fricative [s] before [u], [o], and [a]. Thus, we have a classic case of complementary distribution. There are two possible solutions. First, we can recognize underlying /si, se, su, so, sa/ and posit a rule such as
\[ s \rightarrow s / - \{i\} \]
which converts /si/ and /se/ to [si] and [se], respectively. Or we can recognize underlying /ši, še, šu, šo, ša/ and posit a rule such as
\[ š \rightarrow s / - \{u\} \]
which converts /šu/, /šo/, and /ša/ to [su], [so], and [sa], respectively. The first solution is plausible, while the second solution is implausible. Recognizing only /s/ is plausible, because the rule which derives [s] before /i/ and /e/ is a natural assimilation rule. That is, when /ši/ becomes [ši], the alveolar /s/ assimilates to the frontness (or palatality) of /i/. Similarly, when /še/ becomes [še] the same assimilatory process is observed. On the other hand, if we start with underlying /š/, the rule which is required to derive [š] before /u/, /o/, and /a/ is not a natural assimilation rule. While the process of a palatal consonant becoming nonpalatal before a nonpalatal vowel would appear to be assimilatory in nature, the question is why /š/ should become more fronted (that is, to [š]) rather than back (to, say, [x]) before the back vowels in question. Thus, this rule seems to be unmotivated from a phonetic point of view.

Rule plausibility usually refers to phonetic naturalness. Certain phonological rules are found to occur frequently in languages, and the reason for this frequency is the fact that segments tend to assimilate to neighboring segments, and they do so in fairly predictable ways (see Schachter, 1969; Schane, 1972). The notion which is usually brought forth to explain these phenomena is ease of articulation. It is claimed to be easier to pronounce [ši] than [si], since in the first case both segments agree in palatality.

What this means is that plausible phonological rules are usually unidirectional. Thus, one can use this criterion in phonological analysis and try to establish an inventory of underlying segments from which the surface segments can be derived by plausible rules. This criterion, like the other criteria, is subject to other considerations. In particular, some languages do have implausible or "crazy" rules (Bach and Harms, 1972). As discussed in 5.2.3, the most phonetically natural rule is not necessarily the most simple rule. However, as a general principle, plausibility or rule naturalness is an important criterion in conducting phonological analyses.

4.1 Simplicity, Economy, and Generality

In 3.4.2, the notion of economy was said to be one of the criteria often used as a guide in phonemic analysis. A solution with fewer phonemes is judged more economical than a solution recognizing more phonemes. Similarly, we might say that a solution using fewer rules is more economical than a solution requiring more rules, and so on. Economy, then, is a quantitative measure by which a given solution can be evaluated as requiring fewer or more mechanisms (phonemes, rules, conventions, etc.) than another solution. This notion is characteristic of phonemic approaches to phonology, and, as we shall see, has its application in the history of generative phonology as well.

While one might be tempted to view a solution recognizing fewer phonemes as "simpler" than a solution recognizing more phonemes, there is another view which equates simplicity with generality. In terms of the phonemic inventory, the following argument might be made:
The vowel system of $S_1$ is more economical, because it involves fewer vowel phonemes. The vowel system of $S_2$, on the other hand, is more general, because it makes greater or more general use of the distinctive features of vowels. Looked at a little differently, $S_1$ will require a phonological constraint to the effect that the only front rounded vowel is /ɪ/: $S_2$ will contain no such constraint, since the front rounded series /ʊ, ə, ɛ/ is exactly parallel to the front unrounded series /i, e, ø/ and the back rounded series /u, o, ɑ/. Numerous examples of this sort can be found. For example, compare the consonant systems of the following two hypothetical solutions of the same language:

<table>
<thead>
<tr>
<th></th>
<th>$S_3$</th>
<th>$S_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td>p</td>
<td>p</td>
</tr>
<tr>
<td>t</td>
<td>t</td>
<td>t</td>
</tr>
<tr>
<td>k</td>
<td>k</td>
<td>k</td>
</tr>
<tr>
<td>b</td>
<td>b</td>
<td>b</td>
</tr>
<tr>
<td>d</td>
<td>d</td>
<td>d</td>
</tr>
<tr>
<td>g</td>
<td>g</td>
<td>g</td>
</tr>
<tr>
<td>m</td>
<td>m</td>
<td>m</td>
</tr>
<tr>
<td>n</td>
<td>n</td>
<td>n</td>
</tr>
</tbody>
</table>

(100) 4.1  Phonological Simplicity

In terms of the number of consonant phonemes, $S_3$ is more economical than $S_4$, since it lacks an /ŋ/ phoneme. However, in terms of generality, $S_4$ is simpler than $S_3$, since it makes greater use of the place of articulation features. When applied to phonemic inventories, generality can usually be equated with the notion of pattern congruity discussed in Chapter 3.

Since the more economical phonological systems, that is, those lacking phonemes, often require phonological constraints, they are *uneconomical* in this particular sense. $S_1$ requires a constraint which forbids the feature combination [+ high, - back, + round, V] (that is, /ɔ/ and /ɑ/), and $S_4$ requires a constraint forbidding the feature combination [+ nasal, + back, C] (that is, /ŋ/). Since $S_2$ and $S_4$ do not require any such constraints, they are in this sense more economical.

### 4.1.1 Lexical Simplicity vs. Rule Simplicity

This contrast points out a crucial problem in the assessment of phonological economy/simplicity: an economy in one part of the phonology may create a complexity in another part of the phonology. This means that in order to arrive at some judgment as to the simplicity of an analysis, it is necessary to take into consideration the whole analysis, and not just the inventory of phonemic segments, for instance.

Nowhere is this fact more blatantly clear than in the relationship between simplicity in the lexicon (or phonological level, since lexical items are entered in their phonological form) and simplicity in the phonological rules, which convert the lexical (that is, phonological) representations into phonetic ones. Let us, for example, return to the /ŋ/ solution, which was argued for (see 3.3.1) in preference to an /ŋ/ phoneme in English. First, it is clear that positing /ŋ/ in words such as *sing* [sɨn] and *long* [lɒŋ] permits a great economy, since we do not need an /ŋ/ phoneme, and since we can now equate /ŋ/ with /mb/ and /nd/ and achieve greater generality there. But two complications arise as a result. First, a hole in the pattern is created, as in $S_3$ above, since a phonological constraint will be necessary to rule out the possibility of combining the consonant features [+ nasal] and [+ back] in English. And second, although they turn out to be well-motivated, rules will be required to convert underlying /ŋ/ to [ŋ] in the appropriate environments.

### 4.1.2 The Simplicity Metric

While notions of economy and simplicity have always been implicit in linguistic analysis, the concept of simplicity has gained theoretical significance within the framework of generative grammar (in this case, generative phonology). In Chapter 3, reference was made to the levels of adequacy explicitly differentiated by Chomsky and other generative grammarians. *Observational adequacy* is said to be achieved by a grammar “if it correctly describes the data on which it is based and nothing more—if, in other words, it gives a compact one-one organization of this data” (Chomsky and Halle, 1965:458). If, on the other hand, the grammar achieves the higher goal of capturing the “tacit knowledge” of native speakers, it is said to reach the level of *descriptive adequacy*. In other words, such a grammar is said to be *psychologically* real. In phonology, as in other areas of linguistics, our goal is to write grammars which are psychologically real. In order to do so, our theory of phonology must be developed in such a way that when alternative solutions to a problem are proposed, it leads us to choose the one solution which captures the native competence of speakers. In other words, an *evaluation procedure* is necessary to judge the relative merits of alternative proposals in analyzing a given language.

In the early period of generative theory, an approach similar to Occam’s Razor was outlined. Thus Chomsky (1962:223) wrote: “we must apparently do what any scientist does when faced with the task of constructing a theory to account for a particular subject-matter—namely try various ways and choose the simplest that can be found.” However, in order to do this, it is necessary to have a good idea of what simplicity is, or of what makes one solution simpler than another. As we have already seen, simplicity in one part of the phonology may lead to complexity in another part. Thus the notion of simplicity must be refined and formalized if it is to be of any use in phonological analysis.

Simplicity is a technical term defined by the theory, and not a loosely conceptualized intuitive notion. Originally Chomsky (1955) stated that “simplicity correlates with ‘maximal degree of generalization’.” Linguistic
Phonological Simplicity

4.1

The theory therefore provides a *simplicity metric* which will automatically assign simplicity coefficients to alternate solutions so that the correct solution is chosen. In this way the theory reaches the level of *explanatory adequacy*, that is, it motivates the choice of the best grammar from all the descriptively adequate grammars. In later writings this simplicity metric becomes the second part of a "two-pronged attack":

Suppose that we are concerned to develop a linguistic theory that meets the level of explanatory adequacy. It seems that a two-pronged attack on this problem offers some hope of success. In the first place, we attempt to enrich the structure of linguistic theory so as to restrict the class of grammars compatible with the data given—in other words, we attempt to make the strongest legitimate universal claim about the structure of language. Second, we attempt to construct an evaluation procedure for selecting one among the various grammars permitted by the proposed linguistic theory and compatible with the given data. (Chomsky and Halle, 1965:106-107)

In singling out simplicity as an evaluation procedure, the claim is made that phonologies which are maximally simple (as defined by the theory) are preferred by speakers, or are perhaps more easily learned by children. For as Chomsky (1960) makes clear, linguistic theory is designed "to exhibit the built-in data organizing capacities of the child which lead him to develop the specific linguistic competence characterized in a fully explicit grammar." Thus, every claim about the nature of simplicity is necessarily a claim about the nature of one's innate language faculty.

The ability of children to construct a grammar of their language upon exposure to it has been schematized by Chomsky as follows:

\[
\text{Corpus} \rightarrow \begin{array}{c|c}
\text{LAD} & \text{Grammar} \\
\end{array}
\]

On the basis of a corpus of raw data and guided by the innate constraints on language (as represented by the LAD, that is, language acquisition device), the child constructs a grammar. Since we do not at present have great insight into how children discover this grammar, the possibility of developing a "discovery procedure" was deemed too ambitious a project by Chomsky. Instead as seen in the following schema,

\[
G_1 \rightarrow \begin{array}{c|c}
\text{EM} & G_1 \text{ or } G_2 \\
\end{array}
\]

an evaluation metric (EM) is proposed which, on receiving input from two grammars (two solutions) and the corpus upon which the grammars are based, will tell us which of the two is preferred. This, then, is termed an "evaluation procedure," and the proposed criterion is simplicity. It is hypothesized that the child, upon exposure to a given language, will construct the *simplest* grammar of that language compatible with the data. It is for this reason that so much attention has been paid to simplicity and the simplicity metric.

