2 Phonological Rules and Representations

When generative linguists study the phonology of a language, they try to discover three kinds of generalizations. They look for regularities that help to define the language's inventory of phonological elements: its vowels, consonants, syllables, tones. They determine patterns in the distribution of these elements in the language's representations: may they appear or are they banned from initial, medial, final positions in the word, from stressed/unstressed syllables?; may an element of type *a* immediately precede/follow an element of type *b*?; and so on. Finally, they investigate alternations in the shapes of morphemes composed of these elements within the word and variant pronunciations of words within the sentence. The regularities that emerge from such study are assumed to be the joint product of the principles and parameters of Universal Grammar and the rules and representations that develop through the course of language acquisition. Although other more "experimental" approaches have been proposed from time to time (see Ohala 1986), the study of the "corpus-internal" generalizations just enumerated continues to be the major avenue of research into the speaker's internalized grammar.

The goal of this chapter and the next is to examine some of the basic questions that arise in the study of such generalizations, and the answers that generative phonology has given to these questions. These chapters also introduce the most important analytic techniques used to discover these regularities as well as the concepts and notation required to express them in the grammar.

2.1 Distinctive versus Redundant Features

Let us consider regularities of distribution first, beginning with a simple example from English. Study of English quickly reveals that the articulator feature in stops is in general unpredictable. Labial, coronal, and velar stops occur freely in initial, medial, and final positions. There are numerous examples of morphemes whose pronunciation differs solely in virtue of the point of articulation of the stop: for example, *pin, tin, kin, bun, dun, gun*. Specifications for the features [labial], [coronal], and [dorsal] thus are irreducible, arbitrary features of these particular lexical items. If one knew the rest of English grammar and did not happen to know the pronunciation of a word such as *ketch*, there would be no way to predict that the initial consonant has the value [dorsal]. It could have been [coronal] and pronounced *tetch*. Information of this kind is used to encode the vocabulary of the language and must be memorized in the course of language development.

Consequently, as in most other languages, the point-of-articulation features are *distinctive* for the stop consonants of English.

Now consider the laryngeal features that characterize the English stop system.

(1)

[− voiced, − spread gl]	p t k
[− voiced, + spread gl]	p^h t^h k^h
[+ voiced, − spread gl]	b d g

(2)

initial		medial		final	
[p^h]in	[b]in	ra[p]id	ra[b]id	la[p]	la[b]
[t^h]ot	[d]ot	a[t]om	A[d]am	ma[t]	ma[d]
[k^h]ap	[g]ap	jac[k]et	jag[g]ed	pic[k]	pi[g]

Examination of the data in (2) reveals an important difference in the status of the voicing and aspiration features. Ignoring for the moment contexts adjacent to a consonant, we see that the feature specifications for voicing (i.e., [+ voiced] and [− voiced]) have essentially free distribution: each appears initially, medially, and finally. There are many minimal pairs for this feature. The same is not true for aspiration ([+ spread gl]), however. Careful study of English shows that the distribution of this feature is severely limited. First, [+ spread gl] appears only on voiceless stops. All other consonants (as well as vowels) are [− spread gl]. In addition, [p^h,t^h,k^h] occur only at the beginning of a syllable. At any other position in the syllable voiceless stops are unaspirated. For example, the voiceless stops in *s*[p]*in*, *s*[t]*em*, *s*[k]*in* as well as *la*[p], *ma*[t], and *pic*[k] are not aspirated. Finally, another key distributional regularity is that unaspirated voiceless stops are not found in syllable-initial position (the case of *ra*[p]*id* is discussed below). The sound sets [p^h,t^h,k^h] and [p,t,k] thus have *complementary distributions:* they never appear in exactly the same phonological context.

These distributional regularities are summarized in (3).

(3) a. All segments except for voiceless stops are [− spread gl].
b. [p^h,t^h,k^h] only appear syllable-initially.
c. [p,t,k] do not appear syllable-initially, but freely occur in other positions in the syllable.

These results suggest that, unlike the features for place of articulation and for voicing, aspiration is entirely *redundant*. For any given English sound, we may predict its [± spread gl] value by determining whether it is a voiceless stop (if not, then automatically [− spread gl]); and if it is a voiceless stop, then its location in the syllable. Consequently, in learning English vocabulary, the developing grammar need not list the [± spread gl] value individually for each lexical item. If the generalizations in (3) are known, then in learning a new word such as *ketch*, the grammar can deduce that the initial consonant is aspirated, once having determined that it is a voiceless stop.

It is quite clear that these distributional generalizations follow from the English speaker's internalized grammar. For example, consider what typically happens when English speakers acquire a second language such as French, which lacks aspirated stops. In general, they have little difficulty in pronouncing and distinguishing such idiosyncratic features as [voiced] or point of articulation: they can differentiate *pas* [pa] from *bas* [ba] and *tas* [ta] from *pas* [pa]. However, English speakers typically pronounce French *pas* and *tas* as [p^{h}a] and [t^{h}a] instead of [pa] and [ta]. When this error is pointed out, they often have difficulty hearing the difference between their [t^h]*a* and the French [t]*as*. This pronunciation error cannot be explained simply by saying that unaspirated [p] and [t] are foreign sounds – English speakers have no trouble producing them in *s*[p]*in* and *s*[t]*em*. It's just that they have great difficulty pronouncing unaspirated [p] and [t] at the beginning of a French word. This difficulty can be explained by assuming that the generalization "syllable-initial voiceless stops are aspirated" is part of the grammar of English and that this rule is superimposed on the English speaker's pronunciation of French. In order to pronounce a foreign language accurately, one must often suppress the phonological rules of one's own language.

Slips of the tongue also argue that the principle governing the distribution of aspiration is part of the speaker's internalized grammar. When *tail spin* is transposed to *pail stin,* the aspiration on the voiceless stops is automatically adjusted to accommodate the new location of the shifted sounds so that [t^h]*ail s*[p]*in* → [p^h]*ail s*[t]*in*. If the distribution of aspiration is controlled by the rule that voiceless stops are aspirated syllable-initially, then the otherwise mysterious transformations of [p] to [p^h] and of [t^h] to [t] are explained.

Before asking how to express formally the rule governing the distribution of aspiration, let us consider another example of distinctive versus redundant features. At the level of detail we are considering here, the features defining the tongue body position for the vowel in *tab* [t^{h}æ:b] are idiosyncratic ones that must be memorized. The fact that the vowel of *tab* is [+low] is unpredictable. The word could have been pronounced with a [−low] vowel as [t^{h}e:b] or with a [+high] vowel as [t^{h}i:b]. No rule of English phonology predicts the [+low] and [−back] features in *tab.* On the other hand, the fact that the vowel of *tab* is pronounced long is principled. In stressed syllables, English vowels systematically lengthen before voiced obstruents and shorten before voiceless ones – as shown by minimal pairs such as *tab* [t^{h}æ:b] vs. *tap* [t^{h}æ̆p], *hid* [hɪ:d] vs. *hit* [hɪ̆t], *dug* [dʌ:g] vs. *duck* [dʌ̆k]. If the English speaker knows the rule that vowels lengthen before voiced obstruents and shorten before voiceless ones, then the length of the vowels in these words need not be memorized because it is predictable information.

2.2 Two Levels of Representation

The upshot of the above discussion is that, for English, aspiration in consonants and length in vowels have a different status from such features as consonantal point of articulation or vocalic tongue position. The latter are essentially unpredictable while the former instantiate systematic regularities in the sound pattern of the language. Data from second language acquisition and from speech errors support the contention that speakers of English tacitly know these rules. Granting this, the question now becomes, How is the distinction between the predictable (redundant, nondistinctive) and the unpredictable (contrastive, distinctive) features to be expressed? Generative grammar's answer to this question is based on the hypothesis that the human capacity for language is designed in such a way as to minimize the amount of information that must be stored in the speaker's mental lexicon. If storage space is at a premium, then the more information that can be predicted by simple and general rules, the more space will be available to store additional lexical items. In this way, the acquisition and seemingly effortless deployment of a large vocabulary becomes comprehensible.

If we accept this line of reasoning, then a particular approach toward drawing the distinction between the idiosyncratic and systematic features of pronunciation rather naturally emerges. Let us suppose that the grammar contains two levels of representation for phonological structure. An *underlying* or *phonological representation* will contain all and only the unpredictable (distinctive) information for each lexical item. Predictable features of pronunciation are added to the underlying phonological representation by grammatical rules and principles. These rules operate on the basis of the information in the lexical item's phonological representation and the context in which it is located. For each possible word constructed by the morphology and for each possible sentence constructed by the syntax, the phonological rules will thus "compute" or "derive" a (*surface*) *phonetic representation*. For example, the word *tab* will, at the level of discussion relevant here, have the underlying representation shown in (4).

(4)

	t	æ	b
consonantal	+	−	+
sonorant	−	+	−
continuant	−	0	−
coronal	+	0	−
labial	−	0	+
voiced	−	0	+
nasal	−	0	−
spread gl	0	0	0
low	0	+	0
high	0	−	0
back	0	−	0
long	0	0	0

The pluses and minuses represent unpredictable information that must be memorized when learning *tab*. It is in virtue of this information that one lexical item is distinguished from another. For example, the [−voiced] of the initial consonant distinguishes *tab* from *dab*. Changing the final consonant from [+voiced] to [−voiced] gives the phonological representation of *tap*. And changing the vowel from [+low] to [−low] gives *teb* – a possible but at present unoccupied slot in the English lexicon. The zeros represent information that is predictable and hence are not counted in calculating the complexity of a representation. Instead of being filled with zeros, the predictable values are often left blank to express more graphically the idea that this information entails no storage cost. The zeros (or blanks) will be filled in by rules and principles of the grammar. The discussion in section 2.1 mentioned two of these rules, which are stated in (5).

(5) a. $\begin{bmatrix} -\text{contin} \\ -\text{voiced} \end{bmatrix} \rightarrow [+\text{spread gl}]$ / syllable-initial

b. $\text{V} \rightarrow [+\text{long}]$ / ___ $\begin{bmatrix} +\text{cons} \\ +\text{voiced} \end{bmatrix}$

c. $\text{V} \rightarrow [-\text{long}]$ / ___ $\begin{bmatrix} +\text{cons} \\ -\text{voiced} \end{bmatrix}$

(5a) anticipates later discussion by stating the context for aspirated consonants as the beginning of a syllable. Since the first sound in any word normally begins the first syllable of the word, the context in (5a) will correctly assign aspiration in [tʰ]*ab* and will do much additional work later. This rule does not apply in *s*[t]*ab* since here the voiceless stop does not initiate the syllable. (5b) lengthens vowels before voiced consonants and thus applies in the derivation of *tab*, while (5c) assigns a shortened vowel in [tæ̆p]. Application of the rules in (5) thus changes the [0spread gl] feature of [t] to [+spread gl] and the [0long] of [æ] to [+long]. (In actuality, length is not a feature but an extra position (see section 6.9); we will overlook this fact here, however, since it does not materially affect the point that length is predictable from the voicing of the following consonant.)

Note that our discussion of *tab* is incomplete in several respects. We have not specified the final [b] with respect to aspiration. In addition, we have not specified the consonants [t] and [b] for the tongue body features [high], [low], and [back]; nor have we specified the vowel [æ] for voicing and nasality as well as the features [coronal] and [labial]. These features are predictable by a combination of UG and English rules. But unlike aspiration in voiceless stops and length in vowels, they are not assigned on the basis of the context in which the segment is located. Rather, they are implemented by a series of context-free rules assigning "default" values.

2.3 Default Rules and Unmarked Feature Values

Given that any phonological segment is defined by the UG set of features, we can view the construction of a phonological inventory as the grammar's utilizing a subset of these features to encode the language's lexical representations. Certain features are always selected (e.g., all grammars distinguish high from nonhigh vowels and coronal from noncoronal consonants); other features are language-particular choices that must be learned. For example, consider the laryngeal features of [voiced] and [spread gl]. French encodes its lexicon with the help of the [voiced] feature and thus opposes [t] to [d], while Mandarin Chinese employs [spread gl] to oppose [t] to [tʰ]. Thai utilizes both the [voiced] and the [spread gl] dimensions to make a three-way distinction among [t], [tʰ], and [d]. Hindi utilizes both dimensions fully in its four-way contrast of [p], [pʰ], [b], and [bʰ]. Some languages (e.g., Finnish) choose neither of these features to encode their lexical representations. An important insight due to the Prague School linguists Trubetzkoy and Jakobson is that the two poles of the [voiced], [spread gl], and [constr gl] features are not of equal status. In an important sense [−voiced], [−spread gl], and [−constr gl] are more basic values and denote more stable states. Linguists refer to the fundamental value of a feature as the *unmarked value*. For obstruents, the unmarked value for voicing is [−voiced]. For aspiration, it is [−spread gl]; [+voiced] and [+spread gl] are the *marked* values.

The marked-unmarked distinction for these features is reflected in a number of asymmetries. First, the unmarked values appear in all grammars. According to Maddieson (1984), all languages have at least some voiceless stops. While many languages supplement their stock of obstruents by adding further laryngeal distinctions of aspiration and glottalization, many others do not. For the latter, the unmarked [−voiced], [−constr gl], and [−spread gl] are the values chosen. Thus, Finnish stops are [p,t,k], not [b,d,g]. Second, according to Jakobson (1941), the unmarked values are the first to emerge in language acquisition and the last to disappear in language deficits. Finally, many languages neutralize underlying contrasts among the laryngeal features in particular positions such as the coda of the syllable or the end of the word. In general, it is the unmarked value that emerges in these positions. For example, while Thai contrasts [p,p^h,b] in the syllable onset, only [p] appears in the syllable coda.

Generative phonologists encode the marked-unmarked distinction by supposing that for each feature exhibiting such a distinction, there is a UG rule assigning the unmarked value. This implies the rules shown in (6).

(6) a. [−sonor] → [−voiced]
b. [−sonor] → [−spread gl]
c. [+sonor] → [+voiced]
d. [ ] → [−constr gl]

A number of consequences ensue from assuming the existence of such *default rules*. First, if a particular language does not utilize one of these features to encode its lexical representations, then the relevant default rule will automatically assign the unmarked value for that feature. (This conclusion is based on the assumption that the default rules are present in UG, and that every language develops from UG.) On the face of it, this is somewhat surprising. Since glottalization is essentially irrelevant in English, one might expect the articulators to randomly glottalize or not glottalize any given segment and that the system would simply overlook this feature. But this is not what is found. If a grammar fails to utilize a given phonological dimension to encode its lexical representations, the articulators still assume a definite state with respect to that dimension – in general, the unmarked one predicted by the rules in (6). In other words, in the absence of language-particular information, UG assigns a default value for each articulatory state. (Recent research suggests that this position may be too strong; see section 9.10 for discussion.)

To illustrate this point further, consider that the English plural suffix agrees in voicing with a preceding segment: *lip*[s], *cat*[s], *duck*[s], but *tub*[z], *bed*[z], *rug*[z].

Unlike obstruents, sonorant consonants in English have no underlying voicing opposition. Consequently, rule (6c) assigns these segments their [+voiced] value by default. Note that the plural suffix does not vacillate when added to *gun* or *bell*. It is uniformly pronounced with the same value that emerges after a voiced stop such as [b] or [g]: just as English speakers say *lab*[z] and *bug*[z], they say *gun*[z] and *bell*[z] – never *gun*[s] or *bell*[s]. Consequently, both underlyingly distinctive as well as default values can define the context for phonological rules.

Another consequence of assuming the rules in (6) is that if a particular feature encodes lexical items, then the corresponding default rule must be suppressed. Given that *bit* is lexically distinguished from *pit* by designation of [+voiced], the default rule (6a) must be prevented from applying. This is ensured by stipulating that default rules are restricted to filling in information that is lacking (so-called *feature filling*); they cannot change a feature that has been specified as plus or minus. Finally, even though a given laryngeal feature is not contrastive, that feature may nevertheless be assigned by a language-particular rule in the course of the phonological derivation. In general, we expect such rules to assign the marked value, because the unmarked value will emerge in any case by the UG default rule. For example, aspiration is not contrastive in English. Nevertheless, the grammar has a rule assigning [+spread gl] in syllable-initial position. The [−spread gl] feature is assigned in the complementary contexts by the default rule (6b).

Most of the other zero specifications for *tab* in (4) are assigned plus and minus values by UG default rules that supply the unmarked value for the relevant feature. For example, English does not employ the secondary articulations of palatalization, labialization, or velarization to distinguish lexical items. Consequently, the unmarked, minus values for the tongue body features [high], [low], and [back] are assigned to consonants regardless of context. Similarly, [−nasal] is the unmarked value for nasality in vowels and [+voiced] is the unmarked value for voicing in sonorants. UG default rules assigning these features determine the states of the glottis and the velum in the articulation of the vowel in *tab*. The result is the surface phonetic representation in (7) in which each of the three segments in the string is specified plus or minus for each of the features defined by UG.

(7)

	t^h	æ:	b
consonantal	+	−	+
sonorant	−	+	−
continuant	−	+	−
coronal	+	−	−
labial	−	−	+
voiced	−	+	+
nasal	−	−	−
spread gl	+	−	−
low	−	+	−
high	−	−	−
back	−	−	−
long	−	+	−

To briefly summarize the discussion to this point, we see that the surface phonetic representation is the product of three sorts of information: the unpredictable lexical specifications used to encode the vocabulary, language-particular rules supplying predictable feature specifications on the basis of context, and UG default rules assigning the unmarked value for nondistinctive features.

2.4 Excursus on Unmarked Values

The unmarked values and attendant default rules for the major phonological features are listed in (8). Note first that not all features seem to display the unmarked-marked asymmetry. In general, the major class features [consonantal], [sonorant] show no preference for one value over the other. Every language contrasts vowels and consonants and within the latter, sonorants and obstruents.

(8) a. <u>stricture features</u>

i. $[-\text{sonor}] \rightarrow [-\text{contin}]$

ii. $\begin{bmatrix} +\text{sonor} \\ +\text{nasal} \end{bmatrix} \rightarrow [-\text{contin}]$

iii. $[+\text{sonor}] \rightarrow [+\text{contin}]$

iv. $\begin{bmatrix} -\text{sonor} \\ +\text{contin} \end{bmatrix} \rightarrow [+\text{strid}]$

v. $[\quad\quad] \rightarrow [-\text{strid}]$

b. <u>laryngeal features</u>

i. $[-\text{sonor}] \rightarrow [-\text{voiced}]$

ii. $[+\text{sonor}] \rightarrow [+\text{voiced}]$

iii. $[\quad\quad] \rightarrow [-\text{spread gl}]$

iv. $[\quad\quad] \rightarrow [-\text{constr gl}]$

c. <u>nasality</u>

i. $[\quad\quad] \rightarrow [-\text{nasal}]$

d. <u>consonantal place features</u>

i. $[+\text{cons}] \rightarrow [+\text{coron}]$

ii. $[+\text{coron}] \rightarrow [+\text{anter}]$

iii. $[+\text{cons}] \rightarrow [-\text{high}]$
$\rightarrow [-\text{low}]$
$\rightarrow [-\text{back}]$
$\rightarrow [-\text{round}]$

e. <u>vowel quality</u>

i. $[-\text{cons}] \rightarrow [+\text{high}]$

ii. $\begin{bmatrix} -\text{cons} \\ -\text{low} \\ \alpha\text{back} \end{bmatrix} \rightarrow [\alpha\text{round}]$

(8a) records the default rules for continuancy and stridency. At the points of articulation where a stop-fricative contrast is phonetically possible, [−continuant] is unmarked: segmental inventories generally choose a stop before including a fricative; some languages of Australia lack fricatives entirely; and when a [±continuant] contrast is neutralized nonassimilatorily, it is typically in favor of [−continuant] (e.g., consonants closing the syllable in Korean must be [−continuant] (Kim 1972)). Rule (8ai) encodes the unmarked status of stops over fricatives. If postvocalic spirantization of stops (e.g., Spanish *Ma*[ð]*ri*[ð]) is an assimilatory phenomenon, then [+continuant] is redundant in vowels and may be assigned by (8aiii). The spirantization of [b,d,g] in Spanish also argues that nasals are [−continuant] while liquids are [+continuant]: *hom*[b]*re* 'man' vs. *cal*[β]*o* 'bald', *ver*[ð]*e* 'green'. Since default rules are feature-filling operations, applying (8aii) before (8aiii) preempts application of the latter to a nasal. For stridency, (8aiv) records the observation that the optimal fricative is [+strident]: languages prefer [s] to [θ]. Besides enhancing fricatives, this feature seems to play no role; all other segments are thus [−strident].

For the laryngeal features, we have observed that [+voiced] is unmarked for sonorants while [−voiced] is the optimal value for obstruents. Aspiration and glottalization are marked in both sonorants and obstruents. These observations imply the default rules in (8b). (8c) states that vowels and consonants are optimally oral; [+nasal] is always a marked feature.

For consonantal place of articulation, there is debate as to whether a particular value should be singled out as unmarked, and if so which one (see the discussion in section 9.11). The most popular choice is [+coronal]; (8di) reflects this point. Within the class of coronals, the [+anterior] dental-alveolar category is clearly unmarked in comparison to the [−anterior] alveopalatals. Phonological inventories select [t,s,n] before [č,š,ñ]. The battery of rules in (8diii) indicate that secondary articulations of labialization, palatalization, velarization, and so on, are marked. More generally, phonological segments with complex, multiple articulations are marked in comparison to singly articulated ones.

For vocoids (8e), mid [e] and [o] are marked in comparison to the high vowels [i] and [u]. Glides are typically [+high], and many languages either reject mid vowels entirely in favor of high or limit their appearance to "strong" positions such as stressed or initial syllables. It is difficult to determine an unmarked value for the [±back] feature. Finally, in nonlow vowels, rounding enhances backness: for back vowels [+round] [u,o] are chosen over [−round] [ɯ,ɤ], and [−round] [i,e] are chosen over [+round] [ü,ö]. In section 4.4 we will see the phonetic rationale for this enhancement relation between [round] and [back].

2.5 The Phoneme

One of the factors that initiated the development of phonology as a branch of linguistics distinct from phonetics was the discovery that native speakers often judge sounds to be identical that are clearly distinct phonetically – sometimes quite radically so. For example, in many dialects of American English the coronal stop [t] has as many as eight distinct pronunciations.

(9)	[t]	"plain"	stem
	[t^h]	aspirated	ten
	[ṭ]	retroflexed	strip
	[D]	flapped	atom
	[N]	nasal flap	panty
	[tʔ]	glottalized	hit
	[ʔ]	glottal stop	bottle
	[ ]	zero	pants

Such variants (*allophones*) of the same underlying sound (*phoneme*) are the product of systematic rules that modify the segment depending on the context in which it finds itself. (In structural and early generative phonology, phonemes were distinguished notationally from allophones by enclosing them in slanted brackets. One spoke of the phoneme /t/ with its allophones [t], [t^h], and so on. This notational distinction is no longer systematically enforced; it will be employed here on occasion, when the distinction between an underlying and a derived representation is crucial to the point under discussion.) We are in general unconscious of these rules. But if we come into contact with a speaker who fails to follow them, we may feel that he or she is not "one of us" and speaks with an accent or somehow sounds "funny." In some cases these variant pronunciations are quite puzzling. For example, although we perceive *tents* as having five sounds, phonetic instrumentation reveals that in fact no consonant appears between the [n] and the [s]. The same gap is found in *tends*, yet we feel that the two words are somehow still different. As we saw in the Introduction, Edward Sapir referred to these phenomena as "collective phonetic illusions" – we hear things that objectively are not there, we fail to notice elements that are present, and we judge sounds to be identical that are demonstrably quite different. Thus, in order to speak English, we must learn the rule that omits the [t] and [d] in *tents* and *tends*. Furthermore, it is by virtue of having internalized this rule that we can recognize the defective stimuli [t^{h}ɛ̃ns] as the plural of *tent* and [t^{h}ɛ̃:nz] as *tends*. The mysterious thing is how we learn these rules. No one has taught them to us. We are unable to discover them through introspection. Yet we all tacitly know them if we are native speakers of English. As stated in the Introduction, generative phonologists have set themselves the task of giving a serious, scientific answer to this question.

In this section we augment our sketch of English phonology by examining the rules that account for the varying realizations of the [t] phoneme. The retroflex stops – essentially equivalent to the [ṭ] and [ḍ] of Hindi – have a very limited distribution in English: they appear only before [r]. Compare the plain [t] and [d] of *s*[t]*ay* and [d]*ip* with the retroflex variants in *s*[ṭ]*ray* and [ḍ]*rip*. Since the rhotic [r] of *rip* is produced by curling back the tip of the tongue, it is natural to interpret this sound change as one in which the coronal stops assimilate the [–anterior, –distributed] features of the following [r] (IPA [ɹ]). The rule in (10) expresses this assimilation.

(10) $$\begin{bmatrix} -\text{contin} \\ +\text{coron} \end{bmatrix} \rightarrow \begin{bmatrix} -\text{anter} \\ -\text{distrib} \end{bmatrix} \Big/ \underline{\quad\quad} \begin{bmatrix} +\text{sonor} \\ -\text{anter} \\ -\text{distrib} \end{bmatrix}$$

Another "exotic" segment hidden inside English phonetics is a voiceless lateral, similar to the one found in Welsh. Careful listening to the following paradigms reveals that the liquids are voiceless [r̥] and [l̥] after voiceless stops.

(11)	[r]im	[r]ip	[l]ean
	b[r]im	d[r]ip	g[l]ean
	p[r̥]im	t[r̥]ip	c[l̥]ean
	f[r]om	sh[r]imp	s[l]ip

While we can easily write a rule to devoice liquids after voiceless stops, we may ask why the voiceless fricatives in *f*[r]*om*, *sh*[r]*imp*, and *s*[l]*ip* fail to trigger the change. An attractive answer is suggested by the observation that the voiceless stops in *p*[r̥]*im*, *t*[r̥]*ip*, and *c*[l̥]*ean* are aspirated. This observation in turn suggests that the real process underlying the liquid devoicing is one in which the liquids assimilate the [+spread gl] feature, which will naturally tend to suppress vocal fold vibration in a sonorant. We can test this alternative hypothesis because we know a context in which the voiceless stops are unaspirated: when they are preceded by [s]. The fact that *st*[r]*ip* rhymes with *d*[r]*ip* rather than with *t*[r̥]*ip* confirms this analysis. We express the rule generating the voiceless liquids as (12). It will also aspirate the [w] in *t*[w̥]*elve* and the [y] in *c*[y̥]*ue* (*cue*), whose glides are essentially equivalent to the aspirated ones in [w^h]*ich* and [y^h]*uge* (*huge*).

(12) [+sonor] → [+spread gl] / [+spread gl] ___

Besides the aspirated and the retroflexed variants, the phoneme [t] has two additional realizations in many American dialects: a glottalized version (13a) and a flap (13c).

(13) a. hi[t^ʔ], ho[t^ʔ], ou[t^ʔ]
b. bel[t], raf[t], ap[t]
c. hitter (hi[D]er), hottest (ho[D]est), outing (ou[D]ing)

The glottalized [t^ʔ] is articulated by making simultaneous closures at the glottis and at the alveolar ridge. The flap [D] (IPA [ɾ]) is a variety of [r] quite similar to the tap in Spanish *pero*. Comparison of the forms in (13a) and (13b) suggests that the glottalized allophone arises after a vocalic segment. When the items in (13a) are followed by a vowel, the [t] emerges as a flap (13c) in the dialect in question. Since the flap is a rhotic, and since in many languages [r] arises from an intervocalic dental obstruent (a form of lenition), we will posit a similar analysis for English. Note that the voiced stop [d] of *ride*, *mad*, and *bed* also flaps in the same context.

(14) ride, ri[D]er; mad, ma[D]est; bed, be[D]ing

There is a curious restriction on this lenition process in English. Comparison of the forms *átŏm* [D] vs. *prótòn* [t^h] and *Ádăm* [D] vs. *rádàr* [d] shows that flapping cannot take place when the vowel following the coronal stop is stressed. We know that in *atomic* the [t] is aspirated; and so it might be argued that the aspiration rule applies first to generate a [tʰ]. Flapping could then be defined over the [−spread gl] [t] in the intervocalic environment. However, the fact that the unaspirated [d] of *rádàr* is not flapped either suggests that stress is involved directly in the flapping rule. As we will see in chapter 10, the combination of a stressed plus unstressed syllable forms a metrical constituent in English known as a *foot.* A number of other rules of English phonology are sensitive to this environment: *cónsti*[tʰ]*ùte*, but *constí*[č]*uent*. We therefore express the flapping rule as (15), requiring the flanking vowels to belong to the same metrical foot.

(15) $\begin{bmatrix} -\text{contin} \\ +\text{coron} \end{bmatrix} \rightarrow [+\text{sonor}] \, / \, V_1 \,\_\_\_\, V_2$

(where V_1 and V_2 belong to the same foot)

The rule will now apply to the intervocalic stops in *átom* and *Ádam* but not in *prótòn* and *rádàr* because the flanking stressed vowels of the latter pair belong to separate metrical feet.

Let us now turn to a problem posed by *rá*[p]*id*, *rác*[k]*et* (and *á*[t]*om* in the dialects without flapping). The lack of aspiration in these cases has traditionally been explained by restricting the aspiration rule to the onset of stressed syllables. However, two facts noted by Kahn (1976) call this analysis into question. First, the initial consonants in *Pacífic, terrífic,* and *collápse* are generally perceived as aspirated even though the following vowel is unstressed. A rule assigning aspiration to the onset of stressed syllables would incorrectly overlook these cases. Second, the data in (16) suggest that liquids devoice after a voiceless stop even when they occupy the unstressed syllable of a metrical foot. If the earlier hypothesis that voiceless liquids reflect aspiration of the preceding stop is correct, then aspiration must be assigned to the medial voiceless stops in (16) even though they initiate a stressless syllable.

(16) áp[r̥]on
mát[r̥]on, mát[r̥]ess
ác[r̥]id, ácc[l̥]imate

These arguments indicate that the original generalization – aspirate a syllable-initial voiceless stop – is correct. To explain the lack of aspiration in *rá*[p]*id*, *á*[t]*om*, and *rác*[k]*et*, we can suggest that they reflect another "branch" of the intervocalic lenition process that operates in the flapping dialects. Let us therefore posit the rule in (17).

(17) $[-\text{contin}] \rightarrow [-\text{spread gl}] \, / \, V_1 \,\_\_\_\, V_2$

(where V_1 and V_2 belong to the same metrical foot)

Note that if we accept this analysis, then two rules are competing for the same input. Being syllable-initial, the aspiration rule would assign the [p] of *rapid* the feature [+spread gl]. But since the [p] of *rapid* also finds itself in the context of (17) that assigns [−spread gl], a contradiction arises. In the case of such competing rules, a principle (due originally to the Sanskrit grammarians) generally obtains according to which the rule applying in the more specific context "wins out." In this case, the general aspiration rule is preempted by the more specific lenition rule (17).

To complete the discussion of [t] realizations, note that the glottalization observed in (13a) is also found for many speakers on the [p] of *captain* and the [k] of *buxom*. We account for this fact by invoking the rule in (18), which assigns the feature of glottal closure to a voiceless stop when it belongs to the same syllable as the preceding vowel.

(18) $$\begin{bmatrix}-\text{contin}\\-\text{voiced}\end{bmatrix} \rightarrow [+\text{constr gl}]\ /\ \text{V}\ \_\_\_$$

(where V is tautosyllabic)

Our discussion of flapping abstracts away from certain complications. In many dialects the [t] of *capácity* is flapped (though less uniformly than the [t] of *atom*). In this case each of the flanking vowels is unstressed. If rule (15) is correct, then we might infer that the stressed antepenult heads a metrical foot that takes in both of the following unstressed syllables. Also, many speakers can extend the flapping process to intervocalic stops that arise in sentential contexts: for example, the final [t] of *hit* and *merit* can flap in *hit Ann* and *merit action*. In these cases the requirement that the flanking vowels belong to the same foot is dropped. But since the rule never extends to the [t] of *my tie*, we see that aspiration and flapping are still incompatible. For further discussion, see Kahn 1976, Selkirk 1982b, Harris 1990, Harris and Kaye 1990.

2.6 One or Two Levels of Representation?

The rules of English phonology examined so far (aspiration, flapping, retroflexion, and glottalization) define generalizations that are true of the surface phonetic level. For example, when we examine intervocalic position in the flapping dialect, we consistently encounter a [D] in place of [t] or [d]. We have characterized this state of affairs in terms of abstract representations of [t] that are then assigned the features [+spread gl], [+sonorant], [−distributed], [+constr gl] in the appropriate contexts to produce the allophones [t^h], [D], [t̪], and [$t^ʔ$]. However, it is not obvious that another representation distinct from the phonetic level must be invoked to explain these generalizations of English phonology. We could interpret a rule like (18) as simply an if-then declarative statement of the form "If a voiceless dental stop immediately follows a tautosyllabic vowel, then it is glottalized." Under such an interpretation, no generalization is missed or fact left out. To use a crude analogy, we may state the law that "All swans are white" without appealing to a colorless swan that is "assigned" the property white. As we have seen, generative phonologists have postulated abstract, redundancy-free representations on the assumption that the grammar is designed (through evolution) to minimize the amount of information that must be stored, allowing the development of a large lexicon (Bromberger and Halle 1989). However, with the advent of neural science and more accurate estimates of the capacity of the human brain, this "economy of storage" argument is not compelling in and of itself. Phonologists have proposed two additional types of argument in order to motivate the distinction between phonological and phonetic representations.

First, expressing the difference between distinctive and redundant information as the presence versus absence of material in the phonological representation provides a natural formal basis for understanding why redundant features are often inert. To take a well-known example, in Russian consonants the voicing feature is contrastive among obstruents but redundant in sonorants. The [+voiced] feature of liquids and nasals can thus be supplied by the UG default rule of (8bii). As the paradigms in (19) demonstrate, Russian consonants are devoiced at the end of a word (*vez-u* but *ve*[s]), and in consonant clusters the voicing of the final member is assimilated by nonfinal members (*vez-u* but *ve*[s]-*ti*). Sonorant consonants neither trigger nor undergo the rule (*nes-la, sn-a, son, rt-a*).

(19)	1sg.	vez-u	nes-u	
	masc. past	ve[s]	nes	
	infin.	ve[s]-ti	nes-ti	
	fem. past	vez-la	nes-la	
		'transport'	'carry'	
	nom.sg.	voš	rot	son
	inflected	[f]š-i	rt-a	sn-a
		'louse'	'mouth'	'sleep'

Furthermore, as shown by the preposition *iz-* 'from' in the phrases *iz Moskvi* 'from Moscow' vs. *i*[s] *Mcenska* 'from Mcensk', sonorants fail to interrupt the assimilation of voicing. This set of facts is naturally explained if the final devoicing and voicing assimilation rules operate over representations in which sonorants lack any specification for the feature [voiced]. But if these rules must be defined over the phonetic representations that contain the redundant [+voiced] on the nasals in *Mcensk,* it is unclear why this [+voiced] feature fails to interrupt the propagation of the voicing change through the cluster in *i*[s] *Mcenska* as well as why it neither initiates nor undergoes voicing assimilation itself.

A second reason to posit more than a single level of representation is simply that there are many empirical generalizations that cannot be stated over just a single level. The length of vowels before voiced versus voiceless consonants in English provides a well-known example. Recall that rule (2b) accounts for the difference between *cab* [kʰæ:b] and *cap* [kʰæ̆p]. As shown in (20), many additional pairs of words distinguished solely by the voicing of the final consonant exhibit this vowel length difference: *r*[ɪ:]*b* vs. *r*[ɪ̆]*p; b*[æ:]*g* vs. *b*[æ̆]*ck; b*[ʌ:]*zz* vs. *b*[ʌ̆]*s*. The feature worthy of note is that most dialects preserve the length distinction when the flapping rule merges the underlying contrast between [t] and [d]. This point is illustrated by the paradigms in (20). For speakers of this dialect, *writer* and *rider* are not homophonous. While the distinction is heard in the consonants, this is an illusion, since the words are pronounced the same in this position. Instead, they differ phonetically in the length of the preceding vocalic nucleus.

(20)	wet	[wĕt]	wetting	[wĕDɪŋ]
	wed	[wɛ:d]	wedding	[wɛ:Dɪŋ]
	seat	[sĭt]	seated	[sĭDəd]
	seed	[si:d]	seeded	[si:Dəd]
	write	[răi̯t]	writer	[răi̯Dər]
	ride	[ra:i̯d]	rider	[ra:i̯Dər]

If phonological rules are restricted to stating generalizations defined on and holding over the information contained in the phonetic transcriptions alone, then the vowel length in *rider* cannot be predicted. The phonetic representation of this word is identical to that of *writer* – save for the vowel length. To assign a long vowel in *rider* but not *writer*, the rule must "know" that the flap [D] in *rider* derives from a [d]. In other words, the grammar must define another representation in which the neutralized [t] vs. [d] contrast is restored. But this other representation is automatically supplied – if we agree to characterize the predictable features such as aspiration and glottalization in terms of redundancy-free underlying representations (UR). All we need assume is that the rule assigning vowel length applies before the sonorization rule neutralizing the voicing distinction. Under such an analysis, the following derivations (ordered sequence of rule applications) emerge for *write-writer* versus *ride-rider*.

(21)	/rai̯t/	/rai̯t-ər/	/rai̯d/	/rai̯d-ər/	UR
	răi̯t	răi̯tər	ra:i̯d	ra:i̯dər	length (5b,c)
	—	răi̯Dər	—	ra:i̯Dər	sonorization (15)

If the theory of grammar does not permit a second level of representation, then an important generalization of English phonology is missed. For the labial and velar stops in *r*[ɪ:]*b*-*r*[ɪ:]*bbing* vs. *r*[ĭ]*p*-*r*[ĭ]*pping* or *b*[æ:]*g*-*b*[æ:]*gging* vs. *b*[æ̆]*ck*-*b*[æ̆]*cking*, the presence versus absence of vowel length uniformly correlates with the voicing of the following consonant. Furthermore, this correlation holds for the coronals in *w*[ɛ:]*d* vs. *w*[ĕ]*t*. The one situation in which it appears to break down is before the flap. But this is precisely the context in which another rule neutralizes the contrast between voiced [d] and voiceless [t]. There is, moreover, a regular connection among the words in a paradigm built from the same stem. If a stem ending in [d] has a long vowel (e.g., *w*[ɛ:]*d*), then vowel length emerges in words formed by suffixation (e.g., *w*[ɛ:D]*ing*, *w*[ɛ:D]*ed*). This correlation is also explained if the vowel length generalization is stated before the voicing contrast is neutralized. But if the phonological model is required to define its rules over the phonetic level alone, then the generalization will be lost. Vowel length becomes unpredictable – but only in preflap contexts. Furthermore, it turns out to be a complete accident that the dialect has a rule neutralizing the [t]-[d] contrast in this very context.

In the face of this argument, one might try to salvage the single-level model as follows. We grant that the vowel length contrast in *writer-rider* depends on a prior voicing distinction in the following consonant and thus that the grammar must "compute" representations with a [t] for *writer* and a [d] for *rider*. But we observe that the latter exist in the related words *write* and *ride*. Furthermore, *writer* and *rider* are obviously constructed morphologically from the bases *write* and *ride*. We might thus revamp the single-level model to allow phonological rules to state generalizations over the phonological representations that appear at each layer in the morphological construction of a word. The vowel length rule states a generalization that holds over the information contained in the base forms *write* [răi̯t] and *ride* [ra:i̯d]. The morphology then suffixes the agentive *-ər*, setting up the context for the sonorization rule predicting a flap. The flap rule likewise states a generalization that holds of the surface forms [răi̯Dər] and [ra:i̯Dər]. Thus, while a complex word may have several representations, they are products of the morphology – not the phonology. Only a single phonological representation exists at any given point in the construction of the paradigm. Phonological rules state systematic relations holding within the information structures contained in these representations.

While there may be some merit in the view that vowel length is assigned to the stems in (20) before the suffixes are added, there are many other phonological alternations that cannot be explained in this way. For example, the words *tent* and *tend* contrast in the voicing of their final consonants; this contrast is responsible for the concomitant vowel length difference *t*[ɛ:]*nd* vs. *t*[ɛ̆]*nt*. Consider now the effect of suffixing plural and 3sg. affixes. For many speakers, the dental stops terminating the stem are deleted: they say [t^{h}ɛ̆ns] for *tents* and [t^{h}ɛ:nz] for *tends*. We may account for this alternation by the rule shown in (22).

(22) $$\begin{bmatrix} +\text{coron} \\ -\text{contin} \end{bmatrix} \rightarrow \emptyset \; / \; [+\text{nasal}] \;\_\_\_\; \begin{bmatrix} +\text{coron} \\ +\text{contin} \end{bmatrix}$$

Let us grant that this rule defines a generalization over the surface phonetic representations. We may also assume that, as in the case of *writer* vs. *rider*, the vowel length contrast is defined in [t^{h}ɛ̆ns] vs. [t^{h}ɛ:nz] before the suffix is added. The problem is to explain the voicing difference in the inflectional suffix itself. In the dual-level model, this difference is derived by the general rule that assimilates the voicing of a suffixal obstruent to the preceding segment – the rule that operates in *cup*[s], *hut*[s], *duck*[s] vs. *tub*[z], *bud*[z], *bug*[z]. We need merely stipulate that the voicing assimilation rule is defined on the underlying representation of these words – not on the surface form, where the opposition between [t] and [d] has been neutralized.

(23)	/tɛnt-z/	/tɛnd-z/	UR
	tɛnt-s	———	voicing assimilation
	tɛ̆nt-s	tɛ:nd-z	length
	tɛ̆n-s	tɛ:n-z	deletion (22)

The crucial point is that, unlike in the case of the vowel length, we cannot assign the voicing contrast in the suffix on an earlier morphological stratum. The suffix only appears on the second stratum. It is only by the morphological act of combining stem and suffix that we bring the affixal consonant into phonological proximity with the preceding consonant. But it is precisely at this stratum that the determinant voicing difference in the dental stop is missing from the surface representations: [tʰẽn-s] vs. [tʰɛ:n-z]. It thus looks as if the single-level model must abandon all hope of stating the obviously systematic correlation that exists between the voicing of the suffixal consonant and the voicing of the stem-final consonant.

Another example making the same point is to be found in Mohanan's (1992b) discussion of some rules in the Singapore dialect of English. Like many other varieties of English, the Singapore dialect simplifies consonant clusters composed of a fricative plus stop at the end of the word through deletion of the stop (perhaps reflecting more rigid constraints on syllabification). As a result, words such as *lift, ask,* and *list* are pronounced [lɪf], [ɑ:s], [lɪs]. For many speakers, the underlying stop appears before suffixes beginning with a vowel: *lif*[t]*-ing, as*[k]*-ing, lis*[t]*-ing,* and so on. As in other dialects, the plural suffix *-es* in Singapore takes a reduced schwa-like vowel when the stem ends in a sibilant: *raise-*[ə]*s, hiss-*[ə]*s, watch-*[ə]*s.* After nonsibilants the suffix consists of just the consonant, appearing as [-s] or [-z] as a function of the voicing of the final segment of the stem: *set-s, laugh-s* with [-s] and *bag-s, see-s* with [-z]. The relevant fact here is that the plurals of *list* [lɪs] and *task* [tɑ:s] are not *liss-es* and *tass-es* with a schwa parallel to *hiss-es* but rather simply *list-s* [lɪs] and *task-s* [tɑ:s]. This contrast follows straightforwardly if the rule determining whether or not the plural suffix takes a schwa is controlled by the underlying phonological representation of the stem: /hɪs/ vs. /lɪst/. The former forces the schwa-initial variant of the plural while the latter with its final [t] chooses [-s]. The resultant [lɪst + s] then simplifies to [lɪs] by the cluster reduction process mentioned above. Once again we cannot escape the force of this example by appeal to an earlier morphological cycle. The presence or absence of the schwa is a function of adding the plural suffix. But the proper choice crucially depends on a sound that is not pronounced in the plural.

To summarize the results of this section: We have seen how a variety of different English sounds ([t], [tʰ], etc.) manifest the same phonological category: the phoneme [t]. The phonological rules that realize the [t] operate in a very general and largely unconscious manner. It takes phonetic training for English speakers to realize that the [t]'s of *take* and *stake* are different sounds. They are typically unaware of the flap in *writer,* and no orthographic reformer would propose introducing a special letter to spell this consonant. Such a policy would be impracticable – precisely because most speakers are not aware of this sound. The phonological rules define a given language's sound pattern. In general, speakers of a different language (i.e., one with a different set of rules) will categorize the same sounds differently. For example, while English speakers interpret the intervocalic flap as a dental stop [t] or [d], Spanish speakers assign their flap to a rhotic phoneme. Given these differences among languages, the important question is, How are these assignments made?

When this question was first posed in the early part of the 20th century (in structural linguistics), it was suggested that each phoneme was defined by an invariant set of phonetic features common to all its allophones. For example, the [pʰ] of *pin* and the [p] of *spin* share the features [+labial, −voiced, −continuant]. It was this common core of distinctive features that was thought to define the phoneme [p] for English. But it soon became clear that this answer runs into serious problems. We have seen that the English [t] phoneme has the allophones [t], [tʰ], [ɾ] (flap), [tˀ], [ʔ], and [ ]. Clearly, there is no invariant core here. Furthermore, this set of realizations overlaps the [d] phoneme, which also appears as [ɾ] and [ ]. A procedure that groups allophones on the basis of an invariant core will never arrive at the psychologically correct categorizations. Various solutions to these problems were proposed in the American structuralism of the 1930s and 1940s, but each was found to be seriously defective. In the 1950s Noam Chomsky and Morris Halle concluded that these defects arose from the basic methodological premise of American structuralism that the phonemic structure could only be induced from information present in the (overt) speech act – a strong form of empiricism known as *behaviorism*. Chomsky and Halle rejected this point of view and advocated a mentalist approach to phonology similar to that envisioned by Edward Sapir, for whom a language's phonemic inventory is a conceptual system of "ideal sounds" in terms of which the phonetic segments of speech are perceived and articulated. The phonological representations and rules are thus hypotheses about the information structures of the individual speaker – what he or she has stored in the mind/brain after learning to speak a given language.

If we accept this general point of view, then a number of questions arise. These questions constitute the research program of generative phonology. Descriptively, we ask, For any given language X, what features has it selected from the UG phonetic alphabet to construct its lexicon? What are the phonological representations of language X and its corresponding rules? How do the rules apply to derive (compute) the phonetic representations? At the theoretical level, we ask, What is a possible representation and rule? Since the language faculty has developed for millennia, the number and variety of phonological systems is large. But many aspects seem to be variations on a limited number of themes. The range of possible rules and representations seems to be sharply constrained. If these constraints are wired into the system (if the child "knows" them at birth as part of genetic endowment), they may circumscribe and direct the learning process and make more plausible how such an intricate system as language is acquired in the first place.

2.7 Dual-Level Model: Additional Evidence

2.7.1 *Polish*

In this section we examine two additional cases in which the grammar must postulate an abstract phonological representation in order to state the proper generalizations. The first example comes from the West Slavic language Polish. Inspection of the singular and plural forms of the masculine nouns in (24) reveals two note-worthy alternations.

(24)

sg.	pl.	
klup	klubi	'club'
trup	trupi	'corpse'
dom	domi	'house'
šum	šumi	'noise'
snop	snopi	'sheaf'
žwup	žwobi	'crib'
trut	trudi	'labor'
dzvon	dzvoni	'bell'
kot	koti	'cat'
lut	lodi	'ice'
grus	gruzi	'rubble'
nos	nosi	'nose'
vus	vozi	'cart'
koš	koše	'basket'
nuš	nože	'knife'
wuk	wugi	'lye'
wuk	wuki	'bow'
sok	soki	'juice'
ruk	rogi	'horn'
bur	bori	'forest'
žur	žuri	'soup'
vuw	vowi	'ox'
ul	ule	'beehive'
sul	sole	'salt'
buy	boye	'fight'

In some stems the final obstruent consonant alternates in voicing, appearing as voiced before the plural suffix and as voiceless in word-final position. For example, there is a [p]≈[b] alternation in 'club', a [t]≈[d] alternation in 'labor', an [s]≈[z] alternation in 'rubble', a [š]≈[ž] alternation in 'knife', and a [k]≈[g] alternation in 'lye'. There are two possible descriptions of this alternation, depending on whether the voiced or voiceless obstruent is posited as underlying. If voiced obstruents are underlying, a rule is needed to devoice them at the end of the word (25a). This rule is found in many languages (e.g., German, Russian, Catalan). However, perhaps equally as widespread is the process that voices obstruents intervocalically (25b). (As a notational convenience, we assume that the word boundary symbol # appears at the beginning and end of each phonological word. We can then refer to the final phoneme of the word as the one that stands before #; the initial phoneme immediately follows #.)

