

# CHAPTER 5

## Underlying representations

### PREVIEW

This chapter looks deeper into the nature of underlying forms by

- introducing contrast-neutralizing rules
- seeing how unpredictable information must be part of the underlying form
- learning what factors are most important in establishing an underlying representation
- understanding how underlying forms are different from actually pronounced words

#### KEY TERMS

*alternation*  
*neutralization*  
*predictability*  
*structure*  
*preservation*

A fundamental characteristic of the rules discussed up to this point is that they have described totally transparent allophonic processes, such as aspiration in English or vowel nasalization in Sundanese, where every appearance of a certain segment or class of segments can always be predicted by a rule, starting from some other segment. For such rules, the question of the exact underlying form of a word has not been so crucial, and in some cases a clear decision could not be made. We saw that in Sundanese every vowel becomes nasalized after a nasal sound, and every phonetic nasal vowel appears after a nasal. Nasality of vowels can always be predicted by a rule in this language: all nasal vowels appear in one predictable context, and all vowels are predictably nasal in that context. It was therefore not crucial to indicate whether a given vowel is underlyingly nasal or underlyingly oral. If you assume that vowels are underlyingly oral you can write a rule to derive all of the nasal vowels, and if you contrarily assume that vowels are all underlyingly nasal you could write a rule to derive all of the oral vowels. The choice of underlying sound may make a considerable difference in terms of simplicity and elegance of the solution, and this is an important consideration in evaluating a phonological analysis, but it is possible to come up with rules which will grind out the correct forms no matter what one assumes about underlying representations in these cases. This is not always the case.

### 5.1 The importance of correct underlying forms

**Neutralizing** rules, on the other hand, are ones where two or more underlyingly distinct segments have the same phonetic realization in some context because a rule changes one phoneme into another – thus the distinction of sounds is neutralized. This means that if

you look at a word in this neutralized context, you cannot tell what the underlying segment is. Such processes force you to pay close attention to maintaining appropriate distinctions in underlying forms.

Consider the following examples of nominative and genitive forms of nouns in Russian, focusing on the final consonant found in the nominative.

(1)	<i>Nominative singular</i>	<i>Genitive singular</i>	
	vagon	vagona	‘wagon’
	avtomobil'	avtomobil'a	‘car’
	vet'er	vet'era	‘evening’
	muʃ	muʒa	‘husband’
	karandaʃ	karandaʃa	‘pencil’
	glas	glaza	‘eye’
	golos	golosa	‘voice’
	ras	raza	‘time’
	les	lesa	‘forest’
	porok	poroga	‘threshold’
	vrank	vraga	‘enemy’
	urok	uroka	‘lesson’
	porok	poroka	‘vice’
	t'svet	t'sveta	‘color’
	prut	pruda	‘pond’
	soldat	soldata	‘soldier’
	zavot	zavoda	‘factory’
	xlep	xleba	‘bread’
	grip	griba	‘mushroom’
	trup	trupa	‘corpse’

To give an explanation for the phonological processes at work in these data, you must give a preliminary description of the morphology. While morphological analysis is not part of phonology per se, it is inescapable that a phonologist must do a morphological analysis of a language, to discover the underlying form.

In each of the examples above, the genitive form is nearly the same as the nominative, except that the genitive also has the vowel [a] which is the genitive singular suffix. We could assume as our initial hypothesis that the bare root of the noun is used to form the nominative case, and the combination of a root plus the suffix *-a* forms the genitive. Nothing more needs to be said about examples such as *vagon* ~ *vagona*, *avtomobil'* ~ *avtomobil'a*, or *vet'er* ~ *vet'era*, where, as it happens, the root ends with a sonorant consonant. The underlying forms of these noun stems are presumably /vagon/,

/avtomobilj/ and /vet'er/: no facts in the data suggest anything else. These underlying forms are thus identical to the nominative form. With the addition of the genitive suffix *-a* this will also give the correct form of the genitive.

There are stems where the part of the word corresponding to the root is the same in all forms of the word: *karandas* ~ *karandasa*, *golos* ~ *golosa*, *les* ~ *lesa*, *urok* ~ *uroka*, *porok* ~ *poroka*, *t<sup>č</sup>vet* ~ *t<sup>č</sup>veta*, *soldat* ~ *soldata* and *trup* ~ *trupa*. However, in some stems, there are differences in the final consonant of the root, depending on whether we are considering the nominative or the genitive. Thus, we find the differences *muf* ~ *muža*, ~ *glas* ~ *glaza*, *porok* ~ *poroga*, *vrek* ~ *vraga*, *prut* ~ *pruda*, and *xlep* ~ *xleba*. Such variation in the phonetic content of a morpheme (such as a root) is known as **alternation**. We can easily recognize the phonetic relation between the consonant found in the nominative and the consonant found in the genitive as involving voicing: the consonant found in the nominative is the voiceless counterpart of the consonant found in the genitive. Not all noun stems have such an alternation, as we can see by pairs such as *karandas* ~ *karandasa*, *les* ~ *lesa*, *urok* ~ *uroka*, *soldat* ~ *soldata* and *trup* ~ *trupa*. We have now identified a phonological problem to be solved: why does the final consonant of some stems alternate in voicing? And why do we find this alternation with some stems, but not others?

The next two steps in the analysis are intimately connected; we must devise a rule to explain the alternations in voicing, and we must set up appropriate underlying representations for these nouns. In order to determine the correct underlying forms, we will consider two competing hypotheses regarding the underlying form, and in comparing the predictions of those two hypotheses, we will see that one of those hypotheses is clearly wrong.

Suppose, first, that we decide that the form of the noun stem which we see in the nominative is also the underlying form. Such an assumption is reasonable (it is, also, not automatically correct), since the nominative is grammatically speaking a more “basic” form of a noun. In that case, we would assume the underlying stems /glas/ ‘eye,’ /golos/ ‘voice,’ /ras/ ‘time’ and /les/ ‘forest.’ The problem with this hypothesis is that we would have no way to explain the genitive forms *glaza*, *golosa*, *raza* and *lesa*: the combination of the assumed underlying roots plus the genitive suffix *-a* would give us \**glas-a*, *golosa*, \**ras-a* and *les-a*, so we would be right only about half the time. The important step here is that we test the hypothesis by combining the supposed root and the affix in a very literal-minded way, whereupon we discover that the predicted forms and the actual forms are different.

We could hypothesize that there is also a rule voicing consonants between vowels (a rule like one which we have previously seen in Kipsigis, chapter 3):

$$(2) \quad C \rightarrow [+voice] / V\_V$$

While applying this rule to the assumed underlying forms /glas-a/, /golos-a/, /ras-a/, and /les-a/ would give the correct forms *glaza* and *raza*, it would also give incorrect surface forms such as \**goloza* and \**leza*. Thus, not only is our first hypothesis about underlying forms wrong, it also cannot be fixed by positing a rule of consonant voicing.

You may be tempted to posit a rule that applies only in certain words, such as *eye*, *time* and so on, but not *voice*, *forest*, etc. This misconstrues the nature of phonological rules, which are general principles that apply to all words of a particular class – most generally, these classes are defined in terms of phonological properties, such as “obstruent,” “in word-final position.” Rules which are stated as “only applying in the following words” are almost always wrong.

The “nominative is underlying” hypothesis is fundamentally wrong: our failure to come up with an analysis is not because we cannot discern an obscure rule, but lies in the faulty assumption that we start with the nominative. That form has a consistent phonetic property, which is that any root-final obstruent (which is therefore word-final) is always voiceless, whereas in the genitive form there is no such consistency. If you look at the genitive column, the last consonant of the root portion of the word may be either voiced or voiceless.

We now consider a second hypothesis, where we set up underlying representations for roots which distinguish stems which have a final voiced obstruent in the genitive versus those with a final voiceless obstruent. We may instead assume the following underlying roots.

(3)	<i>Final voiced obstruent</i>	<i>Final voiceless obstruent</i>	
	/muʒ/	‘husband’	/karandaʃ/ ‘pencil’
	/glaz/	‘eye’	/golos/ ‘voice’
	/raz/	‘time’	/les/ ‘forest’
	/porog/	‘threshold’	/porok/ ‘vice’
	/vrag/	‘enemy’	/urok/ ‘lesson’
	/prud/	‘pond’	/tsvet/ ‘color’
	/zavod/	‘factory’	/soldat/ ‘soldier’
	/grib/	‘mushroom’	/trup/ ‘corpse’
	/xleb/	‘bread’	

Under this hypothesis, the genitive form can be derived easily. The genitive form is the stem hypothesized in (3) followed by the suffix *-a*. No rule is required to derive voiced versus voiceless consonants in the genitive. That issue has been resolved by our choice of underlying representations where some stems end in voiced consonants and others end in voiceless consonants. By our hypothesis, the nominative form is simply the underlying form of the noun stem, with no suffix.

However, a phonological rule must apply to the nominative form, in order to derive the correct phonetic output. We have noted that no word in Russian ends phonetically with a voiced obstruent. This regular fact allows us to posit the following rule, which devoices any word-final obstruent.

#### (4) *Final devoicing*

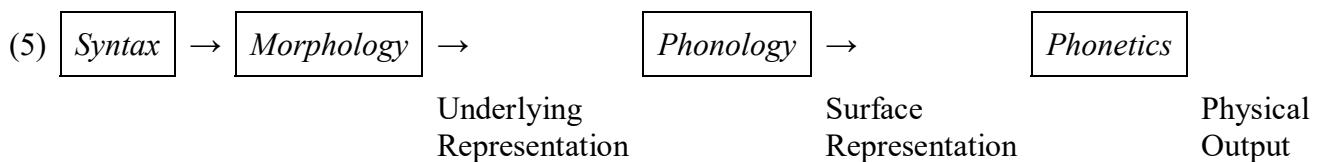
$$[-son] \rightarrow [-voice] / \_ \#$$

By this rule, an obstruent is devoiced at the end of the word. As this example has shown, an important first step in doing a phonological analysis for phenomena such as word-final devoicing in Russian is to establish the correct underlying representations, which encode unpredictable information.