Of course, if linguistic theory becomes sufficiently developed so that the constraints placed on it are strong enough to pick out the right grammar for any language, such an evaluation procedure may not be necessary. While there is much disagreement today among linguists over the merits or usefulness of a simplicity metric (especially as developed so far—see below), most linguists seem to work under the assumption that such a metric is a necessary part of the metatheory.

4.2 Feature Counting

In phonology, simplicity has been equated with the number of features required to capture a phonological generalization. The fewer features required, the simpler (or more highly valued) the phonology. The concept of such an evaluation metric is possible only if we assume that phonological descriptions should be made in terms of (distinctive) features and not in terms of indivisible segments (for example, phonemes). The first statements concerning the simplicity metric dwell on this point. Thus, Halle (1962:381-382) mentions two rules similar to R_1 and R_2 given below:

\[
R_1: \quad k \rightarrow \varepsilon / - \begin{array}{c}
i \\
e \\
a \\
\end{array}
\]

\[
R_2: \quad k \rightarrow \varepsilon / - \begin{array}{c}
p \\
r \\
a \\
\end{array}
\]

In terms of segments, each rule is stated with five symbols. But, as Halle points out, a rule such as R_1 is considerably more highly valued in a phonology than a rule such as R_2. Using indivisible units such as /k/, /i/, /p/, /r/ does not reveal the fact that a phonological process can be conditioned by the front vowels /i, e, e/, but not by the voiceless stop /p/, the liquid /r/, and the low back vowel /a/ functioning as a single class.

Halle notes that if these rules are translated into distinctive features, then the simplicity of R_1 is revealed, as compared to the complexity of R_2:

\[
R_1: \quad k \rightarrow \varepsilon / - \begin{array}{c}
+\text{syl} \\
-\text{back}
\end{array}
\]

\[
R_2: \quad k \rightarrow \varepsilon / - \begin{array}{c}
+\text{back} \\
-\text{syl}
\end{array}
\]
In $R_1$, the environment /i, e, æ/ is expressible in terms of the feature specifications [+syllabic] and [−back], that is, two features. In $R_2$, however, when one attempts to express the environment /p, r, a/ in terms of features, the result is a disjunction involving fourteen feature specifications. A simple formulation is achieved in the first case, but an astounding complexity is found in the second case. This is what is desired. Thus, simplicity can be quantified by counting features, and only a theory which requires that segments are composites of features will differentiate between real and spurious generalizations.

What this procedure reveals is that certain segments constitute natural classes, whereas others do not. Thus, /i, e, æ/ constitute a natural class expressible as [+syll, −back]. Halle states that two (or more) segments constitute a natural class when they can be specified by fewer features than any one member of the class. Thus, /i/ is specified as [+syll, +high, −back] (three features), /e/ is specified as [+syll, −high, −low, −back] (four features), and /æ/ is specified as [+syll, +low, −back] (three features). In each case, at least one more feature is required to specify any one member of the class than the class as a whole.

### 4.2.1 Feature Counting in the Lexicon

There are two places where features have been counted to assess the simplicity of a phonological system: the lexicon (lexical or phonological representation, that is, underlying forms) and the phonological rules. As we saw in Chapter 1, there are numerous phonological constraints characterizing any language. Thus, there are often redundancies created in the phonological representations by constraints on sequences of phonemes. Examples of such sequential constraints in English are: (1) if a word-initial segment is an affricate, that is, either /ʃ/ or /ʒ/, then the following segment must be a vowel; (2) if the second of two word-initial consonants is a stop (oral or nasal), then the first consonant is /s/. Both of these sequential constraints are language-specific, since there are languages which violate them. Thus, the word /dʒrə/ 'to sell' in Ewe breaks the first sequential constraint since the affricate /dʒ/ is followed by something other than a vowel. Similarly, Gwari breaks the second sequential constraint by allowing a variety of /CNV/ sequences, for example, /bmə/ 'to break,' /dənə/ 'to be in.' There are, however, universal sequential constraints which characterize all languages. One possible universal constraint is suggested by Gwari. While Gwari has an implosive /b/ phoneme, for example, /bənə/ 'to beg,' there are no instances of /bənV/ in the language (see Hyman, 1972a:187). This is presumably because of the phonetic complexity which would be involved in pronouncing an egressive nasal consonant after an ingressive implosive (at least in the same syllable).

4.2.1.1 Morpheme Structure Rules (MSRs) Because of sequential constraints, certain features of one segment can be predicted on the basis of features of another segment. That is, certain feature specifications are rendered redundant by sequential constraints. According to the theory of morpheme structure rules proposed by Halle (1959:30ff), redundant feature specifications are to be left blank in the underlying representations of morphemes. Consider the word chat [tʃæt], which has the following phonetic feature specifications:

<table>
<thead>
<tr>
<th>[c]</th>
<th>[æ]</th>
<th>[t]</th>
</tr>
</thead>
<tbody>
<tr>
<td>+cons</td>
<td>−syll</td>
<td>+cons</td>
</tr>
<tr>
<td>−son</td>
<td>+son</td>
<td>−son</td>
</tr>
<tr>
<td>+high</td>
<td>−high</td>
<td>−high</td>
</tr>
<tr>
<td>−low</td>
<td>+low</td>
<td>−low</td>
</tr>
<tr>
<td>−ant</td>
<td>+ant</td>
<td>+ant</td>
</tr>
<tr>
<td>+cor</td>
<td>−cor</td>
<td>+cor</td>
</tr>
<tr>
<td>−voice</td>
<td>+voice</td>
<td>−voice</td>
</tr>
<tr>
<td>−cont</td>
<td>+cont</td>
<td>−cont</td>
</tr>
<tr>
<td>−nasal</td>
<td>−nasal</td>
<td>−nasal</td>
</tr>
<tr>
<td>+del rel</td>
<td>+del rel</td>
<td>+del rel</td>
</tr>
<tr>
<td>−round</td>
<td>−round</td>
<td>−round</td>
</tr>
</tbody>
</table>

1 Sequences of C + N in Gwari can be referred to as single nasally released consonants, i.e., C+N. As argued in Hyman (1972a), a nasally released implosive may be phonetically impossible, since what is involved in the production of an implosive is the rarefaction or lowering of the air pressure inside the mouth by a downward movement of the whole glottis. If the air pressure is lower within the mouth, it should not be possible for air to be released through the nose.
On the basis of the first sequential constraint given above, it can be predicted that any segment following a \([-\text{cont}, +\text{del rel}]\) segment (that is, an affricate) will be a vowel. Thus, the major category features for vowels, that is, \([-\text{cons}, +\text{syll}, +\text{son}]\), are predictable and are therefore left blank in the underlying form of chat. Formally, the unspecified features are entered with zeros, that is, \([0\ \text{cons}, 0\ \text{syll}, 0\ \text{son}]\), which are filled in with pluses and minuses by a morpheme structure rule such as the following:

\[
\begin{align*}
0\ \text{cons} & \rightarrow [-\text{cons}] + [+\text{del rel}] \\
0\ \text{syll} & \\
0\ \text{son} & \\
\end{align*}
\]

However, the distinctive feature matrix obtained after specifying the redundant vowel features as \([0\ \text{cons}, 0\ \text{syll}, 0\ \text{son}]\) is still full of redundancies. In addition to sequential constraints of the type just discussed, languages are characterized by extensive segmental constraints. It has been seen that the feature specifications \([-\text{cons}, +\text{syll}, +\text{son}]\) of /æ/ are redundant as a result of the \([-\text{cont}, +\text{del rel}]\) specifications of /ɛ/. In addition, the specifications \([-\text{cont}, +\text{del rel}]\) allow us to predict all of the remaining features of /ɛ/ except \([-\text{voice}]\). Since there are only two affricates in English, namely /tʃ/ and /fɹ/, we know a lot about a segment once we know that it is an affricate. ( Needless to say, the same does not apply to a language having other affricates in addition to /tʃ/ and /fɹ/, for example, /pʰtʃ/, /tʃʰ/, /kʰtʃ/.) We can predict that it is \([+\text{cons}, -\text{syll}, -\text{son}]\) (that is, an obstruent); that it is \([+\text{high}, -\text{back}, -\text{low}, -\text{ant}, +\text{cor}, -\text{round}]\) (that is, an unrounded alveopalatal); that it is \([-\text{nasal}]\) (that is, oral); and that it is \([+\text{strident}]\) (as opposed to the \([-\text{strident}]\) affricate [tʃ]). Thus, in phonological representations, such features are left unspecified (via zeros) and are filled in by segmental morpheme structure rules such as the following:

\[
\begin{align*}
0\ \text{cons} & \rightarrow [+\text{cons}] \\
0\ \text{syll} & [-\text{syll}] \\
0\ \text{son} & [-\text{son}] \\
0\ \text{high} & [+\text{high}] \\
0\ \text{back} & [-\text{back}] \\
0\ \text{low} & [-\text{low}] \\
0\ \text{ant} & [-\text{ant}] \\
0\ \text{cor} & [+\text{cor}] \\
-\text{cont} & [-\text{cont}] \\
0\ \text{nasal} & [-\text{nasal}] \\
0\ \text{strid} & [+\text{strid}] \\
+\text{del rel} & [+\text{del rel}] \\
0\ \text{round} & [-\text{round}] \\
\end{align*}
\]

Only the \([-\text{voice}]\) of /ɜ/ is not predictable, since the phoneme /ʃ/ also satisfies the segmental constraints of English.

While this segmental morpheme structure rule is designed to capture a redundancy in the segmental inventory of English phonemes, at least two universal redundancies have been confused with the language-specific redundancies. First, affricates are automatically \([-\text{nasal}]\), since it is phonetically impossible to have a nasal affricate. Thus, this part of the redundancy found in English is not a property of English, but rather a property of universal phonetics, and should be stated as such. The following segmental constraint on feature combinations is therefore universal:

\[
\begin{align*}
-\text{cont} & \\
+\text{del rel} & \rightarrow [-\text{nasal}] \\
0\ \text{nasal} & \\
\end{align*}
\]

A second universal segmental constraint concerns the features High and Low. According to these features a segment cannot be \([+\text{high}, +\text{low}]\), since it is impossible for the tongue to be both raised and lowered simultaneously from the neutral position. Thus, two universal segmental morpheme structure rules are required:

\[
\begin{align*}
+\text{high} & \\
0\ \text{low} & \rightarrow [-\text{low}] \\
0\ \text{high} & +\text{low} \rightarrow [-\text{high}] \\
\end{align*}
\]

Since /ʃ/ is \([+\text{high}]\), it is automatically \([-\text{low}]\).