(25) a. $[-\text{sonor}] \rightarrow [-\text{voiced}] \;/\; \underline{\quad\quad} \;\#$

b. $[-\text{sonor}] \rightarrow [+\text{voiced}] \;/\; \text{V} \; \underline{\quad\quad} \; \text{V}$

Thus, on grounds of phonological plausibility, either (25a) or (25b) is a reasonable first guess concerning the rule that underlies the voicing alternation. However, two additional facts motivate the devoicing analysis (25a). The first is that Polish has an equally large number of stems with a final voiceless consonant that never changes to voiced before the plural suffix. The morphemes meaning 'corpse', 'cat', 'nose', and so on, belong to this class. In a grammar with rule (25b), the underlying representations of such lexical items could not be simply [trup], [kot], [nos]. Some additional information indicating that they exceptionally fail to undergo the voicing rule would have to be added. But if (25a) is selected, then the underlying representations do not require this additional information. A learning procedure striving for maximally simple grammars will thus choose (25a) over (25b).

Another reason for selecting this analysis is that Polish has no stems with a final voiced obstruent that remains voiced at the end of the word. Although there are stems such as [trup] with a constant voiceless final consonant, and there are stems such as [klub] where the final obstruent alternates, there are no stems in which the final obstruent is always voiced. The final devoicing rule of (25a) states a generalization of Polish phonetics that is without exception and thus may be said to explain this gap in the distribution of voiced obstruents. If (25b) were the rule, then it would turn out to be a complete accident that Polish has no word-final voiced obstruents. In other words, the grammar incorporating (25b) would fail to relate the distributional gap (no final voiced obstruents) with the alternation exhibited by 'club', 'crib', 'labor', and so on. A separate rule would be required to state the distributional gap. In the alternative analysis, one rule (25a) accounts for the alternation and simultaneously explains the distributional gap. Once again, it is a simpler grammar.

The final devoicing rule (25a) is a *neutralization* rule. Polish utilizes [±voiced] to encode its lexicon. The voicing contrast surfaces before vocalic suffixes such as the plural. But when no suffix follows, the final devoicing rule neutralizes the underlying contrast by changing the specification of an obstruent from [+voiced] to [−voiced]. By virtue of this rule, underlying [b,d,g,z,ž] become identical with underlying [p,t,k,s,š]. Where the [±voiced] specification in the final consonant is the only difference between a pair of morphemes, the two pronunciations merge (e.g., in 'lye' and 'bow'). The following is a good descriptive heuristic. If a morpheme exhibits an alternation between [x] and [y] and one of the two terms of the alternation (say, [x]) is also found in a class of morphemes that fail to alternate in the same way, then try positing the other term of the alternation [y] as underlying.

Let us apply this heuristic to the analysis of the vocalic alternation in (24) between [o] and [u] found in such stems as 'crib', 'ice', 'cart', 'horn'. Since [o] and [u] differ just by the feature [high], the structural change of the rule will be easy to state. However, whether it is [o] or [u] that underlies the alternation is rather more difficult to determine. We may reason as follows. If [u] were underlying, it is unclear why the vowel lowers in *bur, bori* 'forest' and *sul, sole* 'salt' but fails to do so in *žur, žuri* 'soup' and *ul, ule* 'beehive', where we find a stable, nonalternating vowel. [b] and [s] together, excluding [ž], do not form a natural class. If phonological rules are defined over natural classes, we must seek another analysis for the alternation. Our descriptive heuristic suggests that [o] is basic to the alternation and thus that the underlying representations for the alternating morphemes are [žwob], [lod], [voz], [noz], [rog] – essentially, the stem shapes that appear in the plural. To maintain this analysis, a rule is needed to raise [o] to [u] in the singular. But we must formulate the rule so that it does not change the nonalternating [o] in such stems as [snop], [kot], [nos], [koš], and [sok]. The generalization is now obvious. [o] raises before a voiced consonant – but only when that consonant is word-final, since no change occurs in the plural. Furthermore, nasal consonants systematically fail to condition the rule. We may incorporate these observations into the rule stated in (26), which correctly delimits the contexts in which [o] changes to [u] in Polish.

(26) $$\begin{bmatrix} -\text{cons} \\ +\text{back} \\ -\text{low} \end{bmatrix} \rightarrow [+\text{high}] \,/\, \underline{\quad} \begin{bmatrix} +\text{cons} \\ +\text{voiced} \\ -\text{nasal} \end{bmatrix} \#$$

Clearly this rule must apply before devoicing since the [o] → [u] change depends on the voicing contrast that the devoicing rule neutralizes. An ordering restriction "raising precedes final devoicing" guarantees this result. A word such as *vus* 'cart' thus receives the derivation in (27a). If final devoicing had applied first, /voz/ would appear as [vos]. The raising rule would then be unable to apply and we would derive the incorrect *[vos] (27b).

(27) a.	/#voz#/	UR	b.	/#voz#/	UR
	vuz	raising		vos	devoicing
	vus	devoicing		inappl.	raising

The Polish raising rule has a status similar to that of the English vowel-lengthening rule operative in the *writer-rider* contrast. Both rules crucially refer to a voicing distinction in the following consonant that is neutralized by a subsequent rule. But in the case of Polish *vus*, the triggering voiced consonant does not appear on a earlier stratum in the morphological structure of the word. Unlike English *rid-er*, Polish *vus* is a root. It is not derived from any simpler form. Furthermore, unlike in the case of English *tents* (*tĕn*[s]) vs. *tends* (*tɛ:n*[z]), where the vowel length might be considered an enhancement of the obstruent voicing, the [o]≈[u] alternation cannot be construed as incorporating a distinctive feature from the following consonant. The Polish data thus furnish additional evidence for phonological versus phonetic representations. To account for the systematic raising of [o] to [u] in nom. sg. forms of 'cart', 'knife', 'horn', and so on, the grammar must contain representations with a final voiced obstruent ([voz], [nož], [rog]) even though that voiced obstruent is uniformly pronounced voiceless. Such "abstract" representations will be formed by a grammar that seeks to maximize the simplicity of its rules and representations.

Two additional observations are in order. First, we defined the context for both Polish rules in essentially phonological terms instead of referring to the grammatical context in which the stem finds itself (e.g., devoice the final consonant of the stem in the nominative singular). For final devoicing, this decision is motivated by the fact that no word in the language ends in a voiced obstruent, regardless of grammatical category. In addition, there are several other morphological contexts that have no overt suffix and thus place the stem in word-final position. In each case the final devoicing and raising rules apply – as predicted by the formulations in (25a) and (26). One of these is the genitive plural of feminine and neuter nouns; the other is the imperative singular of the verb.

(28)	nom.sg.	gen.pl.		1sg.	imper.	
	swov-o	swuf	'word'	rob'e	rup	'do'
	brod-a	brut	'beard'	vodze	vuć	'lead'
	proz'b-a	prus'p	'request'	odvože	odvuš	'open'
	drog-a	druk	'road'	zwov'e	zwuf	'catch'
	bžoz-a	bžus	'birch'	stoye	stuy	'stand'
	komor-a	komur	'closet'	ogole	ogul	'shave'
	pol-e	pul	'field'			

In general, when rules are formulated in terms of phonological contexts, they make predictions that go beyond the data at hand. In our descriptive work, we may discover additional data that require modification of our initial formulations. Experienced phonologists have strong intuitions about what rules are likely to underlie a given alternation and readily use this "expert system" to develop hypotheses about the rules and representations that constitute the grammar. Our descriptive work constantly seeks to test these hunches and extend the empirical coverage of the proposed rules. Needless to say, one often guesses wrong and the system turns out to be quite different from what was initially expected. Sometimes this is because a crucial piece of information is missing and, once discovered, radically alters the interpretation of the data. At other times the language may be truly odd and not behave the way the theory predicts. This should not be too surprising because any phonological system is also the product of historical evolution whose preceding sound changes may leave arbitrary gaps and create missing links. Finally, descriptive failure may indicate that the underlying theory is simply incorrect and must be revised.

A second point worthy of note is that unlike most of our previous rules, the Polish raising rule cannot readily be assigned to a natural phonological category. It is not a rule of assimilation, weakening, strengthening, enhancement, and so on. Polish raising typifies what Jan Baudouin de Courtenay (1845–1929) called a *sinnlos* ("irrational") alternation. It is unlikely that any other language has a rule of exactly this form. While the environment for the rule forms a natural class and the structural change is very simple, there is no good reason why this particular vowel change ([o] → [u]) should take place in this particular context (before voiced oral consonants). (The answer to this puzzle lies in the history of the language. See Kenstowicz and Kisseberth 1977:64 for discussion.) But even though the [o]≈[u] alternation is phonologically arbitrary, it is clear that it is still a genuine rule of Polish phonology. For example, Bethin (1978) notes that it regularly extends to loanwords in the genitive plural (e.g., *doz-a, dus* 'dose'; *pagod-a, pagut* 'pagoda'), but interestingly not in the nominative singular, where many exceptions have developed (*snop, snob-a* 'snob'; *epizot, epizod-a* 'episode'; *gol* 'goal').

2.7.2 Icelandic

In this section we will examine one more example to motivate the dual-level model of phonological representation. Our discussion focuses on the *u*-epenthesis rule of Modern Icelandic. Orešnik (1972) discusses several independent lines of evidence that converge on the conclusion that the [u] in *dag-ur* and *hest-ur* (29a) is inserted by the rule shown in (29b) and thus that the underlying representation of the nom.sg. suffix in this declension class is [-r].

(29)	a.	nom.sg.	dag-ur	hest-ur	bæ-r
		acc.sg.	dag	hest	bæ
			'day'	'horse'	'farmhouse'

b. $\emptyset \rightarrow u$ / C ___ r #

A rule breaking up final C*r* clusters is a natural process termed *epenthesis* in traditional grammar. But deletion of one vowel after another is also very common. Thus, considerations of rule naturalness do not resolve the analysis of the [-r]≈[-ur] alternation. We must consider the effects the insertion analysis has on other rules of the grammar. Orešnik's first argument runs as follows. If [u] underlies the [u]≈Ø alternation, a rule is needed to delete it after vowel-final stems such as [bæ]. But Icelandic already has a rule of truncation eliminating a V-V cluster for data not considered here. This rule deletes the first vowel in the cluster, however, not the second. Orešnik concludes that epenthesis underlies the alternation since, if the language already has a means to eliminate vowel clusters, why is it not utilized here? This argument is based on the premise that underlying V-V sequences should be eliminated in the same way, regardless of the particular vowels filling the first and second positions. While true in general, this is not very strong justification for an underlying [-r] rather than [-ur] because an additional rule is required in any case. The analysis of the [u]≈Ø alternation as epenthesis receives much stronger support from its interaction with two other phonological alternations.

The first of these is evident in the paradigms of (30a), where stems such as [lyfj] lose their final [j] (a front glide) before a consonant or word boundary. [v] (underlying [w]) has a parallel distribution. As we will see when we develop the theory of syllabification, this is a common alternation. A consonant-glide sequence is difficult to incorporate into the syllable coda. Many languages follow Icelandic in simplifying such clusters through deletion of the second member. This rule can be stated informally as in (30b).

(30)	a.	nom.sg.	lyf-ur	byl-ur	beð-ur	söng-ur
		acc.sg.	lyf	byl	beð	söng
		gen.sg.	lyf-s	byl-s	beð-s	söng-s
		dat.pl.	lyfj-um	bylj-um	beðj-um	söngv-um
		gen.pl.	lyfj-a	bylj-a	beðj-a	söngv-a
			'medicine'	'storm'	'bed'	'song'

b. $[j,v] \rightarrow \emptyset$ / C ___ $\begin{Bmatrix} \# \\ C \end{Bmatrix}$

Returning to the [u]≈Ø alternation, if the underlying representation of the nom.sg. ending is [-ur], then we expect the glides to surface before the vowel – as they do in the dative and genitive plural: *lyfj-um, lyfj-a*. But this is not what we find. Instead, the glide deletion rule uniformly applies: the nominative singular is *lyf-ur*, not *lyfj-ur*. As Orešnik observes, the loss of the glide is explained if zero underlies the [u]≈Ø alternation. Then the nom.sg. suffix will have the same mono-consonantal shape as the gen.sg. [-s] and thus will activate rule (29b). Of course, the glide deletion rule must be ordered before epenthesis so that derivations such as those in (31) obtain.

(31)	/#bylj-um#/	/#bylj-s#/	/#bylj#/	/#bylj-r#/	UR
	——	byl-s	byl	byl-r	glide deletion (30b)
	——	——	——	byl-ur	epenthesis (29b)

If the rules had applied in the opposite order, then [bylj-r] would become [bylj-ur] by epenthesis and the stem-final glide could not delete by (30b).

The general point to emerge from this discussion is that application of the glide deletion rule in the derivation of *byl-ur* crucially depends on positing that the underlying representation of the nom.sg. suffix begins with a consonant [-r]. Yet in every pronunciation of this word (and all stems ending in C-glide), the suffix has the shape [-ur].

The *u*-umlaut rule in Icelandic independently supports this analysis. As illustrated in (32a), many suffixes containing a [u] such as the dative plural, the diminutive, and the first person plural of the verbal inflection mutate the [a] of a preceding syllable to [ö]. This rule is stated informally in (32b).

(32) a.	barn	nom.sg.	baggi		kalla	1sg.
	börn-um	dat.pl.	bögg-ull		köll-um	1pl.
	'child'		'bundle', 'parcel'		'call'	

b. $[a] \rightarrow [ö]$ / ___ C_0 [u]

There is one systematic exception to umlaut: the nom.sg. [-ur]. As the data in (33) demonstrate, [-ur] does not umlaut the preceding stem.

(33)	nom.sg.	hatt-ur	dal-ur	stað-ur
	acc.sg.	hatt	dal	stað
	dat.pl.	hött-um	döl-um	stöð-um
		'hat'	'valley'	'place'

The lack of umlaut has nothing to do with the roots, since they regularly submit to the rule in the dative plural. The assumption that the underlying representation of the nom.sg. suffix is [-r] helps us to understand why this suffix fails to umlaut a preceding root. If the umlaut rule is defined to operate at a point that precedes the epenthesis rule, the suffix will lack a vowel to trigger the umlaut. Under this analysis, *hatt-ur* and *hött-um* receive the following derivations.

(34)

/#hatt + r#/	/#hatt + um#/	UR
———	hött-um	*u*-umlaut (32b)
hatt-ur	———	epenthesis (29b)

Thus, two separate aspects of the *-ur* suffix's behavior are explained if its underlying representation lacks a vowel. If the theory of grammar posits just a single level of phonological representation, we would be unable to explain the systematic behavior of the suffix – and the hundreds of words in which it is attached – in this way. To ensure that the umlaut and glide deletion rules operate on this more abstract representation, we stipulate that each precedes application of epenthesis.

To summarize the results of this section, we have examined several cases indicating that the phonological structure of a given lexical item may require the application of rules that crucially refer to information that is not present in the item's phonetic representation. But in each case the missing information is present in the underlying phonological representation – that is, in the representation that arises from consistent elimination of predictable information and assignment of this information by general phonological rules.

Suggested Readings

Anderson, Stephen. 1985. Phonology in the twentieth century. Chicago: University of Chicago Press. Chapters 11 and 12.

Bromberger, Sylvain, and Morris Halle. 1989. Why phonology is different. Linguistic Inquiry 20.51–70.

Halle, Morris. 1962. Phonology in generative grammar. Word 18.54–72. Reprinted in Phonological theory: Evolution and current practice, ed. by V. Makkai, 393–400. New York: Holt, Rinehart and Winston, 1972.

Kenstowicz, Michael. 1992. American structuralist phonology. International encyclopedia of linguistics, vol. 3, ed. by W. Bright, 215–17. Oxford: Oxford University Press.

Sapir, Edward. 1925. Sound patterns in language. Language 1.37–51. Reprinted in Selected writings of Edward Sapir in language, culture, and personality, ed. by D. Mandelbaum, 33–45. Berkeley: University of California Press, 1949; in Readings in linguistics I, ed. by M. Joos, 19–25. Chicago: University of Chicago Press, 1957; in Phonological theory: Evolution and current practice, ed. by V. Makkai, 13–21. New York: Holt, Rinehart and Winston, 1972.

Exercises

HW

2.1 ***English Allophones***

Review the analysis of (American) English developed in the text. Tell what allophone it assigns to the boldfaced segments in the following words. If the result is different from your dialect, tell what the difference is.

(1)

pin	hug	laugh	rat
stupid	true	Santa Cruz	atlas
martyr	mástodòn	actress	chowder
antler			

2.2 *English Minimal Pairs*

The system of stops and nasals to be found in English is tabulated in (1). These segments are built from combining [±voiced] and [+nasal] with the three articulators [labial], [coronal], and [dorsal].

(1)

	labial	coronal	dorsal
voiceless stops	p	t	k
voiced stops	b	d	g
nasals	m	n	ŋ

Listed in (2) are minimal triples constructed over the three classes of consonants in (1). In these examples the consonants appear in both initial and final position. (2a,b) indicate that the articulator features [labial], [coronal], and [dorsal] are contrastive. (2c,d) suggest that the features [voiced] and [nasal] are also utilized to encode the lexicon. There is of course one major distributional gap: the velar nasal does not appear word-initially (see the discussion in section 3.4).

(2) a. pin, tin, kin; bun, dun, gun; mutt, nut
b. lip, lit, lick; robe, road, rogue; dim, din, ding
c. pat, bat, mat; tip, dip, nip; cut, gut
d. rope, robe, roam; mate, made, main; pick, pig, ping

Try to construct minimal *n*-tuples (pairs, triples, quadruples) over the set of English fricatives listed in (3) for both initial and final position. Which of these segments appear to have restricted distributions?

(3) f θ s š
v ð z ž

2.3 *English s*C Clusters

Although voiced and voiceless stops freely contrast initially, medially, and finally in English, there is one position where the opposition is suspended: after morpheme-initial [s] (e.g., *s*[p]*in*, *s*[t]*em*, *s*[k]*in*). What significance is to be attached to the fact that neither voiced nor aspirated stops are found here instead?

2.4 *Korean*

The liquids [l] and [r] are in complementary distribution in Korean (data from Demers and Farmer 1991). State the context where each is found. What difficulty is a name such as *Lori Roland* likely to present to the Korean learner of English?

(1) mul	'water'	mal	'horse'
mulkama	'place for water'	malkama	'place for horse'
mure	'at the water'	mare	'at the horse'

pal	'foot'	səul	'Seoul'	rupi	'ruby'
pari	'of the foot'	ilkop	'seven'	ratio	'radio'

2.5 *Singapore English*

In his discussion of Singapore English, Mohanan (1992b) notes a process whereby [sp] in the coda of the syllable metathesizes to [ps]; [st] and [sk] clusters reduce by the process mentioned in section 2.6.

(1) lisp	[lips]	list	[lis]	risk	[ris]
crisp	[krips]	mist	[mis]	whisk	[wis]
grasp	[gra:ps]	past	[pa:s]	mask	[ma:s]

Before vocalic suffixes speakers break into the two groups illustrated in (2).

(2)	group 1	group 2
lisping	[lisp-iŋ]	[lips-iŋ]
crispy	[krisp-i]	[krips-i]
grasping	[gra:sp-iŋ]	[gra:ps-iŋ]

Mohanan reports that both groups of speakers treat plurals and 3sg. present verb forms the same, as indicated in (3).

(3) sg.		pl.	
hiss	[his]	hisses	[hisəs]
eclipse	[eklips]	eclipses	[eklipsəs]
lapse	[læps]	lapses	[læpsəs]
lisp	[lips]	lisps	[lips]
grasp	[gra:ps]	grasps	[gra:ps]

Discuss the implications of these data for the issue of single- vs. dual-level models of phonological representation.

2.6 *Friulian*

A. In the Friulian dialect of Italian (Vanelli 1979, 1986), there is an alternation between voiced and voiceless obstruents. Suggest a rule to account for the following voicing alternations.

(1) wárp	'blind'	kwárp	'body'
warb-ít	'sty'	kwarp-út	dimin.
piérd-i	'to lose'	dínt	'tooth'
piért	3sg.	dint-isín	dimin.

B. Accented vowels seem to contrast in length: for example, *lá:t* 'went' vs. *lát* 'milk', *pá:s* 'peace' vs. *pás* 'step'. But in many cases the length is predictable. Suggest a rule to account for the length in the following data. How must your rules be ordered? How does your analysis bear on the issue of one level of representation versus two?

(2)

lá:t	'went'	brút	'ugly'
lád-e	fem.	brút-e	fem.
nervó:s	'nervous'	rós	'red'
nervóz-e	fem.	rós-e	fem.
tróp	'flock'	ló:f	'wolf'
trop-út	dimin.	lov-út	dimin.
sék	'dry'	fí:k	'fig'
séc-e	fem.	fig-ón	'big fig'
vjód-i	'to see'		
vjó:t	3sg.		

C. Friulian has recently borrowed many words from Standard Italian. What bearing do these loanwords have on the analysis?

(3)

Friulian	Italian	
impjegá:t	impiegato	'clerk'
impjegád-e	impiegata	fem.
istitú:t	istituto	'institute'
istitud-út		dimin.
steká:t	steccato	'fenced'
stekad-út		dimin.
afít	affitto	'rent'
afit-út		dimin.

2.7 *Northern Salentino*

As in Standard Italian, the Northern Salentino dialect distinguishes seven vowels in stressed position (Calabrese 1984).

(1)

	i	e	ɛ	a	ɔ	o	u
high	+	−	−	−	−	−	+
low	−	−	−	+	−	−	−
back	−	−	−	+	+	+	+
ATR	+	+	−	−	−	+	+

A. In the following paradigms, the root vowels change in response to the shift of accent – generally on the penultimate syllable. Postulate a rule to account for the alternations in vowel quality that result from the shift of stress. What is the underlying representation? Justify your choice. State the rule, utilizing distinctive features. Show how the analysis works by giving the derivations for ***kanósku*** and ***kanuššímu***.

(2)					
	1sg.	kréu	sέntu	kanósku	trɔ́u
	3sg.	kréti	sέnti	kanóšši	trɔ́a
	1pl.	kritiámu	sintímu	kanuššímu	truámu
		'believe'	'feel'	'recognize'	'find'

B. Some suffixes cause a mutation of the root vowel (known as *metaphony* in Romance linguistics). Examine the following paradigms and suggest a rule to account for the metaphony.

(3)								
	masc.sg.	karúsu	rússu	frísku	fríddu	krútu	vívu	sánu
	fem.sg.	karósa	róssa	fréska	frédda	krúta	víva	sána
	masc.pl.	karúsi	rússi	fríski	fríddi	krúti	vívi	sáni
		'young'	'red'	'cool'	'cold'	'raw'	'alive'	'healthy'

C. In contrast to the suffixes in (3), other suffixes do not cause metaphony. While they may be treated simply as exceptions, a phonological explanation is also possible. Suggest an analysis along the latter lines. Must the rules be ordered? Show how the analysis works by deriving the words *karósi* and *karúsi*.

(4)						
	a.	fem.pl.	karósi	róssi	fréski	fréddi
		masc.pl.	karúsi	rússi	fríski	fríddi
			'young'	'red'	'slow'	'cold'
	b.	sg.	mési	péšši	nóči	króči
		pl.	mísi	píšši	núči	krúči
			'month'	'fish'	'nut'	'cross'
	c.	3sg.	véti	kréti	kanóšši	
		2sg.	víti	kríti	kanúšši	
		1sg.	vétu	kréu	kanósku	
			'see'	'believe'	'recognize'	

2.8 Yakut

Yakut is an Altaic language spoken in Siberia (Kruger 1962). The data for this exercise, shown in (1), consist of nouns in various case forms. Asterisked items have been constructed from attested models to fill out the paradigms.

(1)	gloss	absolute	plural	dative	accusative	partitive	our N	your N
	'father'	aɣa	aɣalar	aɣaɣa	aɣanɯ	aɣata	aɣabɯt	aɣaɣɯt
	'child'	oɣo	oɣolor	oɣoɣo	oɣonu	oɣoto	oɣobut	oɣoɣut
	'lake'	küöl	küöller	küölge	küölü	küölle	küölbüt	küölgüt*
	'horse'	at	attar	akka	atɯ*	atta*	appɯt	akkɯt
	'duck'	kus	kustar	kuska*	kuhu*	kusta*	kusput	kuskut
	'bull'	oɣus	oɣustar	oɣuska*	oɣuhu*	oɣusta	oɣusput	oɣuskut

gloss	absolute	plural	dative	accusative	partitive	our N	your N
'tool'	sep	septer	sepke	sebi	septe	seppit	sepkit
'meat'	et	etter	ekke	eti	ette	eppit	ekkit
'arrow'	ox	oxtor	oxxo	oyu	oxto	oxput	oxxut
'knee'	tobuk	tobuktar	tobukka	tobugu	tobukta	tobukput	tobukkut
'elder brother'	ubay	ubaydar	ubayga*	ubayɯ*	ubayda	ubaybɯt	ubaygɯt
'stallion'	atɯɯr	atɯɯrdar	atɯɯrga*	atɯɯrɯ*	atɯɯrda	atɯɯrbɯt	atɯɯrgɯt
'squirrel'	tiiŋ	tiiŋner	tiiŋŋe*	tiiŋi	tiiŋne	tiiŋmit	tiiŋŋit
'door'	aan	aannar	aaŋŋa	aanɯ	aanna	aammɯt	aaŋŋɯt
'ford'	olom	olomnor	olomŋo	olomu	olomno	olommut	olomŋut

Segment the words into root plus suffix. Note the many alternations. Tabulate the sounds that comprise the phonetic inventory. Only a subset of these appear in underlying representations. Following Kruger, you may postulate the following phonemic inventory.

(2)						
high vowels	i	ü		ɯ	u	
nonhigh vowels	e			a	o	
voiceless stops		p	t	k		
voiceless fricatives			s	x		
nasals		m	n	ŋ		
approximants			l r			
			y			

The point of this exercise is to discover the rules realizing this system of phonemes in the various nominal paradigms. You may orient your analysis around the following questions.

A. Yakut contrasts seven root vowels, distinguished by appropriate choices among [±high], [±back], [±round]. Only one of these three features is contrastive in suffixes. Which one? How are the other features determined?

B. [±voiced] is not underlyingly contrastive. Assuming the appropriate default rules of section 2.4, account for the appearance of [±voiced] in the phonetic representations. You may assume a special rule that deletes the initial *-n* of the accusative suffix after a stem ending in a consonant.

C. Postulate additional rules to account for the remaining alternations. Must they apply in a particular order?

2.9 Kire Nasalization and Orthography

Pryor and Clifton (1987) explain how difficulties with the orthographic registration of vowel nasality in the Papua New Guinean language Kire encouraged a reinterpretation of the phonological structure of the language, which in turn led to a more efficient orthography. Primarily on the basis of such contrasts as *pi* 'to eat' vs. *pĩ:* 'breadfruit tree', Pryor's original analysis of Kire concluded that vowel nasality was phonemic in the language. Accordingly, he proposed the use of the dieresis for the native orthography. The above words were thus originally spelled as *pi* and *piï* (gemination used to indicate long vowels). The authors state (p. 34),

> Though this symbolisation was accepted by the Kire people, problems arose in deciding which vowels should be marked with the dieresis. Literate Kires began putting the dieresis on vowels that had previously been declared oral and omitting them on obviously nasal vowels following or preceding nasal consonants. In addition analysis of the morphophonemics of Kire indicated that at least some nasalisation was predictable.

The authors cite the data in (1a) to illustrate the point that vowel nasality is predictable from a tautosyllabic nasal consonant. This interpretation is also supported by the alternations in vowel nasality in (1b), which respond to changes in syllabification due to the addition of affixes.

(1) a.

vap	'fire'
p^{h}ɨkta	'shoulder'
bẽm	'tree species'
dõmdori	'roll over'
mõmũk	'tree species'
mõnũm	'fish species'
faramẽ	'thumb'

b.

sg.	dual	pl.	
gumã	gumãnĩ	gũmgi	'man'
kĩn	kɨnãnĩ	kɨnĩ	'banana'

Finally, Pryor and Clifton note that while Kire contrasts [m], [n], and [ŋ] in the onset of the syllable, [ŋ] is phonetically absent at the end of a syllable. They propose to analyze such words as those in (2a) with the underlying phonological representations in (2b) and posit the rule (2c) that deletes the velar nasal from the syllable coda.

(2)

a.	b.	
pĩ:	/pi:ŋ/	'breadfruit tree'
bĩ	/biŋ/	'bird species'
t^{h}ã:	/t^{h}a:ŋ/	'bamboo'

c. [ŋ] → ∅ at the end of the syllable

The authors state (p. 41),

> Long before the reanalysis outlined in [(2)] was completed, it had become clear that at least the majority of nasalised vowels were predictable. At the same time the dieresis was proving increasingly problematic. The major problem was the difficulty writers had in using it consistently. . . . Because of these difficulties, the possibility of eliminating the dieresis from the orthography was discussed with those Kires who had been trying to use the dieresis.

During these discussions they indicated their feelings that nasalised vowels were generally conditioned by neighboring nasal consonants, and not "basic" as the use of the dieresis would indicate. After explaining the possibility of using "silent" ŋ as a spelling device to signal nasalised vowels not flanked by a nasal consonant, they were enthusiastic about eliminating the dieresis.

Discuss the relevance of these comments from Pryor and Clifton to the issues of phonetic versus phonological accuracy in orthography. Why did the Kires have so much trouble with the dieresis? Why was the use of the ŋ more readily accepted?