Whether a consonant is voiced cannot be predicted in English ([dɛd] *dead*, [ted] *Ted*, [det] *debt*), and must be part of the underlying form. Similarly, in Russian since you cannot predict whether a given root ends in a voiced or a voiceless consonant in the genitive, that information must be part of the underlying form of the root. That is information about the root, which cannot always be determined by looking at the surface form of the word itself: it must be discovered by looking at the genitive form of the noun, where the distinction between voiced and voiceless final consonants is not eliminated.

## 5.2 Refining the concept of underlying form

It is important to understand what underlying forms are, and what they are not. The nature of underlying forms can be best appreciated in the context of the overall organization of a grammar, and how a given word is generated in a sentence. The structure of a grammar can be represented in terms of the standard block model.



This model implies that the output of one grammatical component forms the input to the next component, so the phonological component starts with whatever the morphological component gives it, and applies its own rules to give the surface representation (which are then subject to principles of physical interpretation within the phonetic component). The output of the morphological component, which is the input to the phonology, is by definition the underlying form, so we need to know a little bit about what the morphological component does, to understand what is presented to the phonology.

The function of the morphological component is to assemble words, in the sense of stating how roots and affixes combine to form a particular word. Thus the morphological component is responsible for combining a noun root [dag] and a plural affix [z] in English to give the word *dog-s* (i.e. /dag-z/), or in Russian the morphology combines a noun root [vagon] with an inflectional ending [a] according to rules of inflection for Russian, to give the genitive word *vagon-a*. Each morpheme is assumed to have a single constant phonetically defined shape coming out of the morphology (there are a few exceptions such as the fact that the third-person-singular form of the verb *be* in English is [ɪz] and the first-person-singular form of that verb is [æm]). The phonetic realization of any morpheme is subject to rules of phonology, so while the morphology provides the plural morpheme *z* (spelled <s>), the application of phonological rules will make that morpheme be pronounced as [s] as in *cats* or [iz] as in *bushes*.

It is very important to understand that the grammar does not formally derive one word from another. (Some languages seem to have special morphological processes, which we

will not be discussing here, that derive one word from another – clipping such as *Sally* → *Sal* would be an example.) Rather, one word derives from a given abstract root plus whatever affixes are relevant, and a related word derives by adding a different set of affixes to the same abstract root. Accordingly, the plural of a noun in English does not derive from the singular, rather, both the singular *and* the plural forms derive from a common root. No suffix is added to the root in the singular, and the suffix /z/ is added to the root in the plural. The Russian genitive [vagona] also does not derive from the nominative, nor does the nominative derive from the genitive. Rather, both derive from the root /vagon/, where the nominative adds no affix and the genitive adds the affix *-a*.

The underlying form of a word is whatever comes out of the morphology and is fed into the phonology, before any phonological rules have applied. The underlying form of the word [kæts] is /kæt-z/, since that is what results in the morphology by applying the rule that combines a noun root such as *cat* with the plural suffix. The underlying form of the plural word [kæts] is *not* /kæt/, because the plural word has to have the plural morpheme. However, /kæt/ is the underlying form of the singular word [kæt]. There is no phonological rule which inserts *z* or *s* in order to form a plural. The principles for combining roots and affixes are not part of the phonology, and thus there is no need to include rules such as “insert [z] in the plural.” Be explicit about what you assume about morphology in a language, i.e. that there is a plural suffix *-z* in English or a genitive suffix *-a* in Russian. As for the mechanics of phonological analysis, you should assume, for example, that the plural suffix is already present in the underlying form, therefore you do not write a rule to insert the plural suffix since that rule is part of morphology. A phonological analysis states the underlying forms of morphemes, and describes changes in the phonological shape of the root or suffix.

We have concluded that the underlying form of the Russian word [prut] ‘pond’ is /prud/. In arriving at that conclusion, we saw how important it is to distinguish the phonological concept of an underlying form from the morphological concept “basic form,” where the singular form, or an uninflected nominative form would be the morphological “basic form.” An underlying form is a strictly phonological concept and is not necessarily equivalent to an actually pronounced word (even disregarding the fundamental fact that underlying forms are discrete symbolic representations whereas actually pronounced words are acoustic waveforms). The underlying form is a representation that constitutes the foundation for explaining the variety of actual pronouciations found in a morpheme, as determined by phonological context.

The morphologically basic form of the Russian word for pond is the unmarked nominative, [prut], composed of just the root with no inflectional ending. In contrast, the phonological underlying form is /prud/, for as we have seen, if we assume the underlying form to be \*/prut/, we cannot predict the genitive [pruda]. The word \*[prud], with a voiced consonant at the end of the word, does not appear as such in the language, and thus the supposition that the underlying form is /prud/ is an abstraction, given that [prud] by itself is never found in the language – it must be inferred, in order to explain the actual data. The logical basis for that inference is the genitive form [pruda], which actually contains the hypothesized underlying form as a subpart. It is important to understand,

however, that the underlying form of a root may not actually be directly attested in this way in any single word, and we will discuss this point in section 5.6.

### 5.3 Finding the underlying form

A similar problem arises in explaining the partitive and nominative forms of nouns in Finnish. The first step in understanding the phonological alternation seen here is to do a standard preliminary morphological analysis of the data, which involves identifying which parts of a word correlate with each aspect of word structure (such as root meaning or grammatical case). The following examples illustrate that the nominative singular suffix is Ø (i.e. there is no overt suffix in the nominative singular) and the partitive singular suffix is -æ, which alternates with -a if there is a back vowel somewhere before it in the word (we will not be concerned with that vowel alternation in the partitive suffix).

(6)	<i>Nominative sg</i>	<i>Partitive sg</i>	
a.	aamu	aamua	‘morning’
	hopea	hopeaa	‘silver’
	katto	kattoa	‘roof’
	kello	kelloa	‘clock’
	kirja	kirjaa	‘book’
	kylmæ	kylmææ	‘cold’
	koulu	koulua	‘school’
	lintu	lintua	‘bird’
	hylly	hyllyæ	‘shelf’
	kømpelø	kømpeløæ	‘clumsy’
	nækø	nækøæ	‘appearance’
b.	joki	jokea	‘river’
	kivi	kiveæ	‘stone’
	muuri	muuria	‘wall’
	naapuri	naapuria	‘neighbor’
	nimi	nimeæ	‘name’
	kaappi	kaappia	‘chest of drawers’
	kaikki	kaikkea	‘all’
	kiirehti	kiirehtiæ	‘hurry’
	lehti	lehteæ	‘leaf’
	mæki	mækeæ	‘hill’
	ovi	ovea	‘door’

posti	postia	‘mail’
tukki	tukkia	‘log’
æiti	æitiæ	‘mother’
englanti	englantia	‘England’
jærvi	jærveæ	‘lake’
koski	koskea	‘waterfall’
reki	rekeæ	‘sledge’
væki	vækeæ	‘people’

We might assume that the underlying form of the root is the same as the nominative (which has no suffix). The problem which these data pose is that in some nouns, the partitive appears to be simply the nominative plus the suffix -æ ~ -a (for example *muuri* ~ *muuria*), but for other nouns the final vowel alternates, with [i] in the nominative and [e] in the partitive (e.g. *joki* ~ *jokea*). It is obvious that the nature of the following vowel does not explain this alternation, since the same surface-quality suffix vowel can appear after either *e* or *i* – compare *jokea*, *nimeæ* where [e] appears before both [a] and [æ], versus *muuria*, *kiirehtiæ* where [i] appears before these same vowels. Nor can the preceding consonant be called upon to predict what vowel will appear in the partitive, as shown by pairs such as *tukkia*, *kaikkea* versus *lehteæ*, *æitiæ*.

This is an area where there is (potentially) a difference between language-learning pedagogy and a formal linguistic analysis. Faced with the problem of learning the inflectional distinction *muuri* ~ *muuria* versus *joki* ~ *jokea*, a second-language class on Finnish might simply have the student memorize a list of words like *joki* ~ *jokea* where the vowel changes in the inflectional paradigm. From the point of view of linguistic analysis this is the wrong way to look at the question, since it implies that this is not a rule-governed property of the language. However, second-language learning is not the same as linguistic analysis: a class in foreign-language instruction has a different goal from a class in analysis, and some students in a language class may receive greater practical benefit from just memorizing a list of words. Thus it is important to distinguish the teaching method where one learns arbitrary lists, and a theoretically based analysis. One simply cannot predict what vowel will appear in the partitive form if one only considers the pronunciation of the nominative. This means: nominative forms are not the same as underlying forms (something that we also know given the previous Russian example). The underlying representation must in some way contain that information which determines whether there will be a vowel alternation in a given word.

In looking for the phonological basis for this vowel alternation, it is important to realize that the alternation in stem-final vowels is not chaotic, for we find precisely two possibilities, either *i* in the nominative paired with *i* in the partitive, or *i* in the nominative paired with *e* in the partitive – never, for example, *i* paired with *u* or *i* paired with *o*. Moreover, only the vowel *i* enters into such a vowel alternation in Finnish, so there are no nouns with *o* in the nominative which is replaced by *u* in the partitive, nor is *u* in the nominative ever replaced by *o* or any other vowel in the partitive. One final fact about the data in (6) suggests exactly how the right underlying representations can explain this

alternation: of the eight vowels of Finnish [i, y, e, ø, æ, u, o, a], all of them appear at the end of the word, except the vowel *e*. Now, since the stem of the word for ‘name,’ which appears as *nimi* in the nominative, actually appears on the surface as *nime-* in the partitive, it is not at all unreasonable to assume that the underlying form of the stem is in fact /nime/. It would be a bit bizarre to assume an underlying form such as /nima/, since the vowel [a] never appears in that position in any form of this word: the most natural assumption to make is that the underlying form of a morpheme is actually composed of segments found in *some* surface manifestation of the morpheme. On the other hand, the stem of the word for ‘wall’ is pronounced *muuri* in both the nominative and the partitive, and therefore there is no reason to assume that it is underlyingly anything other than /muuri/.

We will then assume that the underlying vowel at the end of the stem is actually most faithfully reflected in the partitive form, and thus we would assume underlying representations such as /joke/, /nime/, /kive/, /lehte/, /ove/ and so on, as well as /muuri/, /naapuri/, /kaappi/, /tukki/ and so on. The underlying form of partitive [joke-a] would thus be /joke-a/, that is, no rule at all is required to explain the partitive. Instead, a rule is needed to explain the surface form of the nominative [joki], which derives from /joke/. A very simple neutralizing rule can explain the surface form of the nominative: underlying word-final *e* is raised to *i*.

#### (7) Final vowel raising

$$\begin{bmatrix} +\text{syl} \\ -\text{rd} \\ -\text{back} \\ -\text{lo} \end{bmatrix} \rightarrow [+hi] / \_ \#$$

*This is a natural assumption but not an absolute rule, as we see in chapter 9. Underlying forms can contain segments not found in any form of the word. Only when there is strong evidence for departing from this assumption are you justified in setting up underlying forms with such abstract elements.*

This rule is neutralizing since the distinction between /i/ and /e/ is neutralized – eliminated – by applying this rule: an underlying /e/ becomes phonetic [i].

Apart from illustrating how important correct underlying forms are, these two examples have also shown that it is dangerous, and incorrect in these two cases, to assume that the “most basic” form of a word according to morphological criteria is also the underlying form of the word. To reiterate: the underlying form of a morpheme is a hypothesis set forth by the analyst, a claim that by assuming such-and-such an underlying form, plus some simple set of rules (which need to be discovered by the analyst), the observed variation in the shape of morphemes can be explained.

**Kerewe.** To better understand the reasoning that leads to correct underlying forms, we investigate other examples. Consider the following data from Kerewe (Tanzania).

(8)	<i>Infinitive</i>	<i>1sg habitual</i>	<i>3sg habitual</i>	<i>Imperative</i>
	kupaamba	mpaamba	apaamba	paamba
	kupaanja	mpaanja	apaanja	paanja
	kupima	mpima	apima	pima
	kupuupa	mpuupa	apuupa	puupa
	kupeket <sup>f</sup> a	mpeket <sup>f</sup> a	apeket <sup>f</sup> a	peket <sup>f</sup> a
	kupiinda	mpiinda	apiinda	piinda
	kuhiiga	mpiiga	ahiiga	hiiga
	kuheeka	mpeeka	aheeka	heeka
	kuhaanja	mpaanja	ahaanja	haanja
	kuheeba	mpeeba	aheeba	heeba
	kuhiima	mpiima	ahiima	hiima
	kuhuuha	mpuuha	ahuuha	huuha

*In this example we only have direct evidence for the change after m, so it would be possible to restrict our rule to the more specific context “after m.” But this would run counter to basic assumptions of science, that we seek the most general explanations possible, not the most restricted ones.*

We notice that every infinitive begins with *ku-*, which we surmise is the prefix for the infinitive; the third-singular habitual form has the prefix *a-*, and the first-singular habitual has the prefix *m-*; the imperative involves no prefix. In addition to segmental prefixes, there is a change in the first consonant of the stem in some verbs, in some contexts. The initial consonant of the verb meaning ‘guide’ alternates between [h] and [p], with [p] appearing in the first-singular habitual after [m] and [h] appearing elsewhere. Since this stem appears in two surface variants, [heeba] and [peeba], two plausible hypotheses are immediately possible: the stem is underlyingly /peeba/, or the stem is underlyingly /heeba/. If we assume that the stem is underlyingly /heeba/, we require a rule to explain the divergence between the predicted form of the first-singular habitual form – we would expect \*[mheeba], \*[mhiima], etc. – and the actual form of the verb, [mpeeba], [mpiima] and so on. Since in fact we do not see the sequence /mh/ anywhere in the data, we might assume the following neutralizing rule.

### (9) Postnasal hardening

$$[+spr.\text{glott}] \rightarrow \begin{bmatrix} -\text{cont} \\ +\text{ant} \end{bmatrix} / [+nasal] \_\_$$

If, on the other hand, we assume that the root is underlyingly /peeba/, we would need a rule which changes /p/ into [h] when not preceded by a nasal – in other words, when preceded by a vowel or by nothing. There is no single property which groups together

word-initial position and vowels. Thus, the supposed rule changing /p/ to [h] would have to be a disjunction of two separate environments.

$$(10) \begin{bmatrix} +\text{ant} \\ -\text{cor} \\ -\text{voice} \end{bmatrix} \rightarrow \begin{bmatrix} +\text{spr.glott.} \\ +\text{cont} \\ -\text{ant} \end{bmatrix} / \begin{cases} \{\text{V}\} \\ \{\#\} \end{cases} -$$

This suggests that rule (10) is wrong.

More important than the greater complexity of the rule entailed by assuming that the word for ‘guide’ is underlyingly /peeba/, it is empirically wrong: rule (10) implicitly claims that /p/ should always become [h] word initially or after a vowel, but this is falsified by forms such as *kupaamba*, *apaamba*, *paamba* ‘adorn’ and *kupaanya*, *apaanya*, *paanya* ‘arrange’. If we assume the stems uniformly begin with /p/, then we cannot predict whether the imperative or infinitive has [h] (*kuhaanya*) or [p] (*kupaanya*). On the other hand, if we assume an underlying contrast between initial /h/ and initial /p/ – i.e. *haanya* ‘create’, *paanya* ‘arrange’ – then we can correctly distinguish those stems which begin with /h/ from those which begin with /p/ when no nasal precedes, as well as correctly neutralizing that distinction just in case the stem is preceded by a nasal (*mpaanya* ‘I create’; ‘I arrange’).

The rule statement in (9) exploits a widely-used notion about how rules apply, known as **structure preservation**. Notice that the structural change specified mentions only that /h/ becomes [-cont,+ant], which are two features that characterize the difference between /h/ and [p]. There are two other actual changes in features which are not explicitly mentioned in the rule, namely that the segment becomes [+cons,-son]. These values can be automatically predicted from the fact that in this language, there is only one voiceless [+ant,-cor] stop, namely [p]. The idea underlying structure preservation is that each language defines an inventory of segments, and the structural change of a rule changes from one sound within the inventory to another sound. Specifying that change as [-cont,+ant,+cons,-son] fully specifies what the result of the rule is, but [-cont,+ant] identifies the same unique segment of the language, more economically.

**English plurals.** A further illustration of how to determine the correct underlying representation comes from English. As the following examples illustrate, the surface form of the plural suffix varies between [s] and [z] (as well as [iz], to be discussed later).

(11)	kæps	caps	kæbz	cabs	klæmz	clams
	kæts	cats	kædz	cads	kænz	cans
	kaks	cocks	kagz	cogs	kaɪz	cars
	pruwfs	proofs	huvz	hooves	gəlz	gulls
			flijz	fleas		
			plæwz	plows		
			pj̥iez	purees		

The generalization regarding distribution is straightforward: [s] appears after a voiceless segment, and [z] appears after a voiced one (be it an obstruent, a liquid, nasal or a vowel).

This same alternation can be found in the suffix marking the third singular present-tense form of verbs.

(12)	slæps	slaps	stæbz	stabs	slæmz	slams
	hits	hits	hajdz	hides	kænz	cans
	powks	pokes	dɪgz	digs	hæŋz	hangs
	læfs	laughs	θrajvz	thrives	hɪljz	heals
	pɪθs	piths	bejðz	bathes	hɪljz	hears
			flajz	flies	vjtowz	vetos

If we suppose that the underlying form of the affixes for noun plural and third singular present verbs are /z/, then we would assume the following rule to derive the phonetic variant [s].

$$(13) \quad [-son] \rightarrow [-voice] / [-voice] \underline{\quad}$$

On the other hand, if we were to assume that these suffixes are underlyingly /s/, we would assume the following rule.

$$(14) \quad [-son] \rightarrow [+voice] / [+voice] \underline{\quad}$$

In terms of the simplicity and generality of these two rules, the analyses are comparable. Both formulations require the same number of phonetic specifications to state the rule, and both formulations apply to general and phonetically natural classes. However, the two analyses differ quite significantly in terms of their overall predictions for English. The implicit prediction of the first rule (13) is that there should be no voiced obstruents after voiceless segments in English, since that rule would devoice all such obstruents. This generalization seems to be correct: there are no words like \*[jøkd], \*[pɪfz], \*[sdap]. The implicit prediction of the second rule (14) is different: that rule implies that there should be no voiceless segments after any voiced segments. This is manifestly incorrect, as shown by the existence of words such as [hɪs] *hiss*, [pæθ] *path*, [dæns] *dance*, [fals] *false*. We prefer a hypothesis which makes the correct prediction about the structure of the language as a whole, and thus we select the underlying form /z/ and a rule devoicing obstruents after voiceless segments. Looking for such asymmetries plays an important role in determining which of two hypotheses is the correct one.

The alternation *z* ~ *s* is not limited to the two affixes -*z* ‘plural’ and – *z* ‘3sg present tense.’ The rule of devoicing can also be seen applying to the possessive suffix -*z*.

(15)	<i>Noun</i>	<i>Noun + poss.</i>
	kæt	kæts
	sləg	sləgz
	klæm	klæmz

snow

snowz

snow

Moreover, certain auxiliary verbs such as *has* [hæz] and *is* [ɪz] undergo a reduction in casual speech, so that they appear simply as [s] or [z], the choice between these two being determined by the devoicing rule which we have motivated. There is a difference in the pronunciation of ‘has’ and ‘is’ when the auxiliary is pronounced as [hæz] or [ɪz], but when the auxiliary and noun are fused into one, they are pronounced the same.

(16) <i>Noun + has</i>	<i>Reduced</i>	<i>Noun + is</i>	<i>Reduced</i>	
d <sup>3</sup> æk hæz ij?n̩	d <sup>3</sup> æks ij?n̩	d <sup>3</sup> æk iz ijriŋ	d <sup>3</sup> æks ijriŋ	Jack
pæt hæz ij?n̩	pæts ij?n̩	pæt iz ijriŋ	pæts ijriŋ	Pat
d <sup>3</sup> en hæz ij?n̩	d <sup>3</sup> enz ij?n̩	d <sup>3</sup> en iz ijriŋ	d <sup>3</sup> enz ijriŋ	Jen
bab hæz ij?n̩	babz ij?n̩	bab iz ijriŋ	babz ijriŋ	Bob
d <sup>3</sup> ow hæz ij?n̩	d <sup>3</sup> owz ij?n̩	d <sup>3</sup> ow iz ijriŋ	d <sup>3</sup> owz ijriŋ	Joe

The devoicing rule (13) automatically explains the alternation in the surface shape of the consonant here as well.