We can continue to remove the redundant feature specifications from the underlying representation of chat. Concerning the vowel /æ/, we can predict \([-\text{high}]\) from the \([+\text{low}]\) specification, as just seen, as well as the feature specifications \([-\text{ant}, -\text{cor}, +\text{voice}, +\text{cont}, -\text{nasal}, -\text{strid}, +\text{del rel}]\). All vowels are universally \([-\text{ant}, +\text{cont}, -\text{strid}, +\text{del rel}]\). In addition, all underlying vowels in English are \([-\text{cor}]\) (since there are no underlying \([+\text{cor}]\) retroflex vowels), \([+\text{voice}]\) (since there are no underlying voiceless vowels), and \([-\text{nasal}]\) (since there are no underlying nasalized vowels). Finally, /æ/ is redundantly \([-\text{round}]\), since all \([-\text{round}]\) vowels in English are unrounded, that is, there is no /ə/.

Turning to the /t/ of chat, a number of features are redundant here too. The feature specifications \([+\text{ant}, +\text{cor}]\) tell us that we have an alveolar consonant. The feature specification \([-\text{voice}]\) is necessary to distinguish /t/ from /θ/, and the feature specification \([-\text{cont}]\) is necessary to distinguish it from /ð/ or /s/. All of the remaining features can be predicted from the

\footnote{A nasal affricate (i.e., nasal stop followed by a fricative release) is impossible because of the difficulty of building up oral pressure if the nasal passage allows a steady release of air.}

\footnote{In some recent work, however, Krohn (1972a,b) has suggested that such contradictory feature specifications as \([+\text{low}, +\text{high}]\) be "sequenced" within a segment, as in the diphthong /aɪ/, pronounced [aɪ].}
redundancies of English. The feature specifications [+cons] and [-del rel] are predictable from the [+ant] specification, since only obstruents and liquids can be [+ant] in English, and since the only affricates ([+del rel]) in English are alveopalatals ([+ant]). The features [-syl, -son, -nasal] are predictable from the [-voice] specification, while the features [-high, -back, -low] are all predictable from the [+ant, +cor] specifications. Finally, the [-strid] is redundant, since the segment is neither [+cont] (for example, like /s/) nor [+del rel] (for example, like /ʃ/). Thus, the complete redundancy-free underlying phonological matrix for the word chat is as given below:

<table>
<thead>
<tr>
<th>/ɛ/</th>
<th>/æ/</th>
<th>/t/</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 cons</td>
<td>0 cons</td>
<td>0 cons</td>
</tr>
<tr>
<td>0 syl</td>
<td>0 syl</td>
<td>0 syl</td>
</tr>
<tr>
<td>0 son</td>
<td>0 son</td>
<td>0 son</td>
</tr>
<tr>
<td>0 high</td>
<td>0 high</td>
<td>0 high</td>
</tr>
<tr>
<td>0 back</td>
<td>-back</td>
<td>0 back</td>
</tr>
<tr>
<td>0 low</td>
<td>+low</td>
<td>0 low</td>
</tr>
<tr>
<td>0 ant</td>
<td>0 ant</td>
<td>+ant</td>
</tr>
<tr>
<td>0 cor</td>
<td>0 cor</td>
<td>+cor</td>
</tr>
<tr>
<td>-voice</td>
<td>0 voice</td>
<td>-voice</td>
</tr>
<tr>
<td>-cont</td>
<td>0 cont</td>
<td>-cont</td>
</tr>
<tr>
<td>-nasal</td>
<td>0 nasal</td>
<td>0 nasal</td>
</tr>
<tr>
<td>0 strid</td>
<td>0 strid</td>
<td>0 strid</td>
</tr>
<tr>
<td>+del rel</td>
<td>0 del rel</td>
<td>0 del rel</td>
</tr>
<tr>
<td>0 round</td>
<td>0 round</td>
<td>0 round</td>
</tr>
</tbody>
</table>

4.2.1.2 Morpheme Structure Conditions (MSCs) As pointed out by Stanley (1967) and others, there are a number of problems inherent in this approach to phonological redundancy. This is particularly evident in the above analysis of the redundant feature specifications in the /t/ of /chat/. It turns out that some feature specifications predict more redundancies than others. For example, knowing that a segment is [-voice] automatically tells us that it is a voiceless obstruent in English, since there are no voiceless liquids, glides, nasals, or vowels in the language. (We are considering /h/ to be a fricative). Since this is the case, the [-voice] specification automatically predicts [-syl, -son, -nas], that is, three features. However, the opposite specification, that is, [+voice], does not tell us anything about the redundancies in the segment, since the segment can be either [-syl] or [+syl], [-son] or [+son], and [-nas] or [+nas]. Thus, one value of a given feature often carries more information than the opposite value (see 5.1.1).

In addition, the specification of one feature within a segment frequently carries more information than the specification of another feature. For example, the feature specification [+low] automatically narrows us down to the phonemes /ɛ/, /æ/, and /i/ in English. On the other hand, the feature specification [+cont] includes voiced and voiceless fricatives, liquids, glides, and vowels. The feature specification [-syl] is even more inclusive.

Thus, in assessing the redundancies and presenting them in the framework of morpheme structure rules, it is often necessary to look for those feature specifications from which the greatest number of other specifications can be predicted. In assessing the simplicity in the underlying forms, only pluses and minuses are counted; zeros do not count. Thus, according to the evaluation metric, the more zeros in the phonological representations, the more highly valued the solution. In English, the word chat has a complexity of 9. In a language where a corresponding word chat were possible, the word chat would have a much greater complexity, since so many of the feature specifications of /æ/ are predicted on the basis of the fact that only a vowel can follow word-initial /ɛ/ and /i/ in English. Similarly, in a language permitting other affricates (for example, /pʃ/, /tʃ/), /chat/ would be more complex ("cost more"), because so many of the feature savings in the above analysis depend on the absence of a full series of affricate consonants in English. Thus, by factoring out all of the redundancies from lexical entries, only the idiosyncratic (or unpredictable) features will have to be specified—and counted by the evaluation metric. In this way linguistically significant generalizations are captured by formulating morpheme structure rules which fill in blank (or zero) feature specifications.

A problem sometimes arising within this framework occurs when a feature specification [+F] can be predicted on the basis of a feature specification [+G], and vice-versa. Should [+F] be entered phonologically as [0 F], and be predicted on the basis of [+G], or should [+G] be entered phonologically as [0 G] and be predicted on the basis of [+F]? An example of this arises whenever a language has the typical five-vowel system:

i u e o a

There is a redundancy with respect to the features Back and Round. Both features are predictable in the /a/ case:

| 0 back | +back |
| 0 round | -round |

That is, since /a/ is the only [+low] vowel in the language, it is possible to predict both the [+back] and the [-round] specifications that make up
this vowel. In the [-low] vowels, it is not as straightforward. In the case of nonlow vowels, we have only two possibilities: front unrounded (that is, [+back, -round]) and back rounded (that is, [-back, +round]). There are no front rounded vowels (for example, /ü/) and no back unrounded vowels (for example, /u/) among the nonlow vowels. The question is, should we predict the frontness/backness on the basis of the roundedness/unroundedness, or should we predict the roundedness/unroundedness on the basis of the frontness/backness? We clearly cannot start with [o back, -low, -round], that is, with both features unspecified, since we would have no way of distinguishing [-back, -round] from [+back, +round] in underlying forms.

While phonologists have sometimes asserted that it is possible to determine which feature is dominant or more basic, it is sometimes impossible to provide evidence for choosing one feature over the other. In fact, it is entirely possible that neither feature determines the other, but rather that the two features determine each other. That is, the true generalization may be that the two features agree with one another, and not that one feature is distinctive and the other redundant. Such a notion of agreement of features is difficult to express within the framework of blank-filling morpheme structure rules.

For this and other reasons (mostly formal difficulties associated with MSRs), Stanley (1967) proposed that MSRs be replaced with morpheme structure conditions (MSCs). Stanley pointed out that the blank-filling morpheme structure rules are different from phonological rules in that only the latter are capable of changing features, deleting and adding segments, etc., while the former only express redundancies on the phonological level. In other words, MSRs are basically static in that they do not convert one level of representation into another, but rather simply enumerate the details of the phonological representation. Quite to the contrary, phonological rules convert phonological representations into phonetic ones.

Thus, a crucial distinction was drawn between a constraint on a given level of representation (for example, phonological or phonetic) and a rule converting one level of representation into another level. Morpheme structure conditions were designed to capture the redundancies of the underlying phonological level, but without the notion of blank-filling. Instead, blanks in the underlying matrices were prohibited, thereby making it impossible to have “archiphonemes,” that is, incompletely specified segments (see 3.2.2). While many phonologists still argue for archiphonemes (especially grammatical morphemes such as the incompletely specified /N/ aspect marker in Akan [Schachter and Fromkin, 1968] or the “floating” high tone /i/ associative marker in Igbo [Voorhoeve, Meeussen and de Blois, 1969; Welmers, 1970; Hyman, 1974]), virtually all generative phonologists have given up MSRs for MSCs.

Stanley (1967:426-428) enumerates three kinds of morpheme structure conditions: if-then conditions, positive conditions, and negative conditions. An example of an if-then condition can be found in the above language with the vowel system /i, e, u, o, a/, and is stated as follows:

\[
\text{If } V [-\text{low}] \downarrow \text{Then } [\text{a back}] \\][
\[\text{If } V [-\text{low}] \downarrow \text{Then } [\text{a round}]
\]

This example of a segmental MSC says that if a vowel is [-low] (in this case, anything but /a/), then the features Back and Round agree. This agreement is captured by means of the alpha variable notation. If \( \alpha = + \), then we obtain [+back, +round]; if \( \alpha = - \), then we obtain [-back, -round]. Nothing is said about whether one feature is predictable on the basis of the other. Instead, only the agreement (and not the exact content for any given morpheme) is revealed. This generalization, then, is said to capture a regularity in the underlying forms of this language.

Examples of sequential if-then MSCs were given in Chapter 1. Consider now the example from English stop + /l/ combinations. English allows initial /pl/, /bl/, /kl/, and /gl/, but does not allow */tl/ and */dl/ (for example, *play and *clay, but not *elay). A sequential if-then MSC can be written as follows:

\[
\text{If } C [-\text{cont}] \downarrow \text{Then } [-\text{cor}]
\]

If a word-initial noncontinuant is followed by /l/, then it must be either labial or velar—and not alveolar.

Positive MSCs are used to capture the canonical shapes of underlying forms. As is explicit in the term “morpheme structure condition,” this means the canonical shapes of morphemes. However, since grammatical morphemes (for example, noun prefixes, tense/aspect markers, inflectional markers), which are frequently affixes, often do not show the same phonological shape as lexical morphemes (for example, nouns, verbs, adjectives), it is clear that these regularities refer to so-called “content,” as opposed to “function,” words. The basic assumption in generative phonology has been that the lexicon consists of morphemes which by rules are combined into words. Positive MSCs have been used to capture the phonological shape of morphemes, rather than the derived shape of words. In a model of generative phonology recognizing the word as the structural unit of the lexicon, it would be quite consistent to distinguish between phonological and phonetic
word structure conditions, as opposed to morpheme structure conditions (see 6.1.2.1). An example of a positive MSC is the following from Igbo:

\[ + C(y) V + \]

Each (lexical) morpheme in Igbo consists of an initial consonant, an optional /y/, and a vowel, for example, /bá/ 'enter,' /byá/ 'come.' With few exceptions, morphemes are monosyllabic in Igbo, and the above formula captures the basic underlying generalization characterizing the language. The above positive condition is definitely a morpheme structure condition (that is, a condition on morphemes), since words can be longer than one syllable (and almost always are). Thus, since nouns are typically VCV, we therefore need a word structure condition on nouns of the following kind:

\[ \text{noun} \]

As stated above, many grammatical morphemes do not conform to the positive MSC given above. Thus, the infinitive prefix consists of the single vowel /i/, realized as [i] or [j], depending on vowel harmony.

An example of a negative condition is the following (where ~ = "not"):

\[ \sim \]

This MSC states that there is no phoneme /ŋ/ in this language. Schachter and Fromkin (1968) have suggested that negative conditions are not needed, since they can always be replaced by an if–then condition. This segmental condition can be restated as follows:

If : \[ \begin{array}{c}
+ \text{cons} \\
- \text{syl} \\
+ \text{nasal}
\end{array} \]

Then : \[ \begin{array}{c}
+ \text{back}
\end{array} \]

Thus, it may be that only positive and if–then conditions are required by the theory of morpheme structure conditions.

The abandonment of MSRs in favor of MSCs has had a serious effect on the evaluation of complexity in the lexicon. It is no longer possible to add up zeros and see what kind of a savings is attained by filling in feature values by rule. However, as Stanley (1967:434) himself points out, the savings that were possible in the MSR approach are still recoverable in the MSC approach. He suggests as an evaluation procedure that the "weight" or generality of a morpheme structure condition be judged by the maximum number of feature specifications that could theoretically be removed from an underlying matrix and predicted by an MSC. Thus, while the shift from MSRs to MSCs has actually been accompanied by a shift away from adding up points in underlying representations, the same procedure is still theoretically possible.

### 4.2.2 Feature Counting in Phonological Rules

Relatively little attention has in practice been paid to lexical feature counting as opposed to rule feature counting (see, however, the discussion of Harms, 1966, in 4.4.1). As will be discussed in 4.3, feature counting has had a profound effect on the whole conception of rules in phonology. The basic assumption is that a rule with fewer features specified is a simpler rule than a rule with more features specified. This assumption has led some phonologists to propose serious departures from the standard model of phonology. Thus, Contreras (1969:1) states: "Adherence to the binary principle in phonology conflicts with the simplicity criterion proposed by Halle, in the sense that rules which are intuitively more general are not consistently simpler than less general rules," that is, in terms of feature counting. This assumption has, on the other hand, been challenged by other phonologists. For example, Zimmer (1970:97–98) states: "The fairly widespread assumption that feature counting will automatically lead us to choose the preferable description from two or more competing ones, as long as they use the same features and the same conventions for writing rules, has never, to my knowledge, really been supported by detailed and convincing arguments...." Nevertheless, the idea of a simplicity metric based on feature counting, with the goal of distinguishing linguistically significant generalizations from spurious ones, is one of the trademarks of generative phonology.

It has already been demonstrated that a rule converting /k/ to [c̭] before /b, e, æ/ is simpler in the number of features required to specify it than a rule converting /k/ to [c̭] before /p, r, f/, and /a/. In this particular case, feature counting is capable of discriminating between possible and impossible phonological rules— or, in weaker terms, between "natural," and "unnatural" or "crazy" rules. On the other hand, feature counting has been used to distinguish between phonological rules which are both possible and natural.

The question here is which rule is simpler (more highly valued)?

Consider the following two rules of palatalization:

\[ a \text{ } k \rightarrow c / \text{ } - i \]

\[ b \text{ } k \rightarrow c / \text{ } - i / \text{ } æ \]

In terms of phonemes, rule \( a \) is much simpler than rule \( b \), since fewer symbols
are required (namely, three symbols, as opposed to five). However, when the two rules are translated into distinctive features,

\[
\begin{align*}
 a' \quad k & \to \xi / - [+\text{high}] - \text{back} \\
 b' \quad k & \to \xi / - [-\text{back}]
\end{align*}
\]

rule \textit{a} now requires two features to specify the environment of the rule, while rule \textit{b} requires only one. Thus, the simplicity metric says that rule \textit{b} is simpler than rule \textit{a}. In the sense of simplicity = generality, this is certainly the case, since the environment has been generalized in \textit{b} to include all front vowels. However, as we shall see in Chapter 5, rule \textit{a} is a much more frequent and "natural" rule than rule \textit{b}. If simplicity were the criterion used by children in acquisition, then we would expect rule \textit{b} to be more common. On the other hand, it is conceivable that the simplicity metric is not correct, because it should tell us that \textit{a} is simpler than \textit{b}. Other such examples of where the simplicity metric goes astray will be dealt with in Chapter 5.

Another problem inherent in the simplicity metric, as discussed so far, is that only distinctive features are counted. Special diacritic features such as [+ablaut] and [+noun], which are sometimes needed in phonology, as well as grammatical boundaries, must also be evaluated somehow. While certain proposals have been made much remains to be worked out in this area.

4.3 Consequences of Feature Counting

The decision to base one's judgment of simplicity on feature counting has to a great extent determined the history of generative phonology, since the aim is to make explicit what is a real generalization. In particular, the very design of phonological rules has been determined so as to \textit{minimize} the number of features which will be required to specify them.

4.3.1 Rule Formalisms

A number of formalisms have been introduced into the literature. These formalisms constitute tentative hypotheses concerning the nature of simplicity, which in turn provide an evaluation procedure by which a child, on being exposed to raw data, constructs a phonology of his language.

4.3.1.1 Feature-Saving Formalisms

Phonological rules are written in such a way that unnecessary repetition of feature values is avoided. For example, a rule such as

\[
\begin{align*}
 1 \quad & A \to B / - C \\
 1' \quad & A \to B / - C
\end{align*}
\]

which says that /A/ becomes [B] before /C/, is a simpler way of writing

\[
\begin{align*}
 2' \quad & A \to B / - C \\
 3' \quad & A \to B / - C
\end{align*}
\]

Since two segments, \textit{A} and \textit{C}, are changed, this rule is equivalent to the simultaneous application of two rules:

\[
\begin{align*}
 1' \quad & A \to B / - C \\
 3' \quad & A \to B / - C
\end{align*}
\]

Since \textit{two} separate conditions must be met in order for \textit{AC} to become \textit{BD}, rule \textit{2} represents a complexity over rule \textit{1}. However, in the formalization of the first rule as \textit{1'}, this difference in complexity is not revealed. On the other hand, by stating the environment once, as in \textit{1}, the simplicity metric assigns the right relative values to these rules.

A number of conventions are built into the rule formalism in just this way. Consider, for example, the following rule:

\[
3 \quad [+F] \to [+G] / [+H]
\]

A segment which is \textit{[+F]} acquires the feature specification \textit{[+G]} when it is found after a segment specified \textit{[+H]}. This formalism is quite different from one which is stated in terms of segments (for example, \textit{A \to B / C }). In the latter example, \textit{A} becomes \textit{B}, that is, it is no longer \textit{A}; in the above rule written in features, the segment marked \textit{[+F]} does in fact acquire the feature specification \textit{[+G]}, but it remains \textit{[+F]}. That is, one of the feature-saving formalisms is that features whose values do not change are not repeated on the right of the arrow. Stated somewhat differently, \textit{only those features whose specifications change are included on the right of the arrow}. Thus, the above rule is an abbreviation for the following:

\[
3' \quad [+F] \to [+G] / [+H]
\]
Rule 3 implies that the [+G] segments are still [+F]. It also implies that there were at least some instances of [+F, -G] segments prior to the operation of the rule. Thus, another feature-saving formalism is that when a feature change is stated on the right of the arrow, its opposite (input) value is not stated on the left of the arrow. Rather, it is implicitly there. Thus the original rule is actually an abbreviation for the following:

3'  
\[ [+F] \rightarrow [+F] / [+H] - \]

Finally, note that this rule implies that there were at least some instances of [-H] segments followed by [+F, -G] segments, in which case the input to the rule was not met. If one now states the most expanded redundant formalization of this rule, as follows,

3''  
\[ [+H] [+F] \rightarrow [+H] [+F] / [+H] -[G] \]

it is observed that instead of a rule consisting of three features (as obtained following the conventions just discussed), a rule consisting of six features must be written (if these conventions are not followed). The claim is that rule 3, which expands as in 3'', is more general than rule 4,

4  
\[ [+H] [+F] \rightarrow [+I] [+I] / [+H] -[G] \]

where six different features are involved. Unless the discussed feature-saving conventions are incorporated into the theory, rules 3'' and 4 will be judged of equal complexity by the simplicity metric, and a generalization will have been missed.

4.3.1.2 Abbreviatory Conventions While the above formalism is designed to capture the generality of a rule by minimizing the number of features which need to be expressed, additional conventions have been adopted whose effect is to collapse structurally similar rules into one rule. In a sense the distinctive features serve the purpose of collapsing three subrules into one rule. This is possible only when there is some structural similarity between the subrules, here meaning that the three segments in the environment of the rule constitute a “natural class.” We have already seen that it is only with great difficulty that three equivalent rules can be collapsed when the environment consists of /p/, /r/, and /a/.

The question arises whether two processes are subparts of the same rule or are two separate rules. While it is obvious that the above processes should be analyzed as subparts of one rule, it is equally obvious that the following two processes should be analyzed as separate rules:

6  
\[ k \rightarrow i - i \]

7  
\[ V \rightarrow V - N \]

The palatalization of /k/ to /c/ before /i/ has nothing structurally in common with the nasalization of vowels before nasal consonants. Hence the two rules are not collapsible. Thus, formalisms are sought which permit the collapsing of rules to achieve a real generalization, but which prevent the collapsing of rules when a spurious generalization would result.

4.3.1.2.1 Brace Notation While the palatalization and nasalization rules just given are structurally unrelated, the following two rules of Korean share obvious formal properties with one another:

8a  
\[ r \rightarrow n / \#\# \]

8b  
\[ r \rightarrow n / C \]

In 8a, /r/ becomes [n] at the beginning of a word (that is, directly following the full word boundary # #); in 8b, /r/ becomes [n] after a consonant. These rules put into effect the phonetic sequential constraint in Korean which disallows [r] except when preceded by a vowel (Hyman and Kim, in prep.). Thus, we observe in the following forms that the morpheme /rak/ 'pleasure'

\[ (that is, /æk/) \]

\[ (that is, /æ/) \]

\[ (that is, /ə/) \]
is pronounced [rak] after a vowel, but [nak] at the beginning of a word or after a consonant:

**UNDERLYING**  **PHONETIC**
/kʰwe rak/  /kʰwe rak/  'pleasure-pleasure'
/rak wan/  /nag wan/  'pleasure-garden'
/kuk rak/  /kuñnak/  'extreme-pleasure'

In order to capture the relatedness of rules 8a and 8b, the two rules are conflated into one by means of brace notation, as seen in 8c:

8c  \( r \rightarrow n \left\{ \begin{array}{c} C \\ \end{array} \right\} \)

The fact that 8a and 8b can be abbreviated as in 8c reveals that a solution which includes rules 8a and 8b is more "costly" than a solution with the one rule 8c. In terms of feature counting, the notation in 8c permits a great savings, since it is no longer necessary to state /r/ on the left of the arrow and [n] on the right of the arrow twice. That is, while seven symbols are required to state 8a and 8b, only five symbols are required in the single conflated rule in 8c. The theory therefore requires that 8c occur in the phonology of Korean, and not 8a and 8b. This requirement "forces" the preferred solution, since a phonological analysis with 8a and 8b would miss a generalization.

One of the requirements which must be met in order to conflate two phonological rules is that the rules be **structurally related**, as seen in the Korean example. A second requirement is that no third phonological rule be ordered between the two rules. That is, if it can be demonstrated that there is a third rule, say 8d, which must be ordered after 8a but before 8b, then this would constitute an argument against collapsing the two rules. We would in this case be dealing with two rules rather than one, although 8a and 8b would still exhibit striking structural similarities (namely, the fact that both convert /r/ to [n]). As will be seen in 4.3.2, most work in generative phonology is based on the position that phonological rules must be linearly ordered.

**4.3.1.2.2 Bracket Notation** Braces have been seen to involve a disjunction—thus, in the rule

9  \( A \rightarrow B \left\{ \begin{array}{c} C \\ D \end{array} \right\} \)

A becomes B before either C or D. That is, AC becomes BC and AD becomes BD. In a slightly more complicated example,

10a  \( \left\{ \begin{array}{c} A \\ E \end{array} \right\} \rightarrow B \left\{ \begin{array}{c} C \\ D \end{array} \right\} \)

In the derivation of [kuñnak], /kuk rak/ first becomes intermediate kuuknak by the rule changing /r/ to [n] after a consonant; then a second rule nasalizes /k/ to [g] before a nasal consonant, yielding [kuñnak].
The forms φo, kph, syd, and d'rá reveal that verb roots can have any one of three phonological shapes: CV, CLV, or CGV. In the nouns derived by reduplication of the corresponding verb root, the (prefixed) reduplicated syllable is always CV. Following the formalism developed for Akan by Schachter and Fromkin (1968), the Ewe reduplication rule can be written as follows:

13a \( \text{RED} \rightarrow C_{i} V_{1} / - C_{i} (\{L\}) V_{i} \)

where \( C_{i} = C_{i} \) and \( V_{i} = V_{i} \)

This rule copies a consonant \( (C_{i}) \) and vowel \( (V_{i}) \) identical to those found in the verb root. The parenthesis notation in the environment of this rule indicates the possibility of this rule applying to verb roots of the form \( C_{i} LV_{i} \) and \( C_{i} GV_{i} \). Notice also the brace notation indicating that the segment found between \( C_{i} \) and \( V_{i} \) in verb roots can be either a liquid or a glide. With these notations it is thus possible to state Ewe reduplication as one rule with the following three subparts:

13b \( \text{RED} \rightarrow C_{i} V_{1} / - C_{i} L V_{i} \)

13c \( \text{RED} \rightarrow C_{i} V_{1} / - C_{i} G V_{i} \)

13d \( \text{RED} \rightarrow C_{i} V_{1} / - C_{i} V_{i} \)

where \( C_{i} = C_{i} \) and \( V_{i} = V_{i} \)

4.3.1.2.4 Angled Bracket Notation Angled bracket notation is used to show an interdependency between two optional feature specifications. As an example, consider the following two rules from Nupe (Hyman, 1970b):

14 \( \{\epsilon\} \rightarrow [a] \)

15 \( \{\delta\} \rightarrow [\delta] \)

As originally argued in Hyman (1970a), the two abstract underlying segments \( /\epsilon/ \) and \( /\delta/ \) are realized phonetically as \([a]\), though they respectively palatalize and labialize the preceding consonant (see 3.3.5). In addition, underlying \( /\epsilon/, /\delta/, \) and \( /\dd/ \) are all realized as a nasalized schwa, that is, \([\dd]\). The rule changing both oral and nasalized low vowels to \([\dd]\) or \([\dd]\) is written as follows:

16 \( [+\text{low}] \rightarrow [+\text{back}] \)
\( \text{V} \rightarrow [+\text{round}] \)

A second rule now converts both \( /\dd/ \) and the cases of \([\dd]\) deriving from \( /\epsilon/ \) and \( /\delta/ \) to \([\dd]\), as follows:

17 \( [+\text{nasal}] \rightarrow [-\text{low}] \)
\( \text{V} \)

All nasalized vowels in Nupe are phonetically \([-\text{low}]\). We can now collapse 16 and 17 by means of angled bracket notation:

18a \( \{+\text{nasal}\} \rightarrow [+\text{back}] \)
\( \text{V} \rightarrow [-\text{round}] \)

This rule states that while all \([+\text{low}]\) vowels become \([+\text{back}, -\text{round}]\), if the low vowel is also \([+\text{nasal}]\) it must also become \([-\text{low}]\). That is, this rule schema collapses two rules. First, when the features within angled brackets are evaluated, the following rule converts low nasalized vowels to \([\dd]\):

18b \( [+\text{low}] \rightarrow [+\text{back}] \)
\( \text{V} \rightarrow [-\text{round}] \)

Second, when the angled brackets are not evaluated (since the interdependent features within them are optional), the following rule converts \( /\epsilon/ \) and \( /\delta/ \) to \([a]\):

18c \( [+\text{low}] \rightarrow [+\text{back}] \)
\( \text{V} \rightarrow [-\text{round}] \)

The ordering of 18b before 18c is dictated by the notation, which says first read the rule with the bracketed features and then read it without them. It should be clear that angled bracket notation also leads to an economy of features. Thus, the collapsed Nupe rule is stated with six features, while the two rules taken separately require seven features.

4.3.1.2.5 Alpha Notation Among the other feature-saving devices are alpha notation conventions. Suppose a language has the phonemic vowel inventory \( /i/, e, u, o, a/ \). A common redundancy is that nonlow vowels agree in backness and roundness. The vowels \( /i/ \) and \( /e/ \) are \([-\text{back}, -\text{round}]\),
while the vowels /u/ and /o/ are [+back, +round]. In the absence of an appropriate convention, two segmental constraints would have to be stated:

19a If :  
[−low ]
\[−\text{back}\]

Then :  
[−round]

19b If :  
[−low ]
\[+\text{back}\]

Then :  
[+round]

Eight features are required to state the two constraints. However, these two constraints are clearly related and should be stated as a single constraint. The use of phonological variables permits the collapsing of these two constraints as follows:

20 If :  
[−low ]
\[ x\text{back}\]

Then :  
[x round]

Now only four features are needed (though we avoid the problem of counting pluses and minuses as opposed to alphas). The alpha in this constraint means that either both are + or both are −, that is, all occurrences of alpha carry the same value. Some of the formal uses of variables are summarized below, alongside the feature values they abbreviate:

[aF, aG] :  
[+F, +G] or [−F, −G]

[aF, −aG] :  
[+F, −G] or [−F, +G]

[aF, bG] :  
[+F, +G], [−F, −G], [+F, −G], [−F, +G]

The notation [aF, aG] indicates that the two features must have opposite values, while [aF, bG] simply states that there is no required relationship between the specifications of the two features.

4.3.1.3 The Problem of Notational Equivalence Even with the well-defined formalisms so far developed by the theory, it sometimes is the case that a given phonological process can be formalized in more than one way. As an example, consider the following two structurally related processes:

21a  
/u/ → /o/ / − C S

21b  
/o/ → /o/ / − C S

In Fe?fe?-Bamileke, for instance, /u/ is realized as [ɔ] in “closed” syllables (that is, in syllables ending in a consonant) in many dialects, while /o/ is realized as [ɔ] in closed syllables in all dialects. These two processes can be abbreviated either with angled bracket notation or alpha notation:

22a  
[−low ]
\[−\text{high}\]
\[+\text{low}\]

Rule 22a has the following two expansions:

22b  
[−low ]
\[−\text{high}\]
\[+\text{low}\]

22c  
[−low ]
\[−\text{high}\]

The first expansion converts /o/ to [ɔ], and the second expansion converts /u/ to [ɔ]. 23a has the following two expansions:

23b  
[−low ]
\[+\text{high}\]
\[−\text{low}\]

23c  
[−low ]
\[−\text{high}\]
\[+\text{low}\]

The first expansion converts /u/ to [ɔ] and the second expansion converts /o/ to [ɔ].

In terms of redundancy, both rule formalisms are overspecified. In both 22b and 23c [−high] appears on both the left and right of the arrow, although generally only feature changes are expressed in phonological rules (see 4.3.1.1). This is unavoidable, if the lowering of /u/ and /o/ are to be captured in one rule with the features High and Low. In any case, where there are alternative ways of writing a rule, evidence must be sought to determine which formalization is correct.

It might be argued, on the other hand, that two formalisms, for example, angled bracket and alpha variable notation, are equivalent, that is, they make the same claims about phonological structure, and it will therefore be impossible to argue for one over the other. In attempting to choose one
formalism over the other for the vowel-lowering example, one quickly becomes embroiled in a number of theoretical issues. In 22a, /o/ is first lowered to [ə] and then /u/ is lowered to [o]. The second expansion applies "vacuously" to the [ə], which is derived from the first expansion, since [ə] is already [−high]. In 23a /u/ is first lowered to [o] and then /o/ is lowered to [ə]. However, the [o] which results from the first expansion must not undergo the second expansion or else underlying /u/ will also be realized as [ə].

It is therefore necessary to introduce a principle which has wide acceptance in phonological theory, namely that the two subparts of 23a are disjunctively ordered with respect to each other. If one expansion applies, the other expansion cannot apply to the same form (input). The opposite rule (or two rules which are conjunctively ordered with respect to each other) apply to the same form. From 22a one might conclude that 22b and 22c are either disjunctively or conjunctively ordered, since the same output is obtained in either case. Rules 22a and 23a are in this sense somewhat different.

While the formalism in 23a imposes disjunctive ordering, that in 22a does not. Should it ever be demonstrated that phonological rules are not disjunctively ordered, then the formalism in 22a would necessarily be chosen over that in 23a. On the other hand, it should be noted that rules collapsed by angled brackets have also been claimed to be disjunctively ordered, a position which in the light of evidence can always be reversed.

4.3.1.4 Summary In this section we have seen how various abbreviatory conventions lead to an economy in the number of features required to describe phonological processes. Among the formalisms discussed were brace notation, bracket notation, parenthesis notation, angled bracket notation, and alpha notation. A final formalism, which does not in itself reveal greater simplicity in terms of feature counting, allows us to rewrite rule 24a as 24b:

\[
24a \quad \text{[+F]} \rightarrow \text{[+H]}
\]

\[
24b \quad \text{[+F]} \rightarrow \text{[+H]} / \text{[+G]}
\]

Rule 24a says that a segment which is [+F, +G] changes an understood [−H] specification to [+H]. 24b says that a [+F] segment also becomes [+H] if it is [+G]. That is, placing an environment bar over a feature value indicates that this feature value is part of the specification of the input segment. While these rules are exactly equivalent and involve three features each, this convention is particularly revealing in collapsing rules such as 24b with other rules. Let us say, for instance, that the same language in question has a rule of the following form:

\[
25 \quad \text{[+F]} \rightarrow \text{[+H]} / -X
\]

Rule 25 applies not only to a segment which is [+F, +G] but also to one which is [+F, −G]. While it is not readily collapsible with 24a, it can easily be collapsed by means of brace notation with the equivalent rule 24b:

\[
26 \quad \text{[+F]} \rightarrow \text{[+H]} / \{\text{[+G]}\}
\]

That is, [+F] becomes [+H] if it is either [+G] or followed by X.

As written, rule 26 requires four features, while 24b and 25 require six features in total. While there has been a saving of two features, the convention which allows 24a to be rewritten as 24b is not in itself a feature-saving notation.

4.3.2 Rule Ordering

Consider the following hypothetical rules taken from Schane (1969):

\[
27 \quad /t'i/ \rightarrow [tya]
\]

\[
/te/ \rightarrow [ta]
\]

\[
/ta/ \rightarrow [ta]
\]

\[
/te/ \rightarrow [ta]
\]

In this hypothetical language, when the high vowels /i/ and /u/ are followed by a nonhigh vowel (in this case /a/), they are converted into the respective glides [y] and [w]. Whenever a nonhigh vowel is followed by a vowel, it is deleted. These two processes can be formalized as follows:

\[
28a \quad \text{[+high]} \rightarrow \text{[−syll]} / -V
\]

\[
28b \quad \text{[−high]} \rightarrow \emptyset / -V
\]

As written above, the two rules require eight features (if we count V and Ø as one feature each). These rules can also be applied in either order, since their environments are mutually exclusive. Thus, if 28a is applied first, then /t'ia/ and /t'ua/ become [tya] and [twa], and then 28b applies, converting /te'a/, /to'a/, and /ta'a/ to [ta]. If 28b is applied first, then the same results are obtained, but in reverse order.

9 It turns out that some dialects of Fe?fe? allow historical /tum/ to become /tum/ (and even /tam/). While it has been argued (Hyman, 1972b) that historically *tùm became [tüm] and then optionally lowered again to [tum], the difficulty in discussing this reapplication of the lowering rule in a synchronic framework arises from the problem of maintaining /tüm/ as the underlying form in all dialects. Since there is no alternation, once *tùm is pronounced [tüm] it can just as well be recognized as underlying /tüm/.

10 It should be noted, however, that McCawley (1971) has argued against the use of braces in phonology, especially in such cases as 26.
Such a solution, therefore, does not require any constraint on rule ordering. The rules are written out in such a form that they can be applied in random sequential ordering; that is, whenever the appropriate input is met, they apply.

On the other hand, imposing a definite (or extrinsic) ordering on rules allows us to simplify their structural description, sometimes dramatically. Thus, imagine that rule 28a were to apply before rule 28b. This would mean that all [+high] vowels followed by a vowel would be converted into glides before the operation of 28b, which deletes nonhigh vowels before vowels. Since this is the case, the feature specification [-high] is redundant in 28b. Instead, 28b should be written as follows:

\[ 28b' \quad V \rightarrow \emptyset / - V \]

Since the only VV sequences which could possibly serve as input to this rule have the first vowel [-high] (because of the prior operation of 28a), the correct output is obtained. And in the process one feature specification, namely [-high], is economized.

An appropriate example of this relationship between rule ordering and simplicity comes from Shona (Tom Hinnebusch and Theo Vennemann, personal communication). In this Southern Bantu language there are alternations between [p] and [h], [t] and [h], and [k] and [h], as informally represented by the following subrules:

\[
\begin{align*}
29 \quad p & \rightarrow h / m - \\
\quad t & \rightarrow h / n - \\
\quad k & \rightarrow h / \eta -
\end{align*}
\]

That is, voiceless stops become [h] after homorganic nasals. Since these nasal consonants derive from an underlying /n/ prefix, two rules are required: homorganic nasal assimilation and conversion of voiceless stops to [h]. If the rules are ordered, they can be specified as follows:

\[
\begin{align*}
30a \quad [+\text{nasal}] & \rightarrow [\alpha \text{place}] / - [\alpha \text{place}] & C \\
30b \quad [-\text{voice} - \text{cont}] & \rightarrow h / [+\text{nasal}] - & C
\end{align*}
\]

Rule 30a converts /n/ to [m] before labials and [ŋ] before velars. The notation [\alpha \text{place}] is an abbreviation for the place of articulation features, for example, [\alpha \text{ant}, \alpha \text{cor}], and should therefore be counted as several features rather than as one. Rule 30b says that a voiceless stop becomes [h] after a nasal consonant. Since 30a has already made all preconsonantal nasals homorganic, it is not necessary to specify the nasal consonant as to place of articulation. Thus the following derivations are obtained:\footnote{The rule of homorganic nasal assimilation in 30a applies vacuously to underlying /nt/, since the underlying nasal is already homorganic with the following voiceless stop.}

\[
\begin{align*}
31 \quad /np/ & \rightarrow mp \rightarrow mh \\
\quad /nt/ & \rightarrow nt \rightarrow nh \\
\quad /nk/ & \rightarrow \eta k \rightarrow \eta h
\end{align*}
\]

Rules 30a and 30b follow all the feature-saving conventions discussed in 3.1.1. This is made possible by the imposition of rule ordering. If 30b were to precede 30a, then the following would be the result:

\[
\begin{align*}
32 \quad /np/ & \rightarrow nh - \\
\quad /nt/ & \rightarrow nh - ? \\
\quad /nk/ & \rightarrow nh -
\end{align*}
\]

First, /p, t, k/ would be converted to [h], since they are found after a [+nasal] consonant. But then it is not clear how homorganic nasal assimilation would apply to intermediate nh, since the point of articulation of the following consonant is now glottal, and a "nasal glottal stop" is the only possible output of rule 30a. Thus it is clear that if these rules are to be ordered, 30a must precede rule 30b.

If, on the other hand, more information is incorporated into 30b, then rule ordering is unnecessary:

\[
\begin{align*}
30b' \quad [-\text{voice} - \text{cont}] & \rightarrow h / [+\text{nasal}] - & C
\end{align*}
\]

Rule 30b' says that a voiceless stop becomes [h] when it is preceded by a homorganic nasal consonant. Since the output of 30a is now incorporated into the input of 30b', it is no longer necessary that the two rules have the ordering restriction placed on. Instead, the two rules can apply whenever their structural description is met. If 30a applies before 30b', then of course the derivation is straightforward. If 30b' applies first, then it can only apply to /nt/, since this sequence alone has a consonant following a homorganic stop (as opposed to /np/ and /nk/). But then, after 30b' has applied, 30a can apply and convert /np/ and /nk/ to mp and \eta k, respectively, and now 30b' can reapply, as in the following derivations:

\[
\begin{align*}
33 \quad /np/ & \rightarrow \rightarrow mp \rightarrow mh \\
\quad /nt/ & \rightarrow nh \rightarrow \eta k \rightarrow \eta h
\end{align*}
\]
In the first approach, each rule is designed to apply once at a specific point in the derivation. In the second approach, a rule can apply any time its structural description is met, randomly until there are no longer any forms which are subject to it. This may mean that the rule will apply several times before it has run its course. Even the first approach has recognized the need for so-called "persistent" rules (see Chafe, 1968:131), which can apply at several points in a derivation. In the second approach, all rules operate in this manner.

The consequences are significant. First, while 30b requires only six features to specify it, 30b' requires eight features. Thus, if rules are to be randomly ordered, it will be necessary to complicate the rules—and, most likely, to give up the evaluation measure as so far conceived. Second, the random sequential ordering approach seriously affects the abstractness of underlying forms (see 3.3.5). Consider, for example, the following situation, which is found in Sea and Land Dayak.

Sea Dayak has two rules: 34a, a rule nasalizing vowels after nasal consonants; and 34b, a rule deleting voiced stops after homorganic nasal consonants (see Kisseberth, 1973a:427-428):

\[
\begin{align*}
34a & : & V \rightarrow \tilde{V} / N - \\
34b & : & [b] \rightarrow \emptyset / [m] \\
& & [d] \rightarrow O / [n] \\
& & [g] \rightarrow \emptyset / [\eta] -
\end{align*}
\]

Thus, /nana/ 'to straighten' is pronounced [nâ̄na], while /nanga/ 'to set up a ladder' is pronounced [nâ̄ngâ]. What this means is that the underlying contrast between \( O \) and /g/ is realized on the surface as a nasalized versus an oral vowel, a clear violation of the linearity condition rejected by Chomsky (1964:93). This state of affairs is adequately accounted for by requiring that 34a apply before 34b, as in the following derivations:

\[
34a \quad 34b
\]

\[
\begin{align*}
35 & \quad /nana/ \rightarrow n\dot{a}na \quad \text{‘to straighten’} \\
& \quad /nanga/ \rightarrow n\dot{a}nga \rightarrow n\ddot{a}nga \quad \text{‘to set up a ladder’}
\end{align*}
\]

If, however, 34b were to apply before 34a, then 34a would incorrectly nasalize the second vowel of /nanga/:

\[
34b \quad 34a
\]

\[
\begin{align*}
36 & \quad /nana/ \rightarrow \rightarrow n\dot{a}na \quad \text{‘to straighten’} \\
& \quad /nanga/ \rightarrow n\ddot{a}nga \rightarrow *n\ddot{a}nga \quad \text{‘to set up a ladder’}
\end{align*}
\]

Since phonological rules have access in this theory only to the immediately preceding stage of the derivation, there is no way to nasalize /nana/ without nasalizing the second vowel of the intermediate form naga derived by 34b. Instead, 34a must precede 34b.

Unlike the previous case, there is no way that the rules can be rewritten with random sequential ordering and still maintain the underlying forms /nana/ and /nanga/. The reason is that once /nanga/ becomes naga, there is no way short of rule ordering (but see 4.3.3) to keep 34a from applying to it to yield the incorrect *[nâ̄â̄na] ‘to set up a ladder.’ A theory not allowing extrinsic rule-ordering can be salvaged, however, by recognizing forms such as ‘to straighten’ with underlying vowel nasalization, that is, /nâ̄ngâ/. In this case, 34a is replaced with a phonological constraint stating that in underlying forms, vowels after nasal consonants are automatically [+nasal]:

\[
37 \quad \text{If } V : N V \quad \text{Then } [+\text{nasal}]
\]

This is not necessarily undesirable in itself. In fact, as noted in Chapter 3, there has been a recent shift toward less abstract phonological representations. In this case, adhering to random sequential ordering decreases the distance between the phonological and phonetic representations (see Vennemann, 1973).

There has been considerable discussion concerning the need for extrinsic rule ordering. The original conception of a sequence of ordered rules, each applying once in a derivation, has been seriously challenged (Koutsoudas, Sanders and Noll, 1974; Vennemann, 1973). A distinction has been drawn between intrinsic and extrinsic rule ordering. Intrinsic ordering is that imposed by the system of rules itself; given the form of the two rules, they can only be applied in one way. Extrinsic ordering is imposed by the language in question; given the form of the two rules, one must consult the particular data to see if a given rule precedes or follows another rule.

In order to explicate these notions, it is necessary to draw another distinction often made in the study of rule-ordering relations. Kiparsky (1968b) draws the distinction between feeding and bleeding rule ordering (see Chafe's [1968] equivalent distinction between additive and subtractive rule ordering). A rule a is said to feed into a rule b when it creates new environments for b to apply to. Thus, if [na] deriving from /nga/ were to become [nâ̄], one could say that the rule deleting /g/ feeds into the rule nasalizing [a] to [â̄] after nasal consonants, since it creates new environments for the latter rule's application. (Of course, we saw that this was not the case.) A rule a is said to bleed a rule b if it removes environments that could have undergone rule b. Thus, if our hypothetical language had a rule of the form
by which vowels are deleted word-finally after \([\eta]\), this rule would bleed the rule that nasalizes vowels after nasal consonants, since if this rule had not applied, the vowel in question would have undergone the nasalization rule.

Having drawn this distinction, it is now possible to distinguish absolute feeding and absolute bleeding relationships. A rule \(a\) is said to Absolutely feed a rule \(b\) if it creates all of the inputs to rule \(b\). A rule \(a\) is said to Absolutely bleed a rule \(b\) if it removes all of the inputs to rule \(b\). Absolute bleeding must of course be prohibited in phonology, since if one rule removes all of the inputs to another rule, then there is no need for the second rule. Thus, if we have two rules, and if they stand in a potential absolute bleeding relationship, they must automatically be reordered so that the more general rule applies after the less general rule. This is one type of intrinsic rule ordering.

A good example of such a possibility comes from Schane’s (1968) analysis of French. Two rules are relevant:

\[39a\]

\[V \rightarrow \bar{V} / - N\]

\[39b\]

\[N \rightarrow \emptyset / - S\]

Rule 39a says that a vowel is nasalized before a syllable-final nasal. Rule 39b says that a syllable-final nasal is deleted. If 39a precedes 39b, then the following derivations are correctly predicted:

\[40\]

\[\text{/bon/} \rightarrow \text{b5n} \rightarrow \text{b5} \quad \text{‘good’ (m.)}\]

\[\text{/bonet/} \rightarrow \text{b5nte} \rightarrow \text{b5te} \quad \text{‘goodness’}\]

If, on the other hand, 39b were to precede 39a, the following incorrect derivations would be obtained:

\[41\]

\[\text{/bon/} \rightarrow \text{*bo}\]

\[\text{/bonet/} \rightarrow \text{*bote}\]

That is, the syllable-final nasal would be deleted, and the rule nasalizing vowels before syllable-final nasals would have nothing to apply to. In other words, this would be a case of absolute bleeding, and could therefore not possibly be correct. Thus, given that we know French to have the two rules 39a and 39b, there is only one possible ordering of these rules. In this sense, the ordering can be said to be intrinsic. The rules order themselves.

While this is the definition of intrinsic ordering used by Schane (1969), others have restricted this term to apply only to cases of absolute feeding. Since the above rules can possibly give the wrong output, it is necessary according to this second view to modify 39b so as to permit random sequential ordering. This can be done by incorporating the output of 39a into the input of 39b:

\[39b^*\]

\[N \rightarrow \emptyset / \bar{V} - S\]

Rule 39b’ now states that a syllable-final nasal drops, but only when preceded by a nasalized vowel. Since, according to Schane’s analysis, there are no underlying nasalized vowels, all nasalized vowels derive from 39a. In other words, 39a absolutely feeds 39b. It must apply before 39b or else 39b will have nothing to apply to. But in the random sequential ordering (intrinsic ordering), if 39b is selected first, it does not apply; 39a then applies and creates nasalized vowels; now 39b applies, and so on. The result, again, is that the rule must be complicated to include mention of the nasalized vowel preceding the syllable-final nasal consonant.

While extrinsic rule ordering can be seen as a means of minimizing the number of features required to specify a rule, the more crucial question arises whether there are rules that can only be accounted for by such rule ordering. One case, originally cited by Chomsky (1964:96) in his demonstration against the linearity condition, concerns the pronunciation of the English words writer and rider. Many American English speakers pronounce these words [ra:tor] and [ra:tor], that is, with a vowel length contrast, but no consonant contrast phonetically. Two rules are required:

\[42a\]

\[V \rightarrow V: / - [+\text{voice}]\]

\[42b\]

\[\begin{array}{c}
[t] \\
[d]
\end{array} \rightarrow t / V - V\]

First, a vowel becomes lengthened before a voiced consonant, and second, /t/ and /d/ become \([\text{r}]\) (a voiced tap) intervocalically, when the first vowel is stressed. If the rules are ordered 42a–42b, then the forms [ra:tor] and [ra:tor] are obtained. If they are ordered 42b–42a, then the forms [ra:tor] and [ra:tor] are obtained. Since both possibilities exist, depending on the dialect, it is impossible to determine the ordering intrinsically, that is, on the basis of the form of the rules alone. Rather, one must extrinsically impose the rule ordering depending on which dialect one is describing.

Perhaps one way of avoiding extrinsic ordering in the first dialect (with a vowel-length contrast) is to recognize /at/ and /a:/. Such an analysis is argued by Vennemann (1972d, 1973). For an alternative approach, see Koutsoudas, Sanders and Noll (1974), who argue for simultaneous rule application, maintaining the notion that rules apply only once in a given derivation. Despite all the current research into the nature of rule ordering, the issue seems far from settled.

### 4.3.3 Global Rules

A number of recent studies have proposed that languages have rules which can refer back to earlier (often erased) stages of a derivation (Kisseberth, 1973a,b). In the standard approach to generative phonology, all that is necessary for the application of a phonological rule is the information put
into it from the immediately preceding stage in the derivation. In this modified approach, information from the systematic phonemic level is available at all stages of the derivation. For example, while an earlier rule can delete a vowel in a certain context, a later rule may have to make reference to this vowel, even when it is no longer present at the stage where this later rule applies. This kind of rule is termed a global rule.

While the status of global rules is being debated in current phonological discussions, the effect of this powerful device on phonology is clear. While global rules would still permit the kind of abstract phonological representations made possible by extrinsic rule ordering, it would now be difficult to make any solid argument for such rule ordering— if this alternative is available. Returning to the Sea Dayak example, Kisseberth (1973a:428) and Dinnsen (1974:38) argue that vowel nasalization should be treated as a global rule. As stated by Kisseberth (1973a:428): “a vowel nasalizes in Sea Dayak after a nasal element provided that nasal element does not arise as a consequence of the simplification of clusters of nasal plus voiced stop.” Nasalization of the second vowel of /nangaf/ ‘to set up a ladder’ will therefore never occur, since there is an underlying /ŋa/ between the nasal element /ŋ/ and the potentially affected vowel /a/. Similarly, in the writer:rider distinction, vowel-lengthening before a voiced consonant could be blocked before a voiced consonant which was not voiced at an earlier stage (presumably in the phonological representation). Thus, it appears that global rules can replace extrinsic rule ordering. It is possible that such rules do exist, since the implication is that speakers have access to underlying forms at all stages of the derivation. If the underlying forms are indeed “psychologically real,” then this seems to be a reasonable claim to make.

4.4 An Evaluation of Feature Counting

As has already been said, there is much disagreement over the validity of a simplicity metric based on feature counting. While some phonologists would advocate the rejection of this notion entirely, other phonologists would simply assert that because of serious flaws (see Chapter 5), the simplicity metric should be modified or refined.

The idea of basing one’s judgment of the simplicity of a given analysis on feature counting has serious consequences, since it makes certain claims about the nature of language and human language ability. For example, consider two solutions for the same language, which recognize the following vowel systems:

\[
\begin{array}{c}
\text{i} & \text{u} & \text{i}
\end{array}
\]

\[
\begin{array}{c}
\text{e} & \text{o}
\end{array}
\]

\[
\begin{array}{c}
\text{a}
\end{array}
\]

The first solution posits the three-vowel system /i, u, a/, the second the five-vowel system /i, e, u, o, a/. Now, let us say that this language has a rule palatalizing /k/ to [c] before /i/. In terms of distinctive features, the same rule would be expressed as 43a in the three-vowel system and as 43b in the five-vowel system:

\[
43a \quad k \to \mathbf{c} / -[\text{-back}]
\]

\[
43b \quad k \to \mathbf{c} / -[\text{+high}][\text{-back}]
\]

That is, since in the second solution there is a mid front vowel /e/, it is necessary to include two feature specifications, [+high] and [-back]. The first solution requires only one specification, namely [-back], since there is only one front vowel in the language. Thus, the same rule costs one feature more in the second solution, solely because of the inventory of segments. This is the claim that is made by feature counting. Feature counting always favors more general processes, and where a process is restricted (for example, to only high front vowels), a cost is assigned to it. Perhaps this claim is correct. Like other claims inherent in the simplicity metric, it is subject to empirical verification.

4.4.1 One Phoneme or Two?

Every time a decision is made on the basis of feature counting, an empirical claim is made about language—and this claim must be carefully investigated as to its implications. One appropriate example of this concerns the question of whether a given phonological entity should be analyzed as one or two underlying phonemes (see 3.4.3). As proposed by Harms (1966) (and applied to Igbo by Carroll, 1970), such questions can be resolved by reference to the lexical complexity of the two solutions. Harms argues that in one language it may be more economical to set up /Ch/ vs. /C/, or /C/ vs. /Ch/, in another language it may be more economical to set up /Ch/ and /C/, that is, sequences of two phonemes. As noted in the previous chapter, this question was of considerable importance in phonemic analysis.

Working within the framework of morpheme structure rules (see 4.2.1.1), Harms proposes that indeterminate cases be resolved by calculating the number of features that must be specified within the lexicon in both the one-phoneme and the two-phoneme solutions. Consider, for example, the difference between a solution that recognizes an aspiration contrast between voiceless stops, that is, /C/ vs. /Ch/, and a solution which recognizes a sequence of /C/ + /h/ vs. /C/. This second solution, according to the commutation test discussed in Chapter 3, would have credibility only if there were an independent /h/ in the language, which could occur even if not preceded by /C/.
Now, calculating the number of feature specifications required to determine a given consonant, Harms assigns the arbitrary integer \( n \) to /C/. That is, in the matrix for any given consonant, he is assuming that it will take \( n \) features to specify it. If this is the case, then if an additional feature, say [aspirated], is introduced, it will take \( n + 1 \) features to specify /C/, since it will now contrast with /Ch/, which also requires \( n + 1 \) features. The two will differ in that /C/ is \([-\text{aspirated}]\) while /Ch/ is \([+\text{aspirated}]\). Thus, each will take \( n + 1 \) features, or (taken together) \( 2n + 2 \).

In the /Ch/ vs. /C/ solution, /C/ requires \( n \) features, but now /Ch/ requires \( n \) features (for the /C/), plus however many features are required to unambiguously specify /h/ in the lexicon. Let us say that /h/ requires two features (for example, \([-\text{syll}, +\text{low}]\)). Now /Ch/ will require the \( n \) features for /C/ and two features for /h/, that is, \( n + 2 \). Since the nonaspirated /C/ also requires \( n \) features, /Ch/ + /C/ taken together require \( 2n + 2 \) features, just as in the one-phoneme solution.

Thus, in terms of economy (judged by the number of features which must be specified in the lexicon), the result is a standoff. Harms suggests, at this point, that the relative number of forms exhibiting \( /C^h/ \) or /Ch/ vs. /C/ be incorporated into the calculation. Let us say that in our corpus we have 100 forms with \( /C^h/ \) and 200 forms with /C/. We now calculate as follows:

\[
\begin{array}{c|c|c|c|c}
\hline
 & /C^h/ & /C/ & /Ch/ & /C/ \\
\hline
n + 1 & n + 1 & n + 2 & n \\
\hline
100 Ch & 100n + 200n+ & 100n + 200n & 100n + 200n \\
200 C & 100 & 200 & 200 & 200n+ \\
\hline
\text{Total} & 300n + 300 & 300n + 200 & 300n + 200n+ & 300n & 200n+ \\
\end{array}
\]

As seen from the above calculation, 100 feature specifications can be economized if the opposition is analyzed as one between /Ch/ and /C/. If, on the other hand, we had the opposite proportion (namely, 200 forms with \( /C^h/ \) and 100 forms with /C/) in our corpus, the following tabulation would give the opposite results, as seen below:

\[
\begin{array}{c|c|c|c|c}
\hline
 & /C^h/ & /C/ & /Ch/ & /C/ \\
\hline
n + 1 & n + 1 & n + 2 & n \\
\hline
200 Ch & 200n + 100n+ & 200n + 100n & 200n + 100n \\
100 C & 200 & 100 & 400 & 400 \\
\hline
\text{Total} & 300n + 300 & 300n + 400 & 300n + 100n & 300n + 400 \\
\end{array}
\]

In this case we can economize 100 feature specifications in the lexicon if we analyze the opposition as one between \( /C^h/ \) and /C/. In fact, it will generally work out to be the case that when the consonant with secondary articulation (for example, \( /C^h/, /C^h/ \)) occurs in more forms in the lexicon than the simple consonant, it will be more economical to analyze it as \( /C^h/ \), etc. Whenever it occurs in fewer forms than the simple consonant, it will be more economical to analyze it as /Ch/, etc. Thus this procedure suggested by Harms makes a very strong claim about the way language works—in particular the way children might go about constructing a phonology of their language. This approach claims that children will assess the numerical proportion of forms in assigning a phonological representation to the phonetic sounds they are exposed to.

The question of whether this claim is correct is, of course, difficult to answer. One can imagine various interferences or external factors that might have an effect on this analysis. For example, if /C/ occurs in more basic vocabulary and \( /C^h/ \) only in learned words, one might hypothesize that this could affect the analysis. Also, if the few words that have \( /C^h/ \) in a language are very frequent words, for example, function words like that, this, there, then, those, which all contain the rare English phoneme /ʃ/, this might also be a factor. Briefly, then, while simplicity has been put to the use of deciding between alternate solutions, in this case one vs. two phonemes, there seems to be little empirical support for either the criterion of simplicity or the more specific criterion of feature counting.

### 4.4.2 Derivational Constraints

However, the desire to make common or high valued phonological properties look simple formally has led to a number of other proposals. As will be seen in Chapter 5, Schachter (1969) proposes a formalism for natural rules which is designed in part to show the high value of certain kinds of assimilatory rules as opposed to others. To a great extent, the theory of markedness developed by Chomsky and Halle (1968), which is discussed also in Chapter 5, received its impetus from a desire to make the naturalness of segments, systems, and rules formally explicit.

A further example is provided by the work of Kisseberth (1970a) on phonological "conspiracies." Kisseberth points out that languages frequently have rules which "conspire" to turn out the same output. In Yawelmani, for instance, he describes a rule which deletes short vowels in the following environment:

\[ [-\text{long}] \rightarrow \emptyset \, /V \, C \_ \, C \, V \]

A vowel which is \([-\text{long}]\) is deleted if it is both preceded and followed by a vowel separated from it by exactly one consonant. Thus, a word of the form
CV₁CV₂CV₃ will become CV₁CCV₃ if V₂ is [−long]. If, on the other hand, there is no preceding vowel (that is, there is a word-initial consonant, #C−) or no following vowel (that is, there is a word-final consonant, −C#) the deletion will not occur. Also, if the preceding or following vowel is separated by two consonants, deletion will not occur. These constraints are designed to guarantee that no instance of #CC, CC#, or CCC will result from the deletion rule. These three disallowed sequences have in common the necessity of assigning two successive consonants to the same syllable. A word-initial consonant sequence is automatically syllable-initial, just as a word-final consonant sequence is automatically syllable-final. Finally, any sequence of three consonants must be syllabified with two of the consonants in one syllable. It thus appears that Yawelmani has a surface phonetic constraint against two successive consonants within a syllable.

There is a second rule in Yawelmani which is also related to this constraint (Kisseberth, 1970a:296). A rule of vowel epenthesis (which inserts [i] in the regular case, [a] in the irregular case) applies in the following environment:

\[ \emptyset \rightarrow V / C - C \{ # \} \]

A vowel is inserted in order to break up sequences of CC# and CCC (that is, sequences of two consonants within the same syllable). Kisseberth argues that the rule deleting [−long] vowels and the rule inserting vowels are functionally related in that their form depends crucially on the same phonetic constraint.

We have already seen that various notations have been devised to capture structural relatedness among rules, but there is no formalism to capture functional relatedness. In other words, the rule of vowel deletion (which "costs" seven features) could be just as related to the epenthesis rule as any other rule requiring seven features to specify it. In terms of simplicity, there should, according to Kisseberth’s argument, be some feature-saving formalism for the above two rules, since it should be easier for a child to learn two functionally related rules than two unrelated rules.

To achieve this end, Kisseberth introduces the notion of derivational constraints into phonology. There is a derivational constraint in Yawelmani to the effect that no rule may produce a sequence of #CC, CCC, or CC#. With this derivational constraint in effect, the rule of vowel deletion can be rewritten as follows (Kisseberth, 1970a:304):

\[ [−long] \rightarrow \emptyset / C - C \]

A short vowel is deleted between consonants—the vowels on the far sides of the consonants need not be included in the rule, since the derivational constraint will require that they be there anyway—or else the rule will not apply, since it will violate the constraint.

Thus, if the rule is rewritten in this fashion, the two features required to specify the two vowels in the environment of the rule can be economized. The rule now takes five features to specify it instead of seven. If the proposal were to stop here, the claim would be made that this rule is as related to the rule of epenthesis as is any other rule that takes five features to specify it. But, as Kisseberth hints, it may be possible to devise a formalism to take care of the epenthesis rule as well. In fact, since the epenthesis rule exists only to break up unacceptable clusters, perhaps the whole rule can be economized. Whenever a CCC or CC# is met in a derivation, a vowel is automatically inserted, having been triggered by the derivational constraint.

While derivational constraints have been proposed in a number of recent phonological studies, there is some question whether this functional relatedness between rules should be formally expressed. Herefore the collapsing of rules implied that the two processes were one (inseparable) rule. In this case, the two rules are not subparts of the same rule, but are quite different rules. As suggested by Kiparsky (1972), the bond between two functionally related rules does not seem to be as tight as that between two structurally related rules (which are collapsed). For example, a rule can be ordered between two functionally related rules, and it is apparently possible for a language to lose one rule without losing a functionally related rule. This question, like so many others, has yet to be resolved in phonological theory.