5

The Comparative Method and Linguistic Reconstruction

> Linguistic history is basically the darkest of the dark arts, the only means to conjure up the ghosts of vanished centuries. With linguistic history we reach furthest back into the mystery: humankind.
>
> (Cola Minis 1952: 107 [Euphorion 46])

5.1 Introduction

The comparative method is central to historical linguistics, the most important of the various methods and techniques we use to recover linguistic history. In this chapter the comparative method is explained, its basic assumptions and its limitations are considered, and its various uses are demonstrated. The primary emphasis is on learning how to apply the method, that is, on how to reconstruct. The comparative method is also important in language classification, in linguistic prehistory, in research on distant genetic relationships, and in other areas; these topics are treated in later chapters.

We say that languages which belong to the same language family are *genetically related* to one another: this means that these related languages derive from (that is, 'descend' from) a single original language, called a *proto-language*. In time, dialects of the proto-language develop through linguistic changes in different regions where the language was spoken – all languages (and varieties of language) are constantly changing – and then later through further changes the dialects become distinct languages.

The aim of reconstruction by the comparative method is to recover as much as possible of the ancestor language (the proto-language) from a comparison of the related languages, the descendants of the original language, and to determine what changes have taken place in the various languages that developed from the proto-language. The work of reconstruction usually begins with phonology, with an attempt to reconstruct the sound system; this leads in turn to reconstruction of the vocabulary and grammar of the proto-language. As can be seen from the way languages are classified, we speak of linguistic relationships in terms of kinship; we talk about 'sister languages', 'daughter languages', 'parent language' and 'language families'. If reconstruction is successful, it shows that the assumption

that the languages are related is warranted. (See Chapter 6 for family-tree classification and Chapter 13 for methods of determining whether languages are related.)

With the genealogical analogy of your family tree in mind, we can see how modern Romance languages have descended from spoken Latin (better said, from Proto-Romance, which is reconstructed via the comparative method), illustrated in the family tree for the Romance languages in Figure 5.1. (The biological kinship terms added here under the language names in Figure 5.1 are just a trick to reveal the pedigree of the languages; in this case the focus is on Spanish. This is certainly not conventionally done in linguistic family trees.)

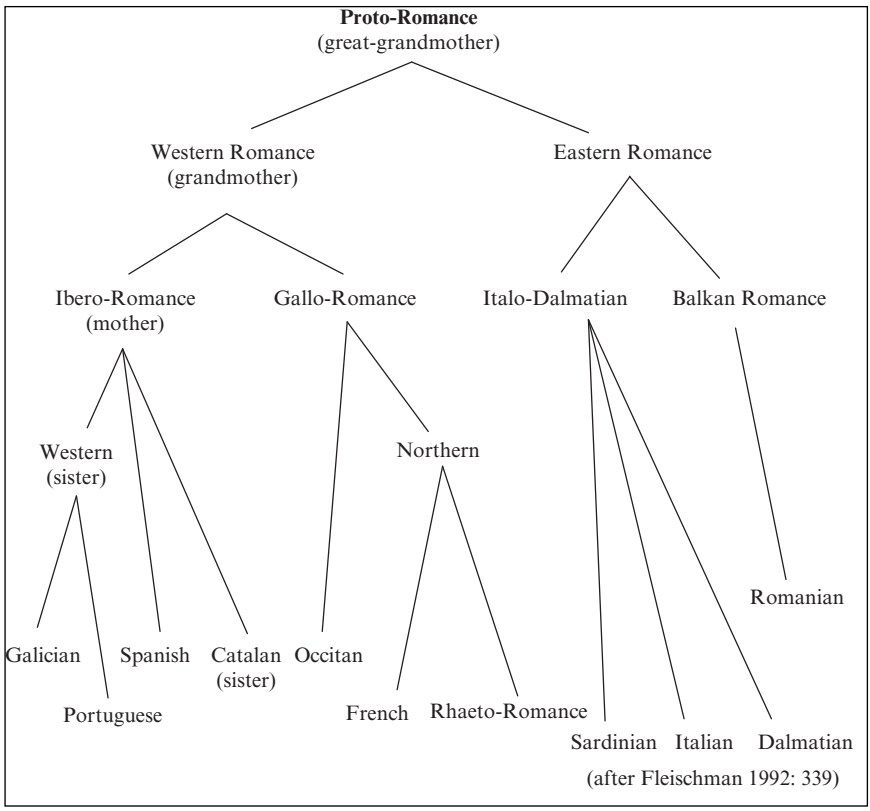

FIGURE 5.1: Proto-Romance family tree (and genealogy of Spanish)

By comparing what these sister languages inherited from their ancestor, we attempt to reconstruct the linguistic traits which Proto-Romance possessed. (Proto-Romance is equivalent to the spoken language at the time when Latin began to diversify and split up into its descendant branches, essentially the same as Vulgar Latin at the time. The 'Vulgar' of Vulgar Latin means 'of the people'.) If we are successful, what we reconstruct for Proto-Romance by the comparative

method should be similar to the Proto-Romance which was actually spoken at the time before it split up into its daughter languages. Of course, our success is dependent upon the extent to which evidence of the original traits is preserved in the descendant languages (daughter languages) which we compare and upon how astute we are at applying the techniques of the comparative method, among other things. In this case, since Latin is abundantly documented, we can check to see whether what we reconstruct by the comparative method accurately approximates the spoken Latin we know about from written sources. However, the possibility of checking our reconstructions in this way is not available for most language families, for whose proto-languages we have no written records. For example, for Proto-Germanic (from which English descends), there are no written attestations at all, and the language is known only from comparative reconstruction.

Currently existing languages which have relatives all have a history which classifies them into language families. By applying the comparative method to related languages, we can postulate what that common earlier ancestor was like – we can reconstruct that language. Thus, comparing English with its relatives, Dutch, Frisian, German, Danish, Swedish, Icelandic and so on, we attempt to understand what the proto-language, in this case called 'Proto-Germanic', was like. Thus, English is, in effect, a much-changed 'dialect' of Proto-Germanic, having undergone successive linguistic changes to make it what it is today, a different language from Swedish and German and its other sisters, which underwent different changes of their own. Therefore, every proto-language was once a real language, regardless of whether we are successful at reconstructing it or not.

5.2 The Comparative Method Up Close and Personal

To illustrate the application of the comparative method, let's begin by applying it briefly in a simplified fashion to some Romance languages. (There are many more Romance languages, but for illustration's sake, this miniature introduction is limited to just a few of the better-known of these.) First, consider some data, the words compared among Romance languages given in Table 5.1. (The first line represents conventional spelling; the second is phonemic.)

Latin is *not* a Romance language; the Latin forms in Table 5.1 are presented only so that ultimately we can check the reconstructions which we postulate for Proto-Romance to see how close they come to the forms in the actual spoken proto-language, which was essentially the same as Latin in this case.

To understand the comparative method and to be able to apply it, we need to control some concepts and technical terms:

Proto-language: (1) the once spoken ancestral language from which aughter languages descend; (2) the language reconstructed by the comparative method which represents the ancestral language from which the compared languages descend. (To the extent that the reconstruction by the comparative method is accurate and complete, (1) and (2) should coincide.)

TABLE 5.1: Some Romance cognate sets

	Italian	*Spanish*	*Portuguese*	*French*	*(Latin)*	*English gloss*
1.	capra /kapra/	cabra /kabra/	cabra /kabra/	chèvre /ʃɛvr(ə)/	capra	'goat'
2.	caro /karo/	caro /karo/	caro /karu/	cher /ʃɛr/	caru	'dear'
3.	capo /kapo/ 'main, chief'	cabo /kabo/ 'extremity'	cabo /kabu/ 'extremity'	chef /ʃɛf/ 'main, chief'	caput	'head, top'
4.	carne /karne/	carne /karne/	carne /karne/	chair /ʃɛr/ (cf. Old French charn /čarn/)	carō/carn-	'meat, flesh'
5.	cane /kane/	can (archaic) /kan/	cão /kãw̃/	chien /ʃjɛ̃/	canis	'dog'

Sister language: languages which are related to one another by virtue of having descended from the same common ancestor (proto-language) are sisters; that is, languages which belong to the same family are sisters to one another.

Cognate: a word (or morpheme) which is related to a word (morpheme)in sister languages by reason of these forms having been inherited by these sister languages from a common word (morpheme) of the proto-language from which the sister languages descend.

Cognate set: the set of words (morphemes) which are related to one another across the sister languages because they are inherited and descend from a single word (morpheme) of the proto-language.

Comparative method: a method (or set of procedures) which compares forms from related languages, *cognates*, which have descended from a common ancestral language (the *proto-language*), in order to postulate, that is *to reconstruct*, the form in the ancestral language.

Sound correspondence (also called *correspondence set*): in effect, a set of 'cognate' sounds; the sounds found in the related words of cognate sets which correspond from one related language to the next because they descend from a common ancestral sound. (A sound correspondence is assumed to recur in various cognate sets.)

Reflex: the descendant in a daughter language of a sound of the proto- language is said to be a *reflex* of that original sound; the original sound of the proto-language is said to be reflected by the sound which descends from it in a daughter language.

For ease of description, we will talk about 'steps' in the application of the comparative method. Strictly speaking though, it is not always necessary to follow all these steps in precisely the sequence described here. In practice, the comparative linguist typically jumps back and forth among these steps.

Step 1: Assemble cognates

To begin to apply the comparative method, we look for potential cognates among related languages (or among languages for which there is reason to suspect relatedness) and list them in some orderly arrangement (in rows or columns). In Table 5.1, this step has already been done for you for the few Romance cognates considered in this exercise. In general, it is convenient to begin with cognates from 'basic vocabulary' (body parts, close kinship terms, low numbers, common geographical terms), since these resist borrowing more than other sorts of vocabulary, and for the comparative method we want to compare only true cognates, words which are related in the daughter languages by virtue of being inherited from the proto-language. For successful reconstruction, we must eliminate all other sets of similar words which are not due to inheritance from a common ancestor, such as those which exhibit similarities among the languages because of borrowing, chance (coincidence) and so on (for details, see Chapter 13). Ultimately, it is the systematic correspondences which we discover in the comparative method (in the following steps) which demonstrate true cognates.

Step 2: Establish sound correspondences

Next, we attempt to determine the sound correspondences. For example, in the words for 'goat' in cognate set 1 in Table 5.1, the first sound in each language corresponds in the way as indicated in SOUND CORRESPONDENCE 1 (here now we concentrate on the phonemic representation of the sound and not on the conventional spelling):

Sound correspondence 1:
Italian *k-* : Spanish *k-* : Portuguese *k-* : French *ʃ-*

Note that historical linguists often use the convention of a hyphen after a sound to indicate initial position, as *k-* here signals initial *k*; a preceding hyphen indicates that the sound is word-final (for example, *-k*); and a hyphen both before and after refers to a medial sound, one found somewhere in the middle of a word but neither initially nor finally (for example, *-k-*).

It is important to attempt to avoid potential sound correspondences which are due merely to chance. For example, languages may have words which are similar only by accident, by sheer coincidence, as the case of Kaqchikel (Mayan) *mes* 'mess, disorder, garbage' : English *mess* ('disorder, untidiness'). To determine whether a sound correspondence such as that of SOUND CORRESPONDENCE 1 is real (reflecting sounds inherited in words from the proto-language) rather than perhaps just an accidental similarity, we need to determine whether the correspondence recurs in other cognate sets. In looking for further examples of this particular Romance sound correspondence, we find that it recurs in the other cognate sets (2–5) of Table 5.1, all of which illustrate SOUND CORRESPONDENCE 1 for their first sound. If we were to attempt to find recurrences of the seeming *m-* : *m-* correspondence between Kaqchikel and English (seen in the comparison of their words meaning 'mess'), we would soon discover that there are no other instances of it, that it does not recur, as illustrated by the compared words of

TABLE 5.2: Kaqchikel–English comparisons

English	*Kaqchikel*
man	ačі
mouse	č'oy
moon	qatiʔt
mother	nan

Table 5.2, where the English forms begin with *m*, but the Kaqchikel forms begin with various sounds.

Of course, in principle in a situation such as this, it is possible that the compared languages could be related but that we accidentally chose the few words to compare in Table 5.2 where one or the other of the related languages has not retained the cognate due to borrowing or lexical replacement. To be certain that this is not the case, we would need to look at many comparisons (not just the handful presented in Table 5.2 for illustration's sake). However, in the case of English and Kaqchikel lexical comparisons, we will never find more than one or two which exhibit what initially might have been suspected of being an *m-* : *m-*correspondence based on the words meaning 'mess' in the two languages, and this is precisely because these two languages are not genetically related and therefore the *m* : *m* matching does not recur and is not a true correspondence. Similarly, we need to attempt to eliminate similarities found in borrowings which can seem to suggest sound correspondences. Usually (though not always), loanwords do not exhibit the sort of systematic sound correspondences found in the comparison of native words among related languages, and loans involving basic vocabulary are much rarer than borrowings in other kinds of vocabulary (see Chapter 13 for details).

Given that SOUND CORRESPONDENCE 1 recurs frequently among the Romance languages, as seen in the forms compared in Table 5.1, we assume that this sound correspondence is genuine. It is highly unlikely that a set of systematically corresponding sounds such as this one could come about by sheer accident in a large number of words so similar in sound and meaning across these languages.

Step 3: Reconstruct the proto-sound

There is no fixed rule about what should be done next. We could go on and set up other sound correspondence sets and check to see that they recur; that is, we could repeat step 2 over and over until we have found all the sound correspondences in the languages being compared. Or, we could go on to step 3 and attempt to reconstruct the proto-sound from which the sound in each of the daughter languages (represented in SOUND CORRESPONDENCE 1) descended. In the end, to complete the task, we must establish all the correspondences and reconstruct the proto-sound from which each descends, regardless of whether we do all of step 2 for each set first and then step 3 for all the sets, or whether we do step 2 followed by step 3 for each set and then move on to the next set, repeating step 2, then step 3.

In either case, as we shall soon see, the initial reconstructions which we postulate based on these sound correspondences must be assessed in steps 5 and 6, when we check the fit of the individual reconstructed sounds which we initially postulate in step 3 against the overall phonological inventory of the proto-language and its general typological fit; it is often the case that some of the reconstructions for sounds postulated in step 3 need to be modified in steps 5 and 6.

The different sounds (one for each language compared) in the sound correspondence set reflect a single sound of the proto-language which is inherited in the different daughter languages; sometimes the sound is reflected unchanged in some daughters, though often it will have undergone sound changes in some (or even all) of the daughter languages which make it different from the original proto-sound. We reconstruct the proto-sound by postulating what the sound in the proto-language most likely was on the basis of the phonetic properties of the descendant sounds in the various languages in the correspondence set. The following are the general guidelines that linguists rely on to help them in the task of devising the best, most realistic reconstruction.

Directionality

The known directionality of certain sound changes is a valuable clue to reconstruction (see Chapter 2). By 'directionality' we mean that some sound changes which recur in independent languages typically go in one direction ($A > B$) but usually are not (sometimes are never) found in the other direction ($B > A$). Some speak of this as 'naturalness', some changes 'naturally' taking place with greater ease and frequency cross-linguistically than others. For example, many languages have changed $s > h$, but change in the other direction, $h > s$, is almost unknown. In cases such as this, we speak of 'directionality'. If we find in two sister languages the sound correspondence *s* in Language$_1$: *h* in Language$_2$, we reconstruct **s* and postulate that in Language$_2$ **s* > *h*. The alternative with **h* and the change **h* > *s* in Language$_1$ is highly unlikely, since it goes against the known direction of change. Usually, the directionality has some phonetic motivation. Some idea of the typical direction of many of the more commonly recurring sound changes can be gathered from a look at the examples considered in Chapter 2.

In the case of SOUND CORRESPONDENCE 1, we know that the direction of change from *k* to *ʃ* is quite plausible and has been observed to occur in other languages, but that *ʃ* essentially never changes to *k*. Actually, even more typical would be for *k* to change to *ʃ* by first going through the intermediate stage of *č*, that is, *k* > *č* > *ʃ*; documentary evidence shows that the sound change in French did go through this intermediate *č* stage. Old French documents had for the words in Table 5.1: *čjɛvr(ə)* 'goat', *čjɛr* 'dear', *čjɛf* 'head', *čarn* 'meat' and *čjɛŋ* 'dog'. This intermediate stage is preserved in many English loans from French from that time, for example, *chief* and *Charles* with [č], where more recent loans from the same French sources have [ʃ], the result of the later French change of *č* > *ʃ*, as in *chef* and *Charlene*, with [ʃ].

In another example of the way in which directionality aids in reconstruction, we know that very often voiceless stops (*p*, *t*, *k*) are voiced (*b*, *d*, *g*) between vowels. If we compare two related languages, Language$_1$ and Language$_2$, and we find intervocalic *-b-* in Language$_1$ corresponding to intervocalic *-p-* in Language$_2$,

then we reconstruct **-p-* and assume that Language$_2$ underwent the common sound change of intervocalic voicing of stops ($p > b$ /V__V, in this case). If we tried to reconstruct **-b-* in this situation, we would have to assume that Language$_2$ had changed *-b-* to *-p-*, but this goes against the direction most commonly taken in changes involving these sounds between vowels. This example comes up in SOUND CORRESPONDENCE 2 (below).

The phonetic motivation for the directionality in this case is clear. It is easy to voice stops between vowels, since vowels are inherently voiced, and therefore the change (1) $p > b$ /V__V is very common, while it is not so easy to make stops voiceless between vowels, which makes the change (2) $b > p$ /V__V very rare indeed – for (2) the vocal cords would be vibrating for the first vowel, then we would need to stop them from vibrating in order to produce the voiceless [p], and then start the vocal-cord vibration up again for the second vowel; for (1) we merely leave them vibrating for all three segments, the two vowels and the intervening [b]. The known directionality, then, with (1) encountered frequently across languages and (2) hardly at all, is natural and phonetically motivated. As a beginning linguist's experience with language changes and phonological systems increases, a stronger understanding of the directionality of changes develops.

Majority wins

Another guiding principle is that, all else being equal, we let the majority win – that is, unless there is evidence to the contrary, we tend to pick for our reconstructed proto-sound the particular sound in the correspondence set which shows up in the greatest number of daughter languages. Since in SOUND CORRESPONDENCE 1, Italian, Spanish and Portuguese all have *k*, and only French diverges from this, with *ʃ*, we would postulate **k* for the Proto-Romance sound, under the assumption that the majority wins, since the majority of the languages have *k* in this correspondence set. This reconstruction assumes that French underwent the sound change **k* > *ʃ*, but that the other languages did not change at all, **k* remaining *k*. The underlying rationale for following the majority-wins principle is that it is more likely that one language would have undergone a sound change (in this case, French **k* > *ʃ*) than that several languages would independently have undergone the sound change. In this case, if **ʃ* were postulated as the proto-sound, it would be necessary to assume that Italian, Spanish and Portuguese had each independently undergone the change of **ʃ* > *k*.

Caution is necessary, however, in the use of the majority-wins guideline to reconstruction. Some sound changes are so common (and languages undergo them so easily) that several languages might undergo one of these kind of changes independently of one another (for example, loss of vowel length, nasalization of vowels before nasal consonants, and so on). It is also possible that only one of the daughter languages might have preserved the original sound unchanged while the others all changed it in some way. It is also possible that all the daughter languages may undergo various changes so that none reflects the proto-sound unchanged. Clearly, in these situations there is no majority to do the winning. Moreover, majority rule may not work if some of the languages are more closely related to one another. If some of the languages belong to the same branch (subgroup) of the family (see Chapter 6), then they have a more

immediate ancestor which itself is a daughter of the proto-language. This intermediate language (a parent of its immediate descendants but itself a daughter of the proto-language) could have undergone a change and then later split up into its daughters, the members of the subgroup, and each of these would then inherit the changed sound that their immediate common ancestor (once a single daughter of the proto-language) had undergone. For example, French, Spanish and Portuguese all share some sounds which are the results of sound changes that took place in Western Romance before it split up further into French, Spanish and Portuguese. Italian does not share these because it comes from a separate branch of Romance. For example, Western Romance changed syllable-final *k* to *i̯*, seen in Spanish, Portuguese and French, which separated from one another only after this Western Romance change had taken place, as in **lakte* > *lai̯te* 'milk', which gives us French *lait*, Portuguese *leite* and Spanish *leche* (where later changes were *ai* > *ei* > *e* in these languages, and *i̯t* > *č* in Spanish); Italian (not a Western Romance language) underwent a different change, *kt* > *tt*, giving *latte* 'milk' – we see the results of these changes in choices of kinds of coffee on menus, with *cafe au lait* (French), *cafe latte* (Italian) and *cafe con leche* (Spanish). Now if we compare Italian *tt* with the *i̯t* of Portuguese, French and formerly also of Spanish, 'majority wins' would seem to suggest **i̯t* as the reconstruction with *i̯* > *t* /__*t* in Italian; but knowing that Portuguese, Spanish and French are closely related, all members of the Western Romance branch, we no longer need to compare three separate instances of *i̯t* to one of *tt*, but only one *i̯t* case (the result of the single change,**kt* > *i̯t*, in Western Romance) to one *tt* case (in Italian). It is only with the aid of other information that we discover that the best reconstruction is **kt*, from which both the Italian and Western Romance languages departed due to their separate sound changes. As will be seen in Chapter 6, it is the results of the comparative method which provide the basis for arriving at the classification which tells us which of the related languages belong to the same branches of the family.

So, 'majority wins' is an important principle, but it is easily overridden by other considerations. Still, it would seem to work in the case of SOUND CORRESPONDENCE 1 above, suggesting **k* as the best reconstruction, since it is found in a majority of the languages compared.

Factoring in features held in common

We attempt to reconstruct the proto-sound with as much phonetic precision as possible; that is, we want our reconstruction to be as close as possible to the actual phonetic form of the sound as it was pronounced when the proto-language was spoken. We can never know for sure how accurately our reconstructed sound matches the actual sound of the formerly spoken proto-language, but in general, the more information available upon which to base the reconstruction, the more likely it is that we may be able to achieve a reasonably accurate reconstruction. We attempt to achieve as much phonetic realism as possible by observing what phonetic features are shared among the reflexes seen in each of the daughter languages in the sound correspondence. We determine which phonetic features are common to the reflexes in the daughter languages (and features which can be derived from others by the known direction of sound changes, in Step 2), and then we attempt to reconstruct the proto-sound by building into it these shared

phonetic features. To illustrate this, let us consider another sound correspondence from Table 5.1, seen to recur here in the words for (1) 'goat' and (2) 'head' (and in many other cognates not given in Table 5.1):

Sound correspondence 2:
Spanish *b* : Portuguese *b* : French *v* : Italian *p*

The reflexes in all four languages share the feature 'labial'; the Spanish, Portuguese and Italian reflexes share the feature 'stop' (phonemically). Factoring the features together, we would expect the proto-sound to have been a 'labial stop' of some sort, a *p* or *b*. Given that the reflex in Spanish, Portuguese and French is 'voiced', under the principle of 'majority wins' we might expect to reconstruct a 'voiced bilabial stop' (**b*). In this case, however, other considerations – especially directionality – override the majority-wins principle. The directionality is that it is easy for *p* to become voiced between voiced sounds (between vowels in cognate set 3, and between a vowel and *r* in cognate set 1 in Table 5.1), but the reverse is very rare. Therefore, by directionality, **p* is a better choice for the reconstruction, phonetically more plausible; Italian maintained *p* while the others underwent the change to voicing (**p* > *b* in Spanish and Portuguese; **p* > *v* in French, actually **p* > *b* > *v*). From directionality, we also know that stops frequently become fricatives between vowels (or between continuant sounds), but that fricatives rarely ever become stops in this environment. Thus, it is very likely that the French reflex *v* is the result of this sort of change. Taking these considerations into account, for correspondence set 2, we reconstruct **p* and postulate that in Spanish and Portuguese **p* > *b*, and French **p* > *v* (or **p* > *b* > *v*). SOUND CORRESPONDENCE 2, then, illustrates how the comparative linguist must balance the various rules of thumb for reconstruction, majority wins, directionality, and factoring in the features shared among the reflexes. (Ultimately, we find out that Western Romance underwent the change of **p* > *b* in this position, and then after Western Romance split up, the change of *b* > *v* in French took place. That is, taking the degree of relatedness (the subgrouping; see Chapter 6) into account, there is no longer a majority with the reflex *b*, but rather only Western Romance *b* as opposed to Italian *p*.)

Economy

What is meant by the criterion of economy is that when multiple alternatives are available, the one which requires the fewest independent changes is most likely to be right. For example, if for SOUND CORRESPONDENCE 1 we were to postulate **ʃ*, this would necessitate three independent changes from **ʃ* > *k*, one each for Italian, Spanish and Portuguese; however, if we postulate **k* for the Proto-Romance sound, we need assume only one sound change, **k* > *ʃ* in French. The criterion of economy rests on the assumption that the odds are greater that a single change took place than that three independent changes took place. Of course, sometimes independent changes do take place, so that the criterion does not always guarantee correct results; but all else being equal, the chances of a reconstruction which embodies more economical assumptions being correct are greater than for a reconstruction which assumes less economical developments. (See below for other examples of the use of the economy criterion.)

The other two general considerations (rules of thumb) which linguists use in

reconstructing sounds involve checking to see whether the individual sounds postulated to represent the various sound correspondences fit the overall phonological pattern of the proto-language and to see whether this reconstructed pattern is consistent with linguistic universals and typological expectations. These are *phonological fit* and *typological fit* respectively (steps 5 and 6, below). These two considerations come into play mostly after the full set of sound correspondences has been dealt with and the overall inventory of reconstructed sounds that are being postulated can be considered. For this reason, let's deal first with the other correspondences of Table 5.1, and then come back to these two considerations later.

Let us continue steps 2 and 3, then, for the forms in Table 5.1, and establish the remaining sound correspondences illustrated in these forms and set up reconstructions for them. It does not matter in which order we investigate the sound correspondences. We could first look only at initial consonants for all of the cognate sets, then medial consonants, then final consonants, and finally the various vowels; or, we could proceed by investigating the sound correspondence representing the next sound (the second) in the first cognate set, then go on to the third sound in that set, and so on until all the sounds of that cognate set have been addressed, and then proceed to the next cognate set, dealing with each of the sound correspondences for each of the sounds found in that set in sequence (though some of these may recur in other cognate sets and thus may already have been established in the consideration of the previous cognate sets already dealt with). We continue in this way until all the recurring sound correspondences have been examined and proto-sounds to represent them have been postulated. In this way, we will eventually come to reconstruct the full inventory of sounds in the proto-language.

In the example in Table 5.1, let us continue with the corresponding sounds in cognate set 1, for 'goat'. The first vowel in the forms in cognate set 1 shows SOUND CORRESPONDENCE 3:

Sound correspondence 3:
Italian *a* : Spanish *a* : Portuguese *a* : French *ɛ*.

We check this to see if it recurs, and we see that it is also found in the other cognate sets of Table 5.1, for 'dear', 'head' and 'meat'. (It is also found again, in effect, in the last vowel of cognate set 1 for 'goat', though we must deal with the later change in French of final *ɛ* to *ə*/*Ø*.) Under the majority-wins principle, for this sound correspondence we reconstruct **a* for the Proto-Romance sound, assuming that French has undergone the sound change **a* > *ɛ*.

The third sound in cognate set 1 'goat' has, in fact, already been dealt with in SOUND CORRESPONDENCE 2 (where we reconstructed **p* for the correspondence set Spanish *b* : Portuguese *b* : French *v* : Italian *p*).

The next sound in the sequence of sounds in the 'goat' cognates gives correspondence set 4:

Sound correspondence 4:
Italian *r* : Spanish *r* : Portuguese *r* : French *r*

SOUND CORRESPONDENCE 4 also recurs, in 'goat', 'dear' and 'meat' (in Table 5.1). For it, we would postulate Proto-Romance **r*, under 'majority wins', since all the languages have this reflex. (To be absolutely accurate, we would have to

deal with the fact that in Standard French the *r* became a uvular, but for now we ignore this detail.)

The last sound in 'goat' in effect repeats SOUND CORRESPONDENCE 3, although French later changed final *ɛ* further (to *ə* or *Ø*). Though technically this must be considered a separate sound correspondence, to make it easier we will just assume here that we would easily discover that the two correspondence sets, for the first and last vowel in the 'goat' cognate set, belong together due to a later conditioned change in French.

To complete the task, we would need to establish the sound correspondences for all the cognate sets and reconstruct sounds to represent them. For example, we would find:

Sound correspondence 5:
Italian *o* : Spanish *o* : Portuguese *u* : French *Ø*.

This recurs, as in 'dear', 'head'. For SOUND CORRESPONDENCE 5, we would reconstruct **o* (majority wins), assuming that Portuguese changed final **o* to *u*, and that French lost final **o*.

With more extensive data (many more cognate sets than presented in Table 5.1), we would confirm these reconstructions, with their attendant sound changes and the conditions under which they took place, and we would eventually find all the sound correspondences and postulate reconstructions for all the sounds of the proto-language and work out its phonemic inventory and phonological patterns.

Step 4: Determine the status of similar (partially overlapping) correspondence sets

Some sound changes, particularly conditioned sound changes, can result in a proto-sound being associated with more than one correspondence set. These must be dealt with to achieve an accurate reconstruction. To see how this is done, we will work through an example. For this, let us consider some additional cognate sets in Romance languages, those of Table 5.3 (numbered to follow those of Table 5.1).

Based on the forms of Table 5.3, we set up a sound correspondence for the initial sound in these forms:

Sound correspondence 6:
Italian *k* : Spanish *k* : Portuguese *k* : French *k*

For SOUND CORRESPONDENCE 6, since all the languages have the same sound, *k*, we would naturally reconstruct **k*. However, SOUND CORRESPONDENCE 6 is quite similar to SOUND CORRESPONDENCE 1 (in Table 5.1), for which we also tentatively reconstructed **k*, repeated here for comparison with SOUND CORRESPONDENCE 6:

Sound correspondence 1:
Italian *k* : Spanish *k* : Portuguese *k* : French *ʃ*

The two sets overlap partially, since both sets share some of the same sounds. In fact, the only difference between the two is in French, which has *k* in SOUND

CORRESPONDENCE 6 but ʃ in SOUND CORRESPONDENCE 1. In cases such as this of similar (partially overlapping) correspondence sets, we must determine whether they reflect two separate proto-sounds or only one which split into more than one sound in one or more of the languages. In the case of SOUND CORRESPONDENCEs 1 and 6, we must determine whether both sets reflect **k*, or whether we must reconstruct something distinct for each of the two. Because we assume that sound change is regular, the options for possible solutions here are restricted to essentially only two. One possible solution would be for us to find evidence to show that the two correspondence sets are different today but represent only a single proto-sound. To show this, it would be necessary to explain away the difference between the two sets, that is, to show how a single original sound could change in ways that would result in the two different correspondence sets. For this, we would need to show that a single original sound ended up as ʃ in certain specific environments in French but as *k* in other circumstances – since the other languages all have only the single reflex, *k*, the most likely candidate is a **k* assumed not to have changed in these languages, but, under this hypothesis, changed to ʃ in French only in specific instances. If we cannot succeed in showing this – in being able to predict where the postulated original **k* became ʃ and where it remained *k* in French – then we cannot reconstruct a single sound for the two sets and we are forced to consider the other possible solution. In this other possible solution, the two correspondence sets represent two distinct sounds in the proto-language which merged to *k* in all contexts in Italian, Spanish and Portuguese, but remained distinct in French.

In this case, we are able to determine the context in which French sometimes but not always changed **k* to ʃ. We notice that in the cognate sets of Table 5.1 which exhibit SOUND CORRESPONDENCE 1, this sound comes before ɛ in French and *a* in the other languages (SOUND CORRESPONDENCE 3), while in SOUND CORRESPONDENCE 6, illustrated by the cognate sets in Table 5.3, the initial sound is not before *a* or ɛ (as in SOUND CORRESPONDENCE 1), but before *o* or *u* (French *u* or *y*). There-fore, we determine that French underwent a conditioned sound change, that **k* > ʃ before the vowel of correspondence set 3 (**a* which became ɛ in French), but retained **k* unchanged before the round vowels seen in the cognates of Table 5.3 (essentially **u* and **o*, though we need to go through

TABLE 5.3: Some additional Romance cognate sets

	Italian	*Spanish*	*Portuguese*	*French*	*(Latin)*	*English gloss*
6.	colore /kolore/	color /kolor/	côr /kor/	couleur /kulœr/	colōre	‘colour’
7.	correre /korere/	correr /kor̃er/	correr /korer/	courir /kuri(r)/	currere	‘to run’
8.	costare /kostare/	costar /kostar/	costar /kostar/	coûter /kuter/	co(n)stāre [‘stand firm’]	‘to cost’
9.	cura /kura/	cura /kura/	cura /kura/	cure /kyr/	cūra [‘care’]	‘cure’

the steps to reconstruct these). So, in spite of two distinct sound correspondences (1 and 6), we reconstruct a single proto-sound and show that one of these (SOUND CORRESPONDENCE 1) is the result of a conditioned change which affected only some of the instances of original **k* in French (those before original **a*) but not the other cases of **k* (those before **u* and **o*).

In some cases, however, we are forced to reconstruct separate protosounds in instances of similar, partially overlapping correspondence sets. Consider for example the two sound correspondences illustrated by the initial sounds in additional cognates in Table 5.4.

TABLE 5.4: Further Romance cognate sets

	Italian	*Spanish*	*Portuguese*	*French*	*(Latin)*	*English gloss*
10.	battere	batir	bater	battre	battuere	'to beat'
	/battere/	/batir/	/bater/	/batr/		
11.	bolla	bola	bola	boule	bulla	'ball, bubble'
	/bolla/	/bola/	/bola/	/bul/		
12.	bontà	bondad	bondade	bonté	bonitāte	'goodness
	/bonta/	/bondad/	/bõdaǰi/	/bõte/		
13.	bev-	beber	beber	boire	bibere	'to drink'
	/bev-/	/beber/	/beber/	Old French beivre		
14.	venire	venir	vir	venir	venīre	'to come'
	/venire/	/benir/	/vir/	/vənir/		
15.	valle	valle	vale	val	valle	'valley'
	/valle/	/baʲe/	/vale/	/val/		
16.	vestire	vestir	vestir	vêtir	vestīre	'to dress'
	/vestire/	/bestir/	/vestir/	/vetir/		

Cognate sets 10 to 13 show the sound correspondence in (7):

Sound correspondence 7:
Italian *b* : Spanish *b* : Portuguese *b* : French *b*

Cognate sets 14 to 16 show the sound correspondence in (8):

Sound correspondence 8:
Italian *v* : Spanish *b* : Portuguese *v* : French *v*

Clearly the best reconstruction for SOUND CORRESPONDENCE 7 would be **b*, since all the languages have *b* as their reflex. SOUND CORRESPONDENCE 8 partially overlaps with this in that Spanish has *b* for its reflex in this set as well, corresponding to *v* of the other languages. As in the case of Proto-Romance **k* (above), either we must be able to explain the difference in these two sets by showing that those languages with *v* changed an original **b* to *v* under some clearly defined circumstances, or we must reconstruct two separate sounds in the proto-language, presumably **b* and **v*, where Spanish would then be

assumed to have merged its original *v* with *b*. In this case, to make a long story short, if we look for factors which could be the basis of a conditioned change in Italian, Portuguese and French, which could explain how a single original **b* could become *v* in certain circumstances but remain *b* in others in these languages, we are unable to find any. We find both *b* and *v* at the beginnings of words before all sorts of vowels, and with more extensive data we would find that both sounds occur quite freely in the same environments in these languages. Since no conditioning factor can be found, we reconstruct **b* for the cognates in correspondence set 7 and **v* for those in correspondence set 8, two distinct proto-sounds. From this, it follows that **v* merged with **b* in Spanish, accounting for why *b* is the Spanish reflex in both cognate sets 14–16 and 10–13 of Table 5.4.

A somewhat more revealing example of the problem of overlapping correspondence sets which prove to contrast and thus require separate sounds to be reconstructed is seen in the example in Table 5.5, from Mayan languages (of which only a few, each representing a major branch of the family, are represented).

TABLE 5.5: Some Mayan cognate sets

	K'iche'	*Tzeltal*	*Yucatec*	*Huastec*	*Proto-Mayan*	*English gloss*
1.	ra:h	ya	yah	yah-	*ra:h	'hot, spicy'
2.	riʔx	yix	yiʔih	yeh-	*riʔix	'old (old man)'
3.	r-	y-	y-	—	*r-	'his/her/its'
4.	raš	yaš	yaʔaš	yaš-	*raʔaš	'green'
5.	war	way	way	way	*war	'to sleep'
6.	ya:x	yah	yah	yaʔ	*ya:h	'sick'
7.	yaš	yaš	—	—	*yaš	'crab, pincers'
8.	k'ay-	k'ay- ['sing']	k'ay- ['sing, sell']	č'ay- ['buy']	*k'ay	'to sell'

(NOTE: y = IPA [j], š = [ʃ], č = [tʃ], C' = glottalized (ejective) consonants.)

Note that the 'dash' (–) is the convention used by linguists to mean that either no cognate is known or the data are unavailable. In such instances, we must rely on information from the other cognate sets in order to determine features of those languages where the forms are missing. (In the examples that follow from Mayan, *y* = IPA [j].)

Cognate sets 1–5 show SOUND CORRESPONDENCE 1:

Sound correspondence 1:
K'iche' *r* : Tzeltal *y* : Yucatec *y* : Huastec *y*

Cognate sets 6–8 show SOUND CORRESPONDENCE 2:

Sound correspondence 2:
K'iche' *y* : Tzeltal *y* : Yucatec *y* : Huastec *y*

Clearly, by our standard criteria, the best Proto-Mayan reconstruction for SOUND CORRESPONDENCE 2 would be **y* (preserved unchanged in all the languages). However, all the languages except K'iche' also have *y* as their reflex in SOUND CORRESPONDENCE 1, whereas K'iche' has *r* in this case. As in the discussion of the Proto-Romance **k* case (above), we must either explain how the difference in these two sets arose by showing that K'iche' had changed original **y* to *r* in some clear set of phonetic circumstances, or we must reconstruct two separate sounds in the proto-language. In this case, to make a long story short, if we look for factors which could be the basis of a conditioned change in K'iche', we are unable to find any. We find both *r* and *y* at the beginning and end of words, before all sorts of vowels, and so on, and basically either sound can occur in any context without restrictions. Since no conditioning factor can be found, we reconstruct **r* for the SOUND CORRESPONDENCE 1 and **y* for SOUND CORRESPONDENCE 2, two distinct proto-sounds. From this, it follows that **r* merged with *y* in Tzeltal, Yucatec and Huastec, accounting for why they have *y* as the reflex also in cognate sets 6–8 of Table 5.5. When we look at still other Mayan languages, we find this distinction further supported, since, for example, Mam has *t* and Motocintlec has *č* where K'iche' has *r* in the cognates that illustrate SOUND CORRESPONDENCE 1, but they both have *y* in cognates where K'iche' has *y* in SOUND CORRESPONDENCE 2. That is, K'iche' turns out not to be the only witness of the distinction between the two sounds of these correspondence sets (Campbell 1977).

TABLE 5.6: Central Algonquian sound correspondences and Bloomfield's reconstruction

	Fox	*Ojibwa*	*Plains Cree*	*Menomini*	*PCA*
1.	hk	sk	sk	čk	*čk
2.	ʃk	ʃk	sk	sk	*ʃk
3.	hk	hk	sk	hk	*xk
4.	hk	hk	hk	hk	*hk
5.	ʃk	ʃk	hk	hk	*çk

There is a famous case which confirms this way of treating partially overlapping sound correspondence sets. Leonard Bloomfield's (1925, 1928) famous proof of the applicability of the comparative method in unwritten ('exotic') languages was based on the correspondence sets from Central Algonquian languages presented with his reconstructions in Table 5.6 (PCA = Proto-Central Algonquian). Bloomfield (1925) postulated the reconstruction of **çk* for set 5 as distinct from the others on the basis of scant evidence, but under the assumption that sound change is regular and the difference in this correspondence set (though exhibiting only sounds that occur in different combinations in the other sets) could not plausibly be explained away. Later, his decision to reconstruct something different for set 5 was confirmed when Swampy Cree was discovered, which contained the correspondence *htk* in the morpheme

upon which set 5 was based, distinct in Swampy Cree from the reflexes of the other four reconstructions. Based on this discovery, Bloomfