological features." The phonological features cannot be chosen arbitrarily, for the phonological component would then have to include a huge number of ad hoc rules of the type

$$\begin{aligned} [+A, -B, -C, +D] &\rightarrow [hát] \\ [-A, -B, -C, +D] &\rightarrow [rát] \\ [-A, +B, -C, +D] &\rightarrow [əlíps] \end{aligned}$$

Moreover, if we represented lexical items by means of an arbitrary feature notation, we would be effectively prevented from expressing in the grammar the crucial fact that items which have similar phonetic shapes are subject to many of the same rules.

We might consider overcoming these difficulties by representing each lexical item in its phonetic representation. However, this solution is not open to us either, for a lexical item frequently has several phonetic shapes, depending on the context in which the item appears. If we chose to represent each lexical item by the set of its phonetic representations, we would be treating all phonetic variations as exceptions and would, in principle, be unable to express within our grammar the phonetic regularities and general phonological processes that determine phonetic form. If, on the other hand, we chose to allow only a single phonetic representation for each item, then we would have to provide some rationale for our selection. Furthermore, it is easily shown that many of the most general and deep-seated phonological processes cannot be formulated as rules that directly relate phonetic representations; rather, these processes presuppose underlying abstract forms.

We therefore can represent lexical items neither in phonetic transcription nor in an arbitrary notation totally unrelated to the elements of the phonetic transcription. What is needed is a representation that falls between these two extremes. Accordingly we propose that each item in the lexicon be represented as a two-dimensional matrix in which the columns stand for the successive units and the rows are labeled by the names of the individual phonetic features. We specifically allow the rules of the grammar to alter the matrix, by deleting or adding columns (units), by changing the specifications assigned to particular rows (features) in particular columns, or by interchanging the positions of columns. Consequently, the matrix that constitutes the phonetic transcription may differ quite radically from the representation that appeared in the lexicon. There is, however, a cost attached to such alterations, for they require the postulation of rules in the phonological component. Such rules are unnecessary in cases where the lexical representation can be accepted as the phonetic representation. In general, the more abstract the lexical representation, the greater will be the number and complexity of the phonological rules required to map it into a phonetic transcription. We therefore postulate abstract lexical entries only where this cost is more than compensated for by greater overall simplification—for example, in cases where the combination of abstract lexical entries and a set of rules permits the formulation of phonological processes of great generality that would otherwise be inexpressible.

Thus, lexical representations and a system of phonological rules are chosen in such a way as to maximize a certain property that we may call the "value" of the grammar, a property that is sometimes called "simplicity." As has been emphasized repeatedly in the literature, the concept of "simplicity" or "value" is an empirical one. There is some correct answer to the question of how lexical items are represented and what the phonological rules are. A particular notion of "value" or "simplicity" will lead to an assumption about lexical items and phonological rules which is either right or wrong, and therefore the validity of the notion must be determined on empirical grounds, exactly as in the case of every other concept of linguistic theory. It may be difficult to obtain crucial empirical evidence bearing on proposed definitions of "simplicity," but this cannot obscure the fact that it is an empirical concept that is involved, and that one can no more employ a priori arguments in determining how "value" should be defined than in determining how to define "set of distinctive features" or "grammatical transformation" or any other concept of linguistic theory.

A specific proposal as to the definition of "value" will make certain assumptions as to what constitutes a linguistically significant generalization, as to what constitutes a "regularity" of the sort that a child will use as a way of organizing the data he is confronted with in the course of language acquisition. The child is presented with certain data; he arrives at a specific grammar, with a specific representation of lexical items and a certain system of phonological

rules. The relation between data and grammar is, we naturally assume, language-independent: there is no basis for supposing that individuals differ genetically in their ability to learn one rather than another natural language. Consequently, the relationship is determined by a principle of universal grammar. Specifically, the definition of "value" or "simplicity" must be part of universal grammar, and a specific proposal will be right or wrong as it does or does not play its part in accounting for the actually existing relation between data and grammar.

Summarizing, we postulate a set of lexical matrices and a system of phonological rules which jointly maximize value, in some sense which will be defined. Phonological representation in terms of lexical matrices (as modified through readjustment rules—see Chapter One, Section 5.1, and Chapter Eight, Section 6.5) is abstract in the sense that the phonological representation is not necessarily a submatrix of the phonetic representation. We do not, in other words, impose the conditions of linearity and invariance (see Chomsky, 1964) on the relation between phonological and phonetic representation. The indirectness of this relation must be purchased at the cost of adding rules to the grammar. Given a definition of "value," we can therefore say that the facts of pronunciation induce the representation of items in the lexicon.⁴

Notice that the phonetic features appear in lexical entries as abstract classificatory markers with a status rather similar to that of the classificatory features that assign formatives to such categories as "noun," "verb," "transitive." Like the latter, the phonological features indicate whether or not a given lexical item belongs to a given category. In the case of the phonological matrices, these categories have the meaning "begins with a voiced stop," "contains a vowel," "ends with a strident nonback obstruent," and so on. In view of the fact that phonological features are classificatory devices, they are binary, as are all other classificatory features in the lexicon, for the natural way of indicating whether or not an item belongs to a particular category is by means of binary features. This does not mean that the phonetic features into which the phonological features are mapped must also be binary. In fact, the phonetic features are physical scales and may thus assume numerous coefficients, as determined by the rules of the phonological component. However, this fact clearly has no bearing on the binary structure of the phonological features, which, as noted, are abstract but not arbitrary categorial markers.⁵

As already noted, the phonetic representation can be thought of formally as a two-dimensional matrix in which the columns stand for consecutive units and the rows stand for individual phonetic features. The phonetic features can be characterized as physical scales describing independently controllable aspects of the speech event, such as vocalicness, nasality, voicing, glottalization. There are, therefore, as many phonetic features as there are aspects under partially independent control. It is in this sense that the totality of phonetic features can be said to represent the speech-producing capabilities of the human vocal apparatus. We shall say that the phonetic representations of two units are distinct if they differ in the coefficient assigned to at least one feature; phonetic representations of sequences of units are distinct if they contain distinct units or if they differ in the number or order of units.

At the level of phonetic representation, utterances are comparable across languages; it thus makes sense to ask whether the phonetic representation of an utterance of language L_1 is distinct from a phonetic representation of an utterance of a different language L_2 . For

⁴ For additional discussion see Chapter Four, Section 2.

⁵ Failure to differentiate sharply between abstract phonological features and concrete phonetic scales has been one of the main reasons for the protracted and essentially fruitless debate concerning the binary character of the Jakobsonian distinctive features.

example, an utterance containing an apical dental stop must have a different phonetic representation from an utterance that is identical except for containing a laminal dental stop in place of the apical dental stop. The representation must differ since the distinction is determined in part by language-specific rules; it is not a case of universal free variation. An interesting example of cross-language contrasts that require a special phonetic feature is provided by the labiovelar consonants found in many African languages. In some languages, such as Yoruba, these consonants are produced with a special clicklike suction, whereas in other languages, such as Late, they are produced without this suction (Ladefoged, 1964, p. 9). Since clicklike suction is clearly an independently controllable aspect of the speech event, the data just cited establish suction as a separate phonetic feature, regardless of the fact that apparently in no language are there contrasting pairs of utterances that differ solely in this feature.

The situation is not always straightforward, however. Since phonetic features are scales which may in principle assume numerous discrete coefficients, the question may arise, under certain circumstances, whether a certain phonetic contrast is to be represented by means of a new phonetic feature or by increasing the number of coefficients that some already extant phonetic feature may be allowed to assume. The latter solution may appear especially attractive in cases where a slight redefinition of some phonetic feature would readily accommodate the proposed solution.

To summarize, the features have a phonetic function and a classificatory function. In their phonetic function, they are scales that admit a fixed number of values, and they relate to independently controllable aspects of the speech event or independent elements of perceptual representation. In their classificatory function they admit only two coefficients, and they fall together with other categories that specify the idiosyncratic properties of lexical items. The only condition that we have so far imposed on the features in their lexical, classificatory function is that lexical representations be chosen in such a way as to maximize the "value" of the lexicon and grammar, where the notion "value" is still to be defined precisely, though its general properties are clear. Apart from this, the representation of a lexical item as a feature complex may be perfectly abstract.

In a later discussion (see Chapter Nine), we will consider significantly heavier conditions on lexical representation. There we will turn to the question of "plausible phonological rules" and, more generally, to ways in which a particular feature may or may not function in the lexicon and in the phonology. These considerations will differentiate features from one another with respect to the role that they can play in the system of rules and in lexical representation. At that point in the development of our theory, considerations beyond maximization of value will enter into the determination of lexical representations.

2. *The phonetic features*

In the remainder of this chapter we attempt to sketch the universal set of phonetic features. Our aim is to cover every inherent phonetic feature regardless of whether it plays a role in the phonetics of English. We are well aware of the many gaps in our knowledge that make the success of this undertaking somewhat problematical, but we feel that general phonetics has been neglected for so long that agreement on even the most elementary propositions of phonetic theory cannot be taken for granted at present.

In the succeeding pages we shall list the individual features that together represent the phonetic capabilities of man. Each feature is a physical scale defined by two points, which are designated by antonymous adjectives: high-nonhigh, voiced-nonvoiced (voiceless), tense-nontense (*lax*). We shall describe the articulatory correlate of every feature and illustrate the feature by citing examples of its occurrence in different languages of the world. We shall speak of the acoustical and perceptual correlates of a feature only occasionally, not because we regard these aspects as either less interesting or less important, but rather because such discussions would make this section, which is itself a digression from the main theme of our book, much too long. We shall consider the phonetic features under the headings given below. (The numbers in parentheses refer to the section in which the feature is discussed.)

Major class features (3)

Sonorant (3.1)

Vocalic (3.2)

Consonantal (3.3)

Cavity features (4)

Coronal (4.1.1)

Anterior (4.1.2)

Tongue-body features (4.2)

High

Low

Back

Round (4.3)

Distributed (4.4)

Covered (4.5)

Glottal constrictions (4.6)

Secondary apertures (4.7)

Nasal (4.7.1)

Lateral (4.7.2)

Manner of articulation features (5)

Continuant (5.1)

Release features: instantaneous and delayed (5.2)

Primary release (5.2.1)

Secondary release (5.2.2)

Supplementary movements (5.3)

Suction (5.3.1)

Velaric suction (clicks)

Implosion

Pressure (5.3.2)

Velaric pressure

Ejectives

Tense (5.4)

Source features (6)

Heightened subglottal pressure (6.1)

Voice (6.2)

Strident (6.3)

Prosodic features (7)

Stress

Pitch

High

Low

Elevated

Rising

Falling

Concave

Length

This subdivision of features is made primarily for purposes of exposition and has little theoretical basis at present. It seems likely, however, that ultimately the features themselves will be seen to be organized in a hierarchical structure which may resemble the structure that we have imposed on them for purely expository reasons.

2.1. THE NEUTRAL POSITION

In most X-ray motion pictures of speech, it can readily be observed that just prior to speaking the subject positions his vocal tract in a certain characteristic manner. We shall call this configuration the "neutral position" and shall describe some of the ways in which it differs from the configuration of the vocal tract during quiet breathing. In the latter state the velum is lowered, thereby allowing air to pass through the nose; in the neutral position, on the other hand, the velum is raised, and the air flow through the nose is shut off. The body of the tongue, which in quiet breathing lies in a relaxed state on the floor of the mouth, is raised in the neutral position to about the level that it occupies in the articulation of the English vowel [e] in the word *bed*; but the blade of the tongue remains in about the same position as in quiet breathing.⁶ Since speech is generally produced on exhalation, the air pressure in the lungs just prior to speaking must be higher than the atmospheric pressure. During quiet breathing, the vocal cords must be widely spread apart since practically no sound is emitted. On the other hand, there is good reason to believe that prior to speaking the subject normally narrows his glottis and positions his vocal cords so that in the neutral position they will vibrate spontaneously in response to the normal, unimpeded air flow. Since this spontaneous vocal cord vibration has been almost totally ignored in the literature, we digress here in order to examine it in somewhat greater detail.

2.2. VOCAL CORD VIBRATION—SPONTANEOUS AND OTHERWISE

The two major factors controlling vocal cord vibration are the difference in air pressure below and above the glottis and the configuration of the vocal cords themselves—their tension, shape, and relative position. The subglottal pressure is that maintained in the trachea by the respiratory muscles. In the absence of a significant constriction in the oral cavity, the supraglottal pressure will be about equal to atmospheric pressure and will, of course, be lower than the subglottal pressure. If, on the other hand, there are significant constrictions in the oral cavity, the supraglottal pressure will rise above the atmospheric pressure since the air being exhaled from the lungs will not be allowed to flow out freely. Part or all of this air will be trapped in the supraglottal cavity, building up the pressure there and thus reducing the

⁶ We follow here Bell, Sweet, D. Jones and other phoneticians in drawing a distinction between the body and the blade of the tongue. See D. Jones (1956, p. 15): "... the part which normally lies opposite the teeth ridge is called the *blade*. The extremity of the tongue is called the *tip* or *point*, and is included in the blade." An almost identical description is given by Westermann and Ward (1933, p. 17).

pressure difference below and above the glottis. This is of importance to us here since, all other things being equal, this pressure difference determines the rate at which the air will flow from the lungs through the glottis, and it is the flow rate which determines whether or not the glottis will vibrate.

In order to initiate vocal cord vibration, it is not necessary that the glottis be totally closed. If the velocity of the air flow through the glottis is high enough, it may reduce the pressure inside the glottis opening (the Bernoulli effect) to the point where the pressure is insufficient to prevent the elastic tissue forces from pulling the vocal cords together and closing the glottis. As soon as the glottis is closed, the subglottal pressure begins to build up and ultimately becomes large enough to overcome the elastic tissue forces pulling the glottis shut. At this point the glottis is opened, and air flows through it again. The air flow is subsequently cut off again since it once more produces a critical pressure drop inside the glottis opening. Obviously the Bernoulli effect can take place only if the vocal cords are appropriately positioned. If they are spread too far apart, as they are in quiet breathing, the pressure drop inside the glottis will not be great enough to pull the vocal cords together and thus initiate vibration.

We have already postulated that in the neutral position the vocal cords are placed so as to vibrate spontaneously in response to the unimpeded air flow. It is, however, a well-known fact that vocal cord vibrations also occur when there is a radical constriction, or even total closure, in the oral cavity. Although direct observations have not as yet been made, there is reason to suppose that the positioning of the vocal cords and their manner of vibration in the presence of a significant constriction in the oral cavity differ in important ways from the position and vibration observed during unimpeded air flow. It thus appears that voicing in obstruents is a rather different matter from that observed in sonorants.

Theoretical investigations by Halle and Stevens (1967) have shown that for sounds with low first formants—i.e., for sounds other than vowels—periodic vocal cord vibrations can be maintained only if the width of the glottal pulse is increased by lengthening the open phase during each glottal vibration over that normally found in vowels and/or if the damping of the first formant is substantially increased by creating a larger average glottal opening. The increased glottal opening would also help to maintain the vibration in the face of the reduced pressure drop across the glottis resulting from the buildup of pressure behind the consonantal constriction in the supraglottal cavity.

Certain well-known observations seem to support the theoretical conclusion that nonspontaneous voicing involves quite different adjustments than does spontaneous voicing. Thus, the air flow in voiced obstruents is noticeably faster than that in sonorants (vowels, glides, liquids, nasals). This fact is readily explained on the assumption that the average glottal opening is larger in obstruents than in vowels. Moreover, studies now in progress indicate that at least in the production of some voiced obstruents, the glottis is partially open during the phonation period. Finally, the very common lengthening of vowels before voiced obstruents can be explained on the grounds that it requires time to shift from the glottis configuration appropriate for vowels to that appropriate for obstruents.

3. Major class features

Reduced to the most rudimentary terms, the behavior of the vocal tract in speech can be described as an alternation of closing and opening. During the closed phase the flow of air from the lungs is either impeded or stopped, and pressure is built up in the vocal tract; during

the open phase the air flows out freely. This skeleton of speech production provides the basis for the major class features, that is, the features that subdivide speech sounds into vowels, consonants, obstruents, sonorants, glides, and liquids. Each of the three major class features—sonorant, vocalic, consonantal—focuses on a different aspect of the open-versus-closed phase.

3.1. SONORANT—NONSONORANT (OBSTRUENT)

Sonorants are sounds produced with a vocal tract cavity configuration in which spontaneous voicing is possible; obstruents are produced with a cavity configuration that makes spontaneous voicing impossible.

As we noted above, spontaneous voicing may be suppressed by narrowing the air passage to a point where the rate of flow is reduced below the critical value needed for the Bernoulli effect to take place. Constrictions more radical than those found in the glides [y] and [w] will have this result. Hence sounds formed with more radical constrictions than the glides, i.e., stops, fricatives, and affricates, are nonsonorant, whereas vowels, glides, nasal consonants, and liquids are sonorant.

In this connection it should be observed that there appear to be differences in the degree of constriction with which [l]- and [r]-sounds are produced. In the better known cases these sounds are produced with a very moderate degree of constriction and are therefore clearly sonorants. There are, however, liquids which are produced with a quite radical constriction and which have to be regarded as obstruents. Such is the case, apparently, in Chipe-wyan, in certain Caucasian languages, and in those languages with strident liquids, such as the Czech [ř].

3.2. VOCALIC—NONVOCALIC

Vocalic sounds are produced with an *oral* cavity in which the most radical constriction does not exceed that found in the high vowels [i] and [u] and with vocal cords that are positioned so as to allow spontaneous voicing; in producing nonvocalic sounds one or both of these conditions are not satisfied.

Vocalic sounds, therefore, are the voiced vowels and liquids, whereas glides, nasal consonants, and obstruents, as well as voiceless vowels and liquids, are nonvocalic.⁷

3.3. CONSONANTAL—NONCONSONANTAL

Consonantal sounds are produced with a radical obstruction in the midsagittal region of the vocal tract; nonconsonantal sounds are produced without such an obstruction.

It is essential to note that the obstruction must be at least as narrow as that found in the fricative consonants and must, moreover, be located in the midsagittal region of the cavity. This feature, therefore, distinguishes liquids and consonants, both nasal and nonnasal, from glides and vowels. It has been observed by Sievers (1901) that an essential characteristic of vowels is their "dorsal articulation"; that is, vowels commonly are produced with the blade of the tongue some distance from the roof of the mouth. When the blade of the tongue is raised close enough to the roof of the mouth to produce the requisite obstruction, the result is a true consonant or a liquid. Thus an [l]-sound is produced when the tip of the tongue touches the roof of the mouth, thereby blocking the midsagittal region of the vocal tract. In the case of the common lingual [r]-sounds, the raised tongue narrows the passage sufficiently to produce a

⁷ Recent work indicates that in place of "vocalicness" the phonetic framework should contain a feature of "syllabicity"—see Chapter Eight, pages 353-55.

consonantal obstruction even if it does not make complete contact with the roof of the mouth. The uvular [R] is produced in a quite similar fashion, but in this case the lowered uvula rather than the raised tongue forms the obstruction in the midsagittal region of the vocal tract.

The presence of an obstruction in the midsagittal region is not necessarily accompanied by sufficient closure of the entire passage to suppress spontaneous voicing. The liquids are therefore consonantal sonorants. In producing the consonantal nonsonorants (obstruents), the passage is narrowed to a point where spontaneous vocal cord vibration is impossible; among the latter types of sounds are the plosives, affricates, and fricatives. On the other hand, not every sound produced with a raised tongue tip is consonantal. The so-called retroflex vowels are formed with a raised tongue tip, which, however, is not close enough to the palate to constitute a consonantal obstruction. These vowels are thus nonconsonantal.

The major class features therefore define the categories of speech sounds shown in Table 1.

TABLE 1. *The major class features*

	sonorant	consonantal	vocalic
voiced vowels	+	—	+
voiceless vowels	+	—	—
glides (I): <i>w, y</i>	+	—	—
glides (II): <i>h, ʔ</i>	+	—	—
liquids	+	+	+
nasal consonants	+	+	—
nonnasal consonants	—	+	—

4. Cavity features

4.1. PRIMARY STRICTURES

There are several ways in which primary strictures have been treated in the phonetic literature. The most widely known approach, that of the International Phonetic Alphabet, utilizes different features to characterize the strictures in vowels and in consonants. Vowel strictures are described with the help of the features "front-back" and "high-low," whereas consonantal strictures are characterized by means of a single multivalued parameter that refers to the location of the constriction. The disadvantage of this method is that it fails to bring out the obvious parallels between vocalic and consonantal strictures. Thus, the difference between palatal and velar consonants clearly parallels that between front and back vowels, for in both cases there are the same differences in the position of the body of the tongue. There is, however, no mechanism in the IPA framework to capture this and similar facts.

One of the many contributions of R. Jakobson is a phonetic framework in which many of these parallels are properly captured. As is well known, the salient characteristic of the Jakobsonian framework is that the same three features—"gravity," "compactness," and "diffuseness"—are used to describe the primary strictures in both vowels and consonants. This complete identification of vowel and consonant features seems in retrospect to have been too radical a solution, for reasons that we briefly outline below. We have therefore made a number of changes in the framework, in particular, with regard to the primary cavity features. The revised framework is quite likely to appear to depart from the earlier framework much more radically than it in fact does. This deceptive impression is the result of the unfortunate need to change terminology once again and replace the by now reasonably familiar

terms "compact," "diffuse," and "grave" in part by totally new terms, in part by terms that are a return to the status quo ante. We discuss the relationship between the two frameworks in Section 4.2.1.

4.1.1. CORONAL—NONCORONAL

Coronal sounds are produced with the blade of the tongue raised from its neutral position; noncoronal sounds are produced with the blade of the tongue in the neutral position.⁸

The phonetic classification effected by this feature is all but self-evident. The so-called dental, alveolar, and palato-alveolar consonants are coronal, as are the liquids articulated with the blade of the tongue. The uvular [R] and the consonants articulated with the lips or with the body of the tongue are noncoronal. The glides [y] and [w] are noncoronal. Finally, the so-called retroflex vowels which are found in some languages of India—e.g., Badaga (H. L. Gleason, personal communication)—as well as in many English dialects in the position before [r] are coronal. Nonretroflex vowels are, of course, noncoronal.

4.1.2. ANTERIOR—NONANTERIOR

Anterior sounds are produced with an obstruction that is located in front of the palato-alveolar region of the mouth; nonanterior sounds are produced without such an obstruction. The palato-alveolar region is that where the ordinary English [ʃ] is produced.

It follows from the proposed characterization that vowels, which are formed without constrictions in the oral cavity, are always nonanterior. Consonants and liquids are anterior when they are formed with an obstruction that is located farther forward than the obstruction for [ʃ]. The consonants that in traditional terminology are described as palato-alveolar, retroflex, palatal, velar, uvular, or pharyngeal are therefore nonanterior, whereas labials, dentals, and alveolars are anterior.

4.2. FEATURES RELATING TO THE BODY OF THE TONGUE: HIGH—NONHIGH, LOW—NONLOW, BACK—NONBACK

The three features "high," "low," "back" characterize the placement of the body of the tongue. Recall that in the neutral position the body of the tongue was assumed to be raised and fronted, approximating the configuration found in the vowel [e] in English *bed*. In characterizing these three features, we shall be concerned with the various displacements of the tongue body from the neutral position.

HIGH—NONHIGH. High sounds are produced by raising the body of the tongue above the level that it occupies in the neutral position; nonhigh sounds are produced without such a raising of the tongue body.

⁸ The term "coronal" is used here in the sense of the German *Vorderzungenslaut* and the Russian *peredne-jazyčnyj*. Sievers (1901) distinguished two types of linguo-palatal sounds with nonlateral articulation:

"(1) Coronale Articulation: die Articulation wird durch den vorderen Zungensaum bewirkt, welcher sich als eine mehr oder weniger scharfe Kante dem Gaumen entgegenstellt . . .

(2) Dorsale Articulation: die nothwendigen Engen bez. Verschlüsse werden durch Emporheben eines Theiles des Zungenrueckens . . . zum Gaumen gebildet" (p. 59).

In much the same way the term is defined by Broch (1911): "Wird die charakteristische Enge oder der Verschluss durch den Vorderrand der Zunge gebildet, wobei sich ihre Oberflaeche gewoehnlich auf einer groesseren oder kleineren Strecke als konkav bezeichnen laesst, so wird die Artikulation koronal genannt" (pp. 11 f.)

We differ somewhat from Sievers and Broch in that we regard as coronal all types of sounds formed with the blade of the tongue; Sievers and Broch did not use this term when speaking of sounds formed with the flat part of the blade (Sweet's "laminal"). (See note 6.) The latter distinction is handled in the present framework with the help of the feature "distributed" (see Section 4.4 of this chapter).

LOW-NONLOW. Low sounds are produced by lowering the body of the tongue below the level that it occupies in the neutral position; nonlow sounds are produced without such a lowering of the body of the tongue.

BACK-NONBACK. Back sounds are produced by retracting the body of the tongue from the neutral position; nonback sounds are produced without such a retraction from the neutral position.

The characterization of the vowels in terms of the three features above is quite straightforward and differs little from that found in most traditional phonetics books. We must observe only that the phonetic characterization of "low" and "high" rules out sounds that are $\left[\begin{array}{l} +\text{low} \\ +\text{high} \end{array} \right]$, for it is impossible to raise the body of the tongue above the neutral position and simultaneously lower it below that level.

The characterization of the consonants in terms of these same features is equally straightforward, though perhaps somewhat unfamiliar. Consider first the consonants where the primary constriction is formed with the body of the tongue, in other words, those that are both noncoronal and nonanterior: the palatals, velars, uvulars, and pharyngeals. These four "points of articulation" are readily captured with the help of the three features under discussion, as shown in Table 2.

TABLE 2.

	palatals	velars	uvulars	pharyngeals
high	+	+	-	-
low	-	-	-	+
back	-	+	+	+

The absence of nonhigh nonback consonants is a direct consequence of the fact that the body of the tongue can form a constriction only if it is high or back.

While no language known to us has all four types of consonants in Table 2, there are quite a number of languages in which three of the four classes are attested. Serer, a West African language, has palatal, velar, and uvular voiceless stops.⁹ Ubykh, a Caucasian language, distinguishes pharyngeal, uvular, velar, and perhaps also palatal obstruents (Vogt, 1963; Allen, 1964). In Ubykh, as in many other languages, such as Gilyak (see Zinder and Matusevič, 1937; Halle, 1957), the difference between velar and uvular points of articulation is paralleled by the difference between nonstrident and strident. This, however, is by no means universal. For instance, the spectrographic evidence published by Ladefoged (1964, p. 22) shows that in Serer the velar and uvular stops are both nonstrident plosives. Distinctions among palatal, velar, and uvular obstruents are also found in Chinook (Boas, 1911) and are mentioned by Trubetzkoy (1958, p. 122) as attested in certain Nilotic languages (Herero, Nuer, Dinka).

We must now inquire into the role that the features "high," "low," and "back" play in the remaining class of consonants, which in terms of the present framework are anterior and/or coronal. We observe that the three features may be used in a natural manner to characterize subsidiary consonantal articulations such as palatalization, velarization, and pharyngealization. These subsidiary articulations consist in the superimposition of vowel-like

⁹ The following contrasting forms are cited by Ladefoged (1964, p. 46; see also pp. 21-22): [k₁it] "gift," [kid] "eyes," [qos] "leg," where the symbol *k*₁ represents the voiceless palatal stop equivalent to the IPA *c*.

articulations on the basic consonantal articulation. In palatalization the superimposed subsidiary articulation is [i]-like; in velarization, [ɨ]-like; and in pharyngealization, [a]-like. The most straightforward procedure is, therefore, to express these superimposed vowel-like articulations with the help of the features "high," "low," and "back," which are used to characterize the same articulations when they appear in the vowels. We shall say that palatalized consonants are high and nonback; velarized consonants are high and back; the pharyngealized consonants (e.g., the Arabic "emphatic" consonants) are low and back. On the other hand, consonants neutral with respect to palatalization, velarization, and pharyngealization are $\begin{bmatrix} -\text{high} \\ -\text{back} \end{bmatrix}$, since such configurations lack a constriction formed by the body of the tongue. Incidentally, it is not clear what role the feature "low" plays in such configurations since we do not know of any language with uvularized dentals or labials. If such consonants exist, however, they will be characterized in terms of our framework as nonhigh, nonlow, and back.

The palato-alveolars differ from the labial and dental consonants in that they are redundantly [+high]. In place of the four-way contrast found in the labials and dentals, the palato-alveolars, therefore, exhibit only a two-way contrast of palatalized ([−back]) and velarized ([+back]). The phonetic contrast can be seen very clearly in the X-ray tracings given by Fant (1960) of the two [ʃ] sounds of standard Russian.

Table 3 (p. 307) gives the feature composition of the most important classes of speech sounds.

4.2.1. ON THE RELATIONSHIP BETWEEN THE FEATURES "DIFFUSENESS," "COMPACTNESS," AND "GRAVITY" AND THE FEATURES OF THE PRECEDING SECTIONS

The features discussed in the preceding sections are basically revised versions of "diffuseness," "compactness," and "gravity," which are well known from earlier presentations of the distinctive feature framework where they served to characterize the main articulatory configurations in the vowels as well as the consonants. As more and more languages were described within this framework, it became increasingly clear that there was a need for modification along the lines discussed in the preceding section. In this section we shall examine some of the problems that arose within the earlier framework and outline the way in which these problems are overcome by the revised features presented above. This question has recently been examined also by McCawley (1967a).

The revisions proposed in the last few pages have the following main effects:

- (1) Features specifying the position of the body of the tongue are now the same for vowels and consonants.
- (2) In the characterization of vowel articulations, the features "high," "low," "back" correspond to the earlier "diffuse," "compact," and "grave," respectively. In consonants, the same three revised features correspond to palatalization, velarization, and pharyngealization in the manner discussed above.
- (3) The feature "anterior" mirrors precisely the feature "diffuse" in consonants.
- (4) The feature "coronal" corresponds most closely to the feature "grave" in consonants but with opposite value. Except for the palatals ([k₁], etc.), consonants that were classified as nongrave in the earlier framework are coronal in the revised framework, whereas those that were classified as grave are noncoronal. The palatals, which in the earlier framework were nongrave, are noncoronal.

We recall that in the earlier framework the feature "diffuse" was used to characterize

TABLE 3. *Feature composition of the primary classes of speech sounds*

	anterior	coronal	high	low	back
CONSONANTS					
labials	+	-	-	-	-
dentals	+	+	-	-	-
palato-alveolars (does not exist)	-	+	+	-	-
palatalized labials	+	-	+	-	-
palatalized dentals	+	+	+	-	-
palatals	-	-	+	-	-
velarized labials	+	-	+	-	+
velarized dentals	+	+	+	-	+
velarized palato-alveolars	-	+	+	-	+
velars	-	-	+	-	+
(?) uvularized labials	+	-	-	-	+
(?) uvularized dentals	+	+	-	-	+
uvulars	-	-	-	-	+
pharyngealized labials	+	-	-	+	+
pharyngealized dentals	+	+	-	+	+
pharyngeals	-	-	-	+	+
VOWELS (nonretroflex)					
high front	-	-	+	-	-
high back	-	-	+	-	+
mid front	-	-	-	-	-
mid back	-	-	-	-	+
low front	-	-	-	+	-
low back	-	-	-	+	+
GLIDES					
y	-	-	+	-	-
w	-	-	+	-	+
h, [?]	-	-	-	+	-
LIQUIDS					
dental	+	+	-	-	-
palatal	-	-	+	-	-
uvular	-	-	-	-	+
palato-alveolar	-	+	+	-	-

both the distinction between high and nonhigh vowels and that between what we have called anterior and nonanterior consonants. As a result the articulatory and acoustical characterization of the feature became quite complex and rather implausible. (See, for example, the discussion of diffuseness in Halle (1964).)

A further consequence of the same fact was the need to characterize palatalization, velarization, and pharyngealization by means of independent features. This, in turn, failed to explain why these subsidiary articulations are not found with consonants that are formed with the body of the tongue, i.e., consonants that are noncoronal and nonanterior in the present framework. In the former framework this was a mere accident; in the revised framework the gap is structurally motivated, as shown in Section 4.2. It is worthy of note that rounding (labialization), which is also a subsidiary articulation, is not subject to similar restrictions. All classes of consonants, including labials, may be rounded.

A related inadequacy of the former framework is that it provided no explanation for the fact that palatalization, velarization, and pharyngealization are mutually exclusive. In

the revised framework the co-occurrence of these articulations is a logical impossibility since a given sound cannot be back and nonback. In the former framework, on the other hand, this is no more than a coincidence.

The former framework, furthermore, did not bring out the fact that palatalization and velarization characteristically occur before front and back vowels, respectively; the connection between palatalization and front vowels and between velarization and back vowels was no more motivated than a connection between glottalization or voicing and front vowels. In the revised framework, on the other hand, palatalization and velarization are obvious cases of regressive assimilation.

The earlier framework failed to account for the appearance of palatal, in place of velar, consonants in precisely the same environments where other classes of consonants undergo palatalization. (Recall that palatalization preserves the point of articulation, whereas the change of velar to palatal constitutes a change in the point of articulation.) In the revised framework these two superficially distinct processes are shown to be a result of the same change, that is, [+back] to [-back]. A parallel argument can be given for the treatment of velarization and pharyngealization in the two frameworks.

The earlier framework made it impossible in principle to distinguish velar from uvular or pharyngeal consonants by means of their points of articulation. Such distinctions instead had to be made by the use of some subsidiary feature such as "stridency." There are, however, languages (Serer, for example—see p. 305 and note 9) in which velar and uvular consonants do not differ in any such subsidiary feature and which therefore could not be accounted for. This shortcoming is easily taken care of in the revised framework, in which the different points of articulation in velar, uvular, and pharyngeal consonants are specified with the help of the features "high," "low," and "back."

4.2.2. DEGREES OF NARROWING IN THE VOCAL TRACT

In our discussion of the features up to this point, we have spoken at length about the location of strictures in the vocal tract but we have said nothing about differences in the degree of narrowing that can readily be observed in the strictures found in different sounds. This omission has been due to the tacit assumption that the degree of narrowing is determinable from other features of a particular sound. This approach is perfectly familiar in phonetics; for example, no phonetics book does more than remark that in rounded vowels the degree of lip narrowing is most radical for high vowels and least radical for low vowels. While degree of narrowing never functions as the sole cue for differentiating two otherwise identical utterances, it is not true that in all languages the degree of narrowing involved in a particular sound is always predictable from universal phonetic principles. This becomes quite clear if we examine velarized consonants, which appear in various languages with radically different degrees of velar constriction.

In Russian a moderate narrowing in the velar region is present in the articulation of the so-called "hard" consonants, where concomitant with velarization we find a certain degree of lip rounding.¹⁰

Velarization with more radical narrowing has been reported by C. M. Doke (1931) as occurring in Shona:

Velarization is brought about by an abnormal raising of the back of the tongue towards the soft palate (velum), instead of the usual slight raising effected in

¹⁰ See Broch (1911, pp. 224 ff.) and X-ray pictures in Fant (1960, pp. 140, 163, 170, 186).