**Jita tone.** It is important to look for correlations which may lead to causal explanations, in analyzing data. Consider the following data from Jita (Tanzania), concentrating on the tones of morphemes (H or high tone is marked with acute accent, L or low-toned syllables are unmarked).

(17)	a.	okuβuma	‘to hit’	okusiβa	‘to block’
		okuβumira	‘to hit for’	okusiβira	‘to block for’
		okuβumana	‘to hit e.o.’	okusiβana	‘to block e.o.’
		okuβumirana	‘to hit for e.o.’	okusiβiranā	‘to block for e.o.’
	b.	okulúma	‘to bite’	okukúβa	‘to fold’
		okulumíra	‘to bite for’	okukuβíra	‘to fold for’
		okulumána	‘to bite e.o.’	okukuβána	‘to fold e.o.’
		okulumírana	‘to bite for e.o.’	okukuβíranā	‘to fold for e.o.’

We can conclude that there is a prefix *oku-* perhaps marking the infinitive, a suffix *-a* appearing at the end of every verb, and two suffixes *-ir-* ‘for’ and *-an-* ‘each other.’ There are also root morphemes: *-βum-* ‘hit,’ *-siβ-* ‘block,’ as well as *-lúm-* ‘bite’ and *-kuβ-* ‘fold.’ We decide that ‘bite’ and ‘fold’ underlyingly have H tones in part based on the fact that there actually is an H tone on the vowels of these roots in the simplest verb forms.

In addition, we observe that the suffixes *-ir-* and *-an-* have H tone when they come immediately after these verb roots. The suffixes do not have H tone after the first set of roots: appearance of H on the suffix is correlated with which morpheme immediately precedes the suffix. Since this unpredictable property is correlated with the preceding root morpheme, it must therefore be an aspect of the underlying form of the preceding morpheme.

We thus explain the H tone on these suffix morphemes by positing that [oku-lum-án-a] derives from underlying /oku-lúm-an-a/, by applying a rule of tone shift which shifts a H tone rightward to the following syllable, as long as the syllable is not word-final. Because of the restriction that H does not shift to a final syllable, the underlying H surfaces unchanged in [okulúma].

Now consider the following data.

(18)	okumuβúma	'to hit him/her'	okumusíβa	'to block him/her'
	okumuβúmira	'to hit for him/her'	okumusíβira	'to block for him/her'
	okut'íβúma	'to hit it'	okut'ísiβa	'to block it'
	okut'íβúmira	'to hit for it'	okut'ísiβira	'to block for it'

When the L-toned roots of (17a) stand after the object prefixes *-mu-* 'him/her' and *-t'i-* 'it,' they have an H tone at the beginning of the root. Again, since the presence of the H is correlated unpredictably with the prefixes *-mu-* and *-t'i-*, we hypothesize that the tones are part of the underlying representation of the prefixes – the prefixes are /mú/ and /t'i/, and the H tone shifts to the right by the tone shift rule which we have already posited.

*Admittedly we did not formalize that rule, because it is difficult to state the rule with the rule-writing devices that we have at this point. This problem will be addressed in chapter 10. When you don't know how to formalize a rule, it is always better to state the rule in plain language, rather than ignore the rule.*

## 5.4 Practice at problem solving

You should now be able to apply this reasoning to data which pose analogous problems; a series of examples are given in this section for practice.

**Chamorro vowel alternations.** There are alternations in the quality of vowels in initial syllables in some contexts seen in the following data from Chamorro (Mariana Islands).

(19)	gwíhən	'fish'	i gwíhən	'the fish'
	gúmə?	'house'	i gímə?	'the house'
	káttə	'letter'	jo? káttə	'a letter (object)'
			i káettə	'the letter'
	t <sup>s</sup> úpa	'cigarettes'	i t <sup>s</sup> ípa	'the cigarettes'
	fíno?	'talk'	mi fíno?	'lots of talk'
	túnū?	'to know'	en túnū?	'you know'
	t <sup>s</sup> úgo?	'juice'	mi t <sup>s</sup> ígo?	'lots of juice'
	sóñsuj	'village'	i séñsuj	'the village'
	húlu?	'up'	sæn hílu?	'upward'
	pét <sup>s</sup> u	'chest'	i pét <sup>s</sup> u	'the chest'
	tómtu	'knee'	i tému	'the knee'

ótdut	'ant'	mi étdut	'lots of ants'
óksu?	'hill'	gi éksu?	'at the hill'
dáŋkulu	'big one'	i dáŋkulu	'the big one'
láhi	'male'	i láhi	'the male'
lágu	'north'	sæn lágu	'toward north'
pulónnun	'trigger fish'	i pulónnun	'the trigger fish'
mundóŋgu	'cow's stomach'	i mundóŋgu	'the cow's stomach'
putamonéda	'wallet'	i putamonéda	'the wallet'

What underlying representations, and what rule or rules, are required to account for these data? When you answer this question, you should consider two hypotheses which differ in terms of what form is taken to be underlying – what are the two most obvious ways of treating these alternations? One of these hypotheses is clearly wrong; the other is the correct hypothesis.

In formalizing the rule, take note of the vowel inventory observed in the language and see how the notion of a structure preserving effect can simplify the statement of your rule.

**Korean.** Now consider the following data from Korean. The first column, the imperative, seems to involve a vowel suffix. One reason to think that there is an imperative suffix is that every imperative ends either in the vowel *a* or in *ə* (the choice between *a* versus *ə* is based on the vowel which precedes that suffix, /a/ or /o/ versus other vowels, and can be ignored here). A second reason comes from comparing the imperative and the plain present forms. Comparing *ana* and *anninta*, or *kama* and *kamninta*, we can see that for each verb, the portions common to both the imperative and the plain present are respectively *an-* and *kam-*. From this we deduce that there must be a suffix, either *-a* or *-ə*, which marks the imperative, and another suffix *-ninta* which marks the plain present.

(20) <i>Imperative</i>	<i>Plain present</i>	
ana	anninta	'hug'
kama	kamninta	'wind'
sinə	sinninta	'wear shoes'
t'atimə	t'atimninta	'trim'
nəmə	nəmninta	'overflow'
nama	namninta	'remain'
t <sup>lh</sup> ama	t <sup>lh</sup> amninta	'endure'
ipə	imninta	'put on'

kupə	kumninta	'bend'
t'əpə	t'əmninta	'fold'
tata	tanninta	'close'
put <sup>h</sup> ə	punninta	'adhere'
t'ot <sup>h</sup> a	t'onninta	'follow'
məkə	məŋninta	'eat'
sək'ə	səŋninta	'mix'
tak'a	tan̩ninta	'polish'
t'ukə	t'uŋninta	'die'
ikə	in̩ninta	'ripen'

What are the underlying forms of these verb stems, and what phonological rule or rules are required to account for the variations that are seen in the surface shape of the various stems?

**Koasati.** What is the underlying form of the first-singular possessive prefix in Koasati (Louisiana), and what phonological rule applies in these examples?

(21) <i>Noun</i>	<i>My N</i>	
apaht <sup>f</sup> á	amapaht <sup>f</sup> á	'shadow'
asikt <sup>f</sup> í	amasikt <sup>f</sup> í	'muscle'
ilkanó	amilkanó	'right side'
ifá	amifá	'dog'
a:pó	ama:pó	'grandmother'
iskí	amiskí	'mother'
pat <sup>f</sup> okkó:ka	ampat <sup>f</sup> okkó:ka	'chair'
towá	antowá	'onion'
kastó	aŋkastó	'flea'
bajá:na	ambajá:na	'stomach'
tá:ta	antá:ta	'father'
t'ofkoní	ajt <sup>f</sup> ofkoní	'bone'
kitiilká	aŋkitiilká	'hair bangs'
toní	antoní	'hip'

**Matuumbi.** What phonological rules pertaining to consonants operate in the following examples from Matuumbi. What are the underlying forms of the stems of the words for 'rope,' 'palm,' 'tongue,' 'piece of wood,' 'pole' and 'covered'? Ignore tonal changes.

(22) <i>Singular</i>	<i>Plural</i>	
lugói	ŋgói	'rope'

lugolóká	ŋgolóká	'straight'
lubáu	mbáu	'rib'
lubágalo	mbagálo	'lath'
lud <sup>3</sup> iíŋgjá	jnd <sup>3</sup> iíŋgjá	'entered'
lulaála	ndaála	'pepper'
lulímí	ndími	'tongue'
lulíndírlá	ndíndírlá	'guarded'
lupaláaí	mbaláaí	'bald head'
lupaálá	mbaálá	'wanted'
lutéelá	ndeelá	'piece of wood'
lukílico	ŋgilígo	'place for initiates'
lukíli	ŋgíli	'palm'
lujímá	jnd <sup>3</sup> íma	'pole'
lujóka	jnd <sup>3</sup> óka	'stomach worm'
lujúsí	jnd <sup>3</sup> úsi	'bee'
lujuwé	jnd <sup>3</sup> úwe	'pumpkin plant'
luwíkíljá	ŋg <sup>w</sup> íkíljá	'covered'

A certain degree of uncertainty regarding the exact underlying form of the plural prefix is expected. However, the underlying form of the stem should be clear, and should be the focus of your analysis. You should be able to explain these alternations with two rules. In formalizing the rules, pay attention to the concept of structure preservation in rules.

## 5.5 Underlying forms and sentence-level phonology

In the examples which we have considered so far, we have been comparing morphologically related words, such as a nominative and a genitive, and we have seen that an underlying distinction may be preserved in one word in a particular inflected form (because in that inflected form the conditions for applying the phonological rule are not satisfied), but the difference is neutralized in a related word where the conditions for the rule are present. We now consider two additional cases where underlying distinctions are neutralized depending on context, and the neutralization takes place within one and the same word, depending on where the word appears in a sentence. What this shows is that phonology is not just about variations in pronunciation between words, but also includes variations in the pronunciation of a word in different sentential contexts.

### 5.5.1 Korean final Cs

The first case is a rule of Korean that nasalizes stops before nasal consonants (a rule that we have seen operating within words in the preceding section). The first set of examples shows the word for 'rice' when said alone, or when it is followed by various words which begin with oral consonants and vowels. In these data, the last consonant of the word for

‘rice’ is pronounced as [p]. In the second set of examples, the word which follows ‘rice’ begins with a nasal, and in that case the final consonant of the word for ‘rice’ is pronounced as [m].

- (23) a. **pap** ‘rice’  
**pap** anməkət’ ‘didn’t eat rice’  
rice didn’t eat
- pap** winmake tuət’ ‘put rice on the upper floor’  
rice on-upper-floor put
- pap** saməkət’ ‘ate rice at a store’  
rice ate-at-store
- pap** totuki humt<sup>h</sup>əkat’ ‘a thief stole rice’  
rice thief-(subj) stole
- b. **pam** mani məkəla ‘eat rice a lot’  
rice lot eat
- pam** mək-imjən ‘if eats rice’  
rice eat-if
- pam** nəmu masik’ e məkət’ ‘I enjoyed rice quite a lot’  
rice very deliciously ate
- pam** nək<sup>h</sup>o ‘add rice’  
rice add

Compare those examples with the following examples with the word for ‘chestnut’.

- (24) a. **pam** ‘chestnut’  
**pam** anməkət’ ‘didn’t eat chestnut’  
chestnut didn’t-eat
- pam** winmake tuət’ ‘put chestnut on the upper floor’  
chestnut on-upper-floor put
- pam** saməkət’ ‘ate chestnut at a store’  
chestnut ate-at-store
- pam** totuki humt<sup>h</sup>əkat’ ‘a thief stole chestnut’  
chestnut thief-(subj) stole
- b. **pam** mani məkəla ‘eat chestnut a lot’  
chestnut lot eat
- pam** mək-imjən ‘if eats chestnut’  
chestnut eat-if
- pam** nəmu masik’ e məkət’ ‘I enjoyed chestnut quite a lot’  
rice very deliciously ate

chestnut very deliciously ate	
pam nək <sup>h</sup> o	‘add chestnut’
chestnut add	

In fact the (b) phrases above are actually ambiguous as to whether the word being pronounced means ‘chestnut’ or ‘rice.’

The last consonant of the word for ‘chestnut’ is always [m], so we would presume that the underlying form of that word is /pam/. Since the word for ‘rice’ varies between [pap] and [pam], and since we know that the underlying form cannot be /pam/ (this is the underlying form of ‘chestnut,’ and ‘chestnut’ cannot have the same underlying form as ‘rice’ since they do not behave the same), we conclude that the underlying form of the word for ‘rice’ is /pap/, and a nasalization rule changes /p/ (in fact, all stops) to nasals before a nasal. Whether a word undergoes that rule depends on what follows the final consonant. One and the same word can be pronounced differently depending on the properties of the phrase in which it appears.

### 5.5.2 Matuumbi tone

In the Korean case which we just considered, it happens that the underlying form of the word is the same as the way the word is pronounced when it is said alone. This situation does not hold in Matuumbi, where one has to know how a word is pronounced when it is *not* at the end of an utterance, in order to determine the underlying form of the word. The words in (25) have an H tone (marked with an acute accent) on the second vowel from the beginning of the word when said alone. When another word follows, they seem to lose that H tone.

(25)	kiwíkíljo	‘cover’	ŋga kiwíkíljo lí	‘it isn’t a cover’
	lubágalo	‘lath’	ŋga lubagalo lí	‘it isn’t a lath’
	mikóta	‘sugar canes’	ŋga mikota lí	‘it isn’t sugar canes’
	ŋguúŋguni	‘bedbug’	ŋga ŋguúŋguni lí	‘it isn’t a bedbug’
	lukólogo	‘brewery’	ŋga lukologo lí	‘it isn’t a brewery’
	mabáando	‘thighs’	ŋga mabaando lí	‘it isn’t thighs’
	kikóloombe	‘shell’	ŋga kikoolombe lí	‘it isn’t a shell’
	lipítanuŋgu	‘rainbow’	ŋga lipitanuŋgo lí	‘it isn’t a rainbow’

In contrast, the words of (26), which also have an H tone on the second vowel from the beginning of the word when the word is said alone, keep their H tone when another word follows.

(26)	lukóŋgobe	‘wood’	ŋga lukóŋgobe lí	‘it’s not wood’
	kitókotoku	‘quelea bird’	ŋga kitókotoku lí	‘it’s not a quelea’
	diíwai	‘wine’	ŋga diíwai lí	‘it’s not wine’
	lukóongono	‘chicken leg’	ŋga lukóongono lí	‘it’s not a leg’

lukóongowe	'marble'	ŋga lukóongowe lí	'it's not marble'
matógolo	'waterbucks'	ŋga matógolo lí	'it's not waterbucks'
mivíriiŋgo	'circles'	ŋga mivíriiŋgo lí	'it's not circles'
kijógojo	'bird (sp)'	ŋga kijógojo lí	'it's not a bird'
kikálaango	'pan'	ŋga kikálaango lí	'it's not a pan'

There are no words in Matuumbi which are toneless when said by themselves, thus *\*kitekeljo* said by itself is an unattested kind of word. There is a clear contrast in tonal behavior between the words in (25) where the presence of an H tone on the second vowel depends on whether the word is said alone or is followed by another word, and those in (26) where the second vowel always has an H tone. The solution to this puzzle is that the words in (26) have an underlying H tone on their second vowel, and thus nothing happens to that tone; but the words in (25) have no underlying H, and instead get an H at the end of an utterance by a rule that assigns an H tone to the second vowel of a toneless word which comes at the end of an utterance. Thus in the case of Matuumbi tone, the contrast between underlyingly toneless words and words with underlying H is best revealed by looking at the word when it appears *not* by itself: it is the citation form of the word that undergoes the neutralization rule, which is the opposite of the situation we just encountered in Korean.

## 5.6 Underlying forms and multiple columns in the paradigm

The following data from Samoan illustrate the very important point that it is wrong to think of deriving underlying forms by chopping off affixes from some single column of data. In the first set of examples, our initial task is to deduce the underlying form of each of the verb roots and the affix for the perfective form.

(27) Simple	Perfective	
olo	oloia	'rub'
lafo	lafoia	'cast'
aŋa	aŋaia	'face'
usu	usuia	'get up and go early'
tau	tauia	'reach a destination'
taui	tauiia	'repay'
sa:ili	sa:iliia	'look for'
vanai	vanaia	'face each other'
paŋi	paŋia	'touch'
naumati	naumatia	'be waterless'
sa:uni	sa:unia	'prepare'
seŋi	seŋia	'be shy'
lele	lelea	'fly'

su?e	su?ea	'uncover'
ta?e	ta?ea	'smash'
tafe	tafea	'flow'
ta:upule	ta:upulea	'confer'
palepale	palepalea	'hold firm'

Examples such as *oloia*, *ayaia* and *usuia* suggest that the perfective suffix is *-ia*, and the simple form of the verb reflects the underlying form of the root. Examples such as *seyi* ~ *seyia* or *lele* ~ *lele* suggest a phonological rule, since the combination of the presumed stems *seyi* and *lele* with the perfective affix *-ia* would result in the incorrect forms *\*seyia*, *\*leleia*. However, this problem can be corrected by positing a phonological rule which deletes a front vowel when it is preceded by a front vowel. In the formalization of the rule, we say that the second front vowel is replaced by zero, which means that it is deleted.

(28) *Vowel-cluster reduction*

$$\left[ \begin{array}{l} +\text{syl} \\ -\text{back} \end{array} \right] \rightarrow \emptyset / \left[ \begin{array}{l} +\text{syl} \\ -\text{back} \end{array} \right] \_$$

An alternative hypothesis would be that [i] is inserted between a back vowel and the vowel [a], if we were to presume that the perfective suffix is underlyingly /a/.

$$(29) \quad \emptyset \rightarrow \left[ \begin{array}{l} +\text{syl} \\ +\text{hi} \\ -\text{back} \end{array} \right] / \left[ \begin{array}{l} +\text{syl} \\ +\text{back} \end{array} \right] \_ [+low]$$

This latter rule is more complicated than (28). Additional data will show that, in addition, this would just be plain wrong. We abandon the idea of inserting the vowel *i* and conclude that the underlying form of the perfective suffix must be *-ia*, hence there must be a rule deleting a front vowel after a front vowel. We would then conclude that the underlying representation of roots is best revealed in the simple verb, rather than the perfective, since the simple form of the verb shows whether the stem ends with /i/, a vowel which may be deleted in the perfective.

A rather different conclusion about arriving at underlying forms would have to be drawn from the following additional Samoan examples.

(30)	<i>Simple</i>	<i>Perfective</i>	<i>Simple</i>	<i>Perfective</i>		
	tu:	tu:lia	'stand'	au	aulia	'flow on'
	tau	taulia	'cost'	ma:tau	ma:taulia	'observe'
	?alo	?alofia	'avoid'	ili	ilifia	'blow'
	oso	osofia	'jump'	ulu	ulufia	'enter'
	sao	saofia	'collect'	tajo	tajofia	'take hold'
	asu	asunja	'smoke'	soa	soanjia	'have a friend'

pole	poleñia	'be anxious'	fesili	fesiliñia	'question'
ifo	ifonja	'bow down'	?ote	?otenja	'scold'
ula	ulañia	'mock'	tofu	tofunja	'dive'
milo	milosia	'twist'	la?a	la?asia	'step'
valu	valusia	'scrape'	taji	tajisia	'cry'
vela	velasia	'be cooked'	motu	motusia	'break'
api	apitia	'be lodged'	mata?u	mata?utia	'fear'
e?e	e?etia	'be raised'	sau	sautia	'fall'
lava:	lava:tia	'be able'	o?o	o?otia	'arrive'
u:	u:tia	'grip'	ufi	ufitia	'cover'
puni	punitia	'be blocked'	tanu	tanumia	'cover up'
si?o	si?omia	'be enclosed'	mo?o	mo?omia	'admire'
ŋalo	ŋalomia	'forget'	tao	taomia	'cover'
sopo	sopo?ia	'go across'	fana	fana?ia	'shoot'

Here, we see that the perfective form of the verb contains a consonant which is not present in the simple form. That consonant can be any one of *l*, *f*, *ŋ*, *s*, *t*, *m* or *?* , given these data. An attempt to predict the nature of that consonant by an insertion rule proves fruitless. We could attempt to insert an appropriate consonant on the basis of the preceding vowel: but while *l* appears after *u*, so do *f* ([ulufia]), *ŋ* ([tofunja]) and *s* ([valusia]); and while *f* appears after *o*, so do *ŋ* ([ifonja]), *m* ([ŋalomia]), and *s* ([milosia]). In short, it is simply impossible to predict from anything in the environment what the consonant of the perfective is going to be, if we start with the simple form as the underlying form: that consonant must be part of the underlying representation of the root. Thus the underlying forms of this second set of roots would be as follows.

(31) tu:l	'stand'	aul	'flow on'
taul	'cost'	ma:taul	'observe'
?alof	'avoid'	ilif	'blow'
osof	'jump'	uluf	'enter'
saof	'collect'	ta?of	'take hold'
asuŋ	'smoke'	soaŋ	'have a friend'
poleñ	'be anxious'	fesiliñ	'question'
ifonj	'bow down'	?oteñ	'scold'
ulañ	'mock'	tofunj	'dive'
milos	'twist'	la?as	'step'
valus	'scrape'	tajis	'cry'
velas	'be cooked'	motus	'break'

apit	'be lodged'	mata?ut	'fear'
e?et	'be raised'	saut	'fall'
lava:t	'be able'	o?ot	'arrive'
u:t	'grip'	ufit	'cover'
punit	'be blocked'	tanum	'cover up'
si?om	'be enclosed'	mo?om	'admire'
ŋalom	'forget'	taom	'cover'
sopo?	'go across'	fana?	'shoot'

The postulation of underlying consonants at the end of these roots entails the addition of a phonological rule, in order to account for the surface form of the simple verb where there is no final consonant. Noting that no word ends in a consonant phonetically in these examples, we can postulate the following rule of final consonant deletion.

### (32) Final consonant deletion

$$C \rightarrow \emptyset / \_ \#$$

The underlying forms of these verbs can be heuristically derived by eliminating the perfective affix *-ia* from the perfective form. However, notice that we made a different heuristic assumption for the first group of roots, which underlyingly ended in a vowel. The point is that an underlying representation is whatever is required to correctly predict all of the surface variants of a given morpheme: it does not necessarily derive from any one column in a paradigm.

It is also important to understand the difference between saying that the underlying form *is* the simple form, or *is* the perfective form, and saying that we may best *learn* what the underlying form is by looking at the perfective, or simple form, or some other form. The underlying form of the word for 'stand' is /tu:l/. We learn that this is the underlying form by comparing the simple form [tu:] and the perfective [tu:lia] and understanding that the perfective form preserves important information about the underlying form that is lost in the simple form. But the perfective form itself is [tu:lia] – this is not the underlying form.

**Palauan.** The language Palauan provides a second clear illustration of the point that one cannot always arrive at the correct underlying representation by looking at any single column in the paradigm. In this language, the underlying form of the word does not actually surface as such in any form of a word. Consider the following examples:

(33)	Present middle	Future innovative	Future conservative	
	mədáŋəb	dəŋəbáll	dəŋóbl	'cover'
	mətē?əb	tə?əbáll	tə?íbl	'pull out'
	məŋétəm	ŋətəmáll	ŋətóml	'lick'
	mətábək	təbəkáll	təbákl	'patch'
	mə?árəm	?ərəmáll	?əróml	'taste'

məsésəb	səsəbáll	səsóbl	‘burn’
---------	----------	--------	--------

The prefix for the present middle is apparently /mə/, the future suffix (found in the future conservative and the future innovative) is *-l*, and the innovative suffix is *-al*. The position of stress can be predicted by a simple rule: the final syllable is stressed if it ends in two consonants, otherwise the second to last (penultimate) syllable is stressed.

The fundamental problem of Palauan is how to predict vowel quality in the root. Notice that the root meaning of the word for ‘cover’ has three surface realizations: *dáŋəb*, *dəŋəb* and *dəŋób*. Looking at all of the data, we notice that the only full vowel in the word is the stressed vowel, which suggests that unstressed vowels are neutralized to schwa.

(34) *Unstressed vowel reduction*

$$\begin{bmatrix} +\text{syl} \\ -\text{stress} \end{bmatrix} \rightarrow \begin{bmatrix} -\text{hi} \\ -\text{lo} \\ +\text{back} \\ -\text{rd} \end{bmatrix}$$

Note that this rule has no context: it does not matter what precedes or follows the unstressed vowel.

In order to predict that the stressed first vowel in the word for ‘cover’ is [a], that choice of vowel must be part of the underlying representation, giving the partial solution /daŋVb/. In contrast, the first vowel of the word for ‘pull out’ must be specified as [e], since that is the vowel which appears in this word when the first vowel is stressed, giving /teʔib/. By the same reasoning, the second vowel of the word for ‘cover’ must be [o], since that is the realization which the vowel has when it is stressed, and the second vowel of the word for ‘pull out’ must be [i]. Thus, the underlying forms of the stems given above would be the following.

(35)	daŋob	‘cover’	teʔib	‘pull out’
	ŋetom	‘lick’	tabak	‘patch’
	?arom	‘taste’	sesob	‘burn’

The underlying form of a verb in Palauan is a rather abstract object, something which is never revealed in any single surface form. Rather, it must be deduced by looking at information which is manifested in a number of different morphologically related words derived from a single stem.

**English.** A similar example can be found in English, as the following examples show. We will ignore other alternations and focus only on vowel alternations. Thus for example, alternations such as the one between *k* and *s* can be ignored. There are many idiolectal differences in the pronunciation of certain words such as *economy*, where some people pronounce the word as [i:j'kanəmij] and others pronounce it as [ə'kanəmij]: only attempt to account for the pronunciations given here.

(36)	'manətown	‘monotone’	mə'natənij	‘monotony’
------	-----------	------------	------------	------------

'teləgræf	'telegraph'	tə'legrəfij	'telegraphy'
'ɛpəgræf	'epigraph'	ə'pigrəfij	'epigraphy'
'rələtɪv	'relative'	rə'lejʃən	'relation'
ə'kanəmij	'economy'	,ɛkə'namik	'economic'
'dijfekt	'defect (noun)'	də'fektiv	'defective'
'dəməkræt	'democrat'	də'makrəsij	'democracy'
'itəlij	'Italy'	ə'tæljən	'Italian'
'hamənim	'homonym'	hə'manəmij	'homonymy'
fə'nɛtiks	'phonetics'	,fownə'tiʃən	'phonetician'
stə'tistiks	'statistics'	,stætə'stiʃən	'statistician'
rə'siprəkl̩	'reciprocal'	,resə'prasətij	'reciprocity'
fə'nalədʒij	'phonology'	,fownə'ladʒəkl̩	'phonological'
'ladʒɪk	'logic'	lə'dʒɪʃn	'logician'
'sinənim	'synonym'	sə'nanəmij	'synonymy'
ə'rɪstəkræt	'aristocrat'	,erə'stakrəsij	'aristocracy'

As in Palauan, there is an alternation between stressed full vowel and unstressed schwa. We assume underlying stems with multiple full vowels, e.g. /manatown/, /teləgræf/, /ɛpəgræf/, /dəməkræt/, /fownalədʒij/, etc. But not every unstressed vowel is reduced: cf. for example *'relətɪv*, *'dijfekt*, *'manənim* where the unstressed vowel is in a closed syllable (followed by one or more consonants within that syllable).

**Tonkawa: reaching the analysis step-by-step.** The following examples will illustrate the logic that leads to seeing the correct underlying forms, in explaining variations found in the verb root in Tonkawa (Texas). You must first give a morphological analysis of the data, identifying the morphemes for progressive, present, first-singular object, and third-plural object; you must also set forth initial hypotheses about the underlying forms of roots. The data to be accounted for are as follows.

(37) picno?	'he cuts'	picnano?	'he is cutting'
wepceno?	'he cuts them'	wepcenano?	'he is cutting them'
kepceno?	'he cuts me'	kepcenano?	'he is cutting me'
notxo?	'he hoes'	notxono?	'he is hoeing'
wentxo?	'he hoes them'	wentoxono?	'he is hoeing them'
kentxo?	'he hoes me'	kentoxono?	'he is hoeing me'
netlo?	'he licks'	netleno?	'he is licking'
wentalo?	'he licks them'	wentaleno?	'he is licking them'
kentalo?	'he licks me'	kentaleno?	'he is licking me'
naxco?	'he makes fire'	naxceno?	'he is making fire'

wenxaco?	'he makes them fire'	wenxaceno?	'he is making them fire'
kenxaco?	'he makes me fire'	kenxaceno?	'he is making me fire'
jamxo?	'he paints a face'	jamxano?	'he is painting a face'
wejmaxo?	'he paints their face'	wejmaxano?	'he is painting their face'
kejmaxo?	'he paints my face'	kejmaxano?	'he is painting my face'
nawlo?	'he spreads'	nawleno?	'he is spreading'
wenwelo?	'he spreads them'	wenweleno?	'he is spreading them'
kenwelo?	'he spreads me'	kenweleno?	'he is spreading me'

Every word in this set ends with *o?*, and the verb forms have a third-person subject, suggesting that *-o?* is mark third-person subject. Comparing the habitual present and present progressive form, we see that the present progressive is marked by a suffix, *-n-* or *-Vn-* before the suffix *-o?*. An object is marked by a prefix, *we-* for third-plural object and *ke-* for first-singular object. What remains is the verb root.

We have two unresolved questions: whether the suffix for the progressive is *-n-*, or is there a vowel in the suffix; and, what is the underlying form of the verb root. To resolve the first question, we look at verbs with no object:

(38)	picno?	picnano?
	notxo?	notxono?
	netlo?	netleno?
	naxco?	naxceno?
	jamxo?	jamxano?
	nawlo?	nawleno?

We might think that the vowel before *-n-* is part of the progressive suffix, but if it were part of that suffix, it should have a constant underlying form and all surface variants of that vowel should be derived by some simple rule. Clearly, the vowel before *n* ranges over *a*, *o* and *e*, and there is no reasonable way to predict which vowel is present. Since that information is governed by which root appears before the suffix, the vowel is part of the underlying form of the verb root. Thus, we arrive at the following partial answer to the question about the underlying forms of the verb roots:

(39)	/picna/	'cut'
	/notxo/	'hoe'
	/netle/	'lick'
	/naxce/	'make a fire'
	/jamxa/	'paint a face'
	/nawle/	'spread'

The progressive form of the verb can be derived straightforwardly by adding the two affixes *-n-* and *-o?*. The habitual present involves the application of a further

phonological process. Based on our hypotheses regarding the underlying forms of the verb roots, we predict the following underlying forms for the habitual forms.

(40) <i>Predicted form</i>	<i>Actual surface form</i>	
picnao?	picno?	‘cut’
notxoo?	notxo?	‘hoe’
netleo?	netlo?	‘lick’
naxceo?	naxco?	‘make a fire’
jamxao?	jamxo?	‘paint a face’
nawleo?	nawlo?	‘spread’

The underlying form is whatever is given by the morphological component, so in this case it would be the root plus progressive suffix, followed by the suffix *-o?*. Our initial hypothesis is that the underlying form should be identical to the surface form until we have evidence that phonological rules change the underlying forms in predictable ways. The difference between the predicted form and the actual surface realization of the verb is that the underlying form has a cluster of vowels which is not found in the surface form. The data do not provide any examples of surface vowel clusters, and this fact allows us to state a very simple rule accounting for the surface form: the first of two consecutive vowels is deleted.

#### (41) *Vowel cluster reduction*

$$V \rightarrow \emptyset / \_ V$$

Now we turn to the alternations in the shape of the stem that arise between the plain forms of the verb and the verb with an object prefix. Verbs with the prefix *ke-* behave exactly like verbs with the prefix *we-*. Disregarding the suffixes *-n-* and *-o?*, we arrive at the following surface variations in the shape of the stem.

(42) <i>Stem without prefix</i>	<i>Stem with CV prefix</i>	
picna	pcena	‘cut’
notxo	ntoxo	‘hoe’
netle	ntale	‘lick’
naxce	nxace	‘make a fire’
jamxa	jmaxa	‘paint a face’
nawle	nwele	‘spread’

In form without a prefix, there is a vowel between the first two consonants and none between the second and third consonants; in form with a CV prefix, there is no vowel between the first two consonants but there is a vowel between the second and third consonants. One alternative is that this vowel is epenthetic (inserted); the other is that the vowel is part of the underlying vowel of the stem and is deleted in some phonological context. Just as there is no way to predict what vowel will appear between the first and second consonants, it is also impossible to predict what vowel will appear between the

second and third consonants, and therefore the vowel cannot be epenthetic. In short, the underlying representation must contain unpredictable vowels after each consonant.

(43)	picena	'cut'
	notoxo	'hoe'
	netale	'lick'
	naxace	'make a fire'
	jamaxa	'paint a face'
	nawele	'spread'

The underlying forms of prefixed and unprefixed forms would thus be as follows (illustrating with the progressive form of the verb).

(44) <i>Unprefixed</i>	<i>Prefix</i>	<i>Prefix</i>
/picenano?/	/kepicenano?/	'cut'
/notoxono?/	/kenotoxono?/	'hoe'
/netaleno?/	/kenetaleno?/	'lick'
/naxaceno?/	/kenaxaceno?/	'make a fire'
/jamaxano?/	/kejamaxano?/	'paint a face'
/naweleno?/	/kenaweleno?/	'spread'

Compare this with the surface form of the verbs:

(45) <i>Unprefixed</i>	<i>Prefix</i>	<i>Prefix</i>
picnano?	kepcenano?	'cut'
notxono?	kentoxono?	'hoe'
netleno?	kentaleno?	'lick'
naxceno?	kenxaceno?	'make a fire'
jamxano?	kejmaxano?	'paint a face'
nawleno?	kenweleno?	'spread'

The relation between the underlying forms in (44) and surface forms in (45) is simple. Each is subject to a rule deleting the second vowel of the word.

$$(46) V \rightarrow \emptyset / \# CVC \underline{\phantom{C}}$$

Whether the first or second stem vowel is deleted depends on whether a prefix is present.

Apart from illustrating the point that underlying forms of words may not correspond to any single column in a word's paradigm, this discussion of Tonkawa illustrates two important characteristics of a phonological analysis. The first is that one analyzes data by advancing an initial hypothesis, and then refining the hypothesis when it becomes clear that the initial hypothesis doesn't work perfectly. Thus we began with the hypothesis that the underlying forms were /picna/, /notxo/, /netle/ and so on, and were able to account for

a certain amount of data based on that hypothesis, but later modified our hypothesis about underlying forms to be /picena/, /notxo/, /netale/ and so on. In other words, although our first hypothesis turned out to be wrong, it was close to right, and we were able to identify the source of the problem in our hypothesis and correct it.

The second characteristic of our analysis is that we always seek ways to test the predictions of our hypotheses. The hypothesis that the stems are underlying /picna/, /notxo/, /netle/, etc. makes a prediction that if a vowel were ever to appear between the second and third consonants (for example due to a rule of vowel insertion), it would always be a single consistent and predictable vowel (since we are saying that it is not in the underlying form). The fact that a *different* vowel appears in *wepceno?*, *wentoxo?*, *wentalo?* and *wenxaco?* shows that the prediction of this hypothesis is wrong, and this forced us to consider the alternative hypothesis that the underlying form contains a vowel between the second and third consonants: this hypothesis proved to be correct. The most basic form of hypothesis testing that is done in phonology is combining presumed forms of roots and affixes, and mechanically applying the rules which we assume in the analysis. If the wrong form is produced by this test, something is wrong with the hypothesis – either the underlying forms are wrong, or the rules are stated incorrectly (or the rules are being applied in the wrong order, a point we get to in the next chapter).

## Summary

Establishing the correct underlying representation for a morpheme is the most important first step in giving a phonological analysis of data. A correct underlying representation unifies surface variants of a morpheme, giving recognition of the basic “sameness” of a morpheme, regardless of variations in pronunciation which arise because phonological rules have applied. The underlying form and the system of rules are thus connected: by making the right choice of underlying form, and given the right system of rules, the rules will correctly operate on just those segments which participate in the alternation. The key to making the right decision about underlying forms is to carefully consider different hypotheses: if a segment in a morpheme has two or more surface realisations, it is often necessary to consider two or more possibilities for what is underlying – is variant [a], [b] or [c] the right choice? The main issue relevant to answering this question is knowing which variant preserves important distinctions and which neutralizes distinctions. The underlying form may not even be seen directly in any one pronunciation of a morpheme: it may be a form inferred from considering a number of specific instantiations of the morpheme.

## Exercises

### 1 Axininca Campa

Provide underlying representations and a phonological rule which will account for the following alternations:

toniro	‘palm’	notoniroti	‘my palm’
jaarato	‘black bee’	nojaaratoti	‘my black bee’

kanari	'wild turkey'	nojanariti	'my wild turkey'
kosiri	'white monkey'	nojosiriti	'my white monkey'
pisiro	'small toucan'	nowisiroti	'my small toucan'
porita	'small hen'	noworitati	'my small hen'

## 2. Xavante

What is the underlying form of the prefix meaning 'your'? Formalize the rule which accounts for the change in that prefix.

<i>N</i>	<i>your N</i>	
du	?addu	'stomach'
?ra	?aj?ra	'child'
hi?rāti	?ajhi?rāti	'knee'
tō	?attō	'eye'
?wa	?aj?wa	'tooth'
brō	?ajbrō	'wife'
ſe:re	?aſſe:re	'hair'
pa:ra	?ajpa:ra	'foot'
bā:bā	?ajbā:bā	'father'

## 3 Kuria I

What are the underlying forms of the prefixes marking the infinitive and the objects 'you', 'it (cl. 3)' (which refers to one of the noun classes in the language)? Give a rule that explains the non-underlying pronunciation of the prefix. Tone can be ignored.

'to V'	'to V you'	'to V it (cl. 3)'	gloss
okorééndá	ogokóreendá	okogóreendá	guard
ogosóóká	okogósooká	okogósooká	respect
ogoséékérá	okogóseekéra	okogóseekéra	close for
ogotááŋgátá	okogótaaangáta	okogótaaangáta	lead
ogotát'órá	okogótát'orá	okogótát'orá	tear
okohéétóká	ogokóheetóka	okogóheetóka	remember
ogokáraangéra	okogókáraáŋgéra	okogókáraáŋgéra	fry for
okobéémérá	ogokóbeeméra	okogóbeeméra	measure for
okogóótá	ogokógoötá	okogógoötá	hold
okományá	ogokómányä	okogómányä	know

## 4 Korean

Give the underlying representations of each of the verb stems found below; state what phonological rule applies to these data. (There is a vowel harmony rule explaining the variation between final *a* and *ə* in the imperative, which you do not need to be concerned with.)

<i>Imperative</i>	<i>Conjunctive</i>	
ipə	ipk'o	'wear'
kupə	kupk'o	'bend'
kap <sup>h</sup> a	kapk'o	'pay back'
t'ip <sup>h</sup> ə	t'ipk'o	'feel the pulse'
tata	tatk'o	'close'
put <sup>h</sup> ə	putk'o	'adhere'
məkə	məkk'o	'eat'
t'ukə	t'ukk'o	'die'
ikə	ikk'o	'ripen'
tak'a	takk'o	'polish'
k'ak'a	k'akk'o	'reduce expenses'
sək'ə	səkk'o	'mix'

## 5 Zoque

Provide the necessary phonological rules to describe the alternations found in the following data. Give the underlying representation for each noun stem and state what the underlying form of the morpheme meaning "my" is.

<i>N</i>	<i>my N</i>	<i>gloss</i>	<i>N</i>	<i>my N</i>	<i>gloss</i>
waka	nwaka	basket	huki	nhuki	cigarette
disko	ndisko	record	yomo	nyomo	woman
buru	mburu	burro	gayu	ŋgayu	rooster
pama	mbama	clothes	plato	mblato	plate
kaju	ŋgaju	horse	t'o?ŋgoja	nd <sup>3</sup> o?ŋgoja	rabbit
tatah	ndatah	father	t'ima	nd <sup>z</sup> ima	calabash
faha	faha	belt	sək	sək	beans
sapun	ſapun	soap	lawus	lawus	nail

## 6 Hungarian

Explain what phonological process affects consonants in the following data (a vowel harmony rule makes suffix vowels back after back vowels and front after front vowels, which you do not need to account for). State what the underlying forms are for all morphemes.

<i>Noun</i>	<i>In N</i>	<i>From N</i>	<i>To N</i>	
kalap	kalabban	kalapto:l	kalapnak	'hat'
ku:t	ku:dban	ku:tto:l	ku:tnak	'well'

ʒa:k	ʒa:gban	ʒa:kto:l	ʒa:knak	‘sack’
re:s	re:zben	re:stø:l	re:snek	‘part’
ʃro:f	ʃro:vban	ʃro:fto:l	ʃro:fnak	‘screw’
laka:ʃ	laka:ʒban	laka:sto:l	laka:ʃnak	‘apartment’
ketret <sup>s</sup>	ketred <sup>z</sup> ben	ketret <sup>s</sup> tø:l	ketret <sup>s</sup> nek	‘cage’
test	tezdben	testtø:l	testnek	‘body’
rab	rabban	rapto:l	rabnak	‘prisoner’
ka:d	ka:dban	ka:tto:l	ka:dnak	‘tub’
meleg	melegben	melektø:l	melegnek	‘warm’
vi:z	vi:zben	vi:stø:l	vi:znek	‘water’
vara:ʒ	vara:ʒban	vara:ʃto:l	vara:ʒnak	‘magic’
a:g <sup>j</sup>	a:g <sup>j</sup> ban	a:k <sup>j</sup> to:l	a:g <sup>j</sup> nak	‘bed’
sem	semcen	semtø:l	semnek	‘eye’
byn	bynben	byntø:l	bynnek	‘crime’
toron	toronjban	toronto:l	toronnak	‘tower’
fal	falban	falto:l	falnak	‘wall’
ø:r	ø:rben	ø:rtø:l	ø:rnek	‘guard’
sa:j	sa:jban	sa:jto:l	sa:jnak	‘mouth’

## 7 Kuria II

Provide appropriate underlying representations and phonological rules which will account for the following data:

<i>Verb</i>	<i>Verb for</i>	
suraŋga	suraŋgera	‘praise’
taaŋgata	taaŋgatera	‘lead’
baamba	baambera	‘fit a drum head’
reenda	reendera	‘guard’
rema	remera	‘cultivate’
hoora	hoorera	‘thresh’
roma	romera	‘bite’
sooka	sookera	‘respect’
tat'ora	tat'orera	‘tear’
siika	seekera	‘close’
tiga	tegera	‘leave behind’
ruga	rogera	‘cook’

suka	sokera	'plait'
huuta	hootera	'blow'
riiŋga	reeŋgera	'fold'
siinda	seendera	'win'

**8: Isthmus Zapotec.** Provide an analysis of the following alternations, formalizing your rules using features and listing the underlying forms of all morphemes.

pan	ʃpanbe	'bread'
taburete	ʃtaburetebe	'chair'
tanguju	ʃtangujube	'clay doll'
kuba	ʃkubabe	'corn dough'
luuna?	hluuunabe	'bed'
neza	hnezabe	'road'
mani?	hmanibe	'horse'
diid <sup>3</sup> a?	ʃtiid <sup>3</sup> abe	'word'
daa	ʃtaabe	'mat'
doo	ʃtoobe	'rope'
geta	ʃketabe	'tortilla'
gjna	ʃkjnabe	'trunk'
gamiža?	ʃkamizabe	'shirt'
bere	ʃperebe	'chicken'
biuuza?	ʃpiuuzabe	'guest'
ba?du?	ʃpa?dube	'child'

## 9. Lezgian

Give underlying forms of all morphemes and formalize the rules which account for the following data:

<i>sg</i>	<i>pl</i>	<i>gloss</i>	<i>sg</i>	<i>pl</i>	<i>gloss</i>
tar	tarar	tree	pel	peler	forehead
q'el	q'eler	salt	qaw	qawar	roof
beden	bedenar	body	penzer	penzerar	window
ʂyl	ʂyler	husband	luw	luwar	wind
lif	lifer	pigeon	t <sup>h</sup> ur	t <sup>h</sup> urar	sword
qyl	qyler	wheat	bulut <sup>h</sup>	bulut <sup>h</sup> ar	cloud
k' <sup>w</sup> at <sup>h</sup>	k' <sup>w</sup> at <sup>h</sup> ar	foot	tark <sup>wh</sup>	tark <sup>wh</sup> ar	pestle
mark <sup>wh</sup>	mark <sup>wh</sup> ar	stack	rang	rangar	color
fend	fender	trick	k'alub	k'alubar	form
k'arab	k'arabar	bone	p <sup>h</sup> elenj	p <sup>h</sup> elenjar	tiger
qejd	qejder	remark	rab	rapar	needle
rib	riper	awl	t <sup>h</sup> eb	t <sup>h</sup> eper	clay

mug	mukar	nest	t <sup>s</sup> eg <sup>w</sup>	t <sup>s</sup> ek <sup>w</sup> er	ant
meg	meker	hair	jad	jatar	water
q <sup>w</sup> ed	q <sup>w</sup> eter	partridge	kard	kartar	falcon
nek <sup>h</sup>	nek'er	milk	wirt <sup>h</sup>	wirt'er	honey
met <sup>h</sup>	met'er	knee	net <sup>h</sup>	net'er	louse
wik <sup>h</sup>	wik'er	yoke	wak <sup>h</sup>	wak'ar	pig
haq <sup>h</sup>	haq'ar	truth	req <sup>h</sup>	req'er	way

## 10 Estonian

Give the underlying representations of the following nouns, and formalize the phonological rule which applies in the following data:

<i>nom. sg</i>	<i>gen. pl.</i>	<i>gloss</i>
kubu	kubude	bundle
tiba	tibade	drop
sadu	sadude	falling
ləbu	ləbude	fun
abi	abide	help
soga	sogade	mud
nibu	nibude	point
kivi	kivide	stone
lugu	lugude	story
saba	sabade	tail
sula	sulade	thawed
təde	tədede	truth
kude	kudede	web
viht	vihtade	bath
vakk	vakkade	birch bark box
arg	argade	cowardly
vəlg	vəlgade	debt
søöt	søötade	fodder
mets	metsade	forest
laisk	laiskade	lazy
hoob	hoobade	lever
haav	haavade	wound
lill	lillede	flower
kərb	kərbede	forest
laht	lahtede	gulf
neem	neemede	headland
leep	leepede	sloping
luik	luikede	swan
jæl <sup>ŋ</sup> g	jæl <sup>ŋ</sup> gede	trace
kaart	kaartide	card

kər <sup>j</sup> b	kər <sup>j</sup> bide	copperish red
paar	paaride	pair
kook	kookide	pothook
kepp	keppide	stick
pyks	pykside	trousers
lind	lindude	bird
koon	koonude	chin
kark	karkude	crutch
hull	hullude	lunatic
arv	arvude	numeral
orb	orbude	orphan
roog	roogude	reed

## 11 Tibetan

Numbers between 11 and 19 are formed by placing the appropriate digit after the number 10, and multiples of 10 are formed by placing the appropriate multiplier before the number 10. What are the underlying forms of the basic numerals, and what phonological rule is involved in accounting for these data?

d <sup>3</sup> u	'10'	d <sup>3</sup> ig	'1'	d <sup>3</sup> ugd <sup>3</sup> ig	'11'
ʃi	'4'	d <sup>3</sup> ubʃi	'14'	ʃibd <sup>3</sup> u	'40'
gu	'9'	d <sup>3</sup> urgu	'19'	gubd <sup>3</sup> u	'90'
ŋa	'5'	d <sup>3</sup> uŋa	'15'	ŋabd <sup>3</sup> u	'50'

## 12 North Saami

Posit appropriate underlying forms and any rules needed to explain the following alternations. The emphasis here should be on correctly identifying the underlying form: the exact nature of the changes seen here is a more advanced problem.

Nominative sg.	Essive	
varit	varihiń	'2-year-old reindeer buck'
oahpis	oahpisin	'acquaintance'
t'óarvvuſ	t'óarvvuſin	'antlers and skullcap'
lottæ:ʃ	lottæ:d <sup>3</sup> in	'small bird'
t'uojvvat	t'uojvvagin	'yellow-brown reindeer'
ahhkut	ahhkubin	'grandchild of woman'
suohkat	suohkaðin	'thick'
he:d <sup>3</sup> oʃ	he:d <sup>3</sup> od <sup>3</sup> in	'poor guy'
æ:dd <sup>3</sup> ut	æ:dd <sup>3</sup> ubin	'grandchild of man'
bissobeah̚t <sup>s</sup> et	bissobeah̚tehin	'butt of gun'
t'eaht <sup>s</sup> it	t'eaht <sup>s</sup> ibin	'children of elder brother of man'

jæ:?min	jæ:?mimin	‘death’
læ:geʃ	læ:ged <sup>3</sup> in	‘mountain birch’
gahpir	gahpirin	‘cap’
gæ:wht <sup>s</sup> is	gæ:wht <sup>s</sup> isin	‘8 people’
æ:slat	æ:slagin	[man’s name]
baðo:jgæ:tt <sup>s</sup> et	baðo:jgæ:tt <sup>s</sup> ebin	‘tailed tit’
ahhkit	ahhkiðin	‘boring’
bahæ:næ:lat	bahæ:næ:lagin	‘badly behaved’
beſſtor	beſſtorin	‘bird type’
he:veměahhtun	he:veměahhtunin	‘inappropriate’
bejo:t	bejo:hin	‘white reindeer’
bissomeahtun	bissomeahtumin	‘unstable’
laðæ:s	laðæ:sin	‘something jointed’
heiatusmielat	heiatusmielagin	‘unhappy’
heanjkkan	heanjkkanin	‘hanger’
jamæ:n	jamæ:nin	‘something that makes noise’

### 13 Logoori (online)

The recordings at <https://languagedescriptions.github.io/IP3/Ch5.html> provide data for you to transcribe and analyze phonologically.

### Further reading

Inkelas 1994; Kaisse and Shaw 1985; Kenstowicz and Kisseeberth 1977; Stanley 1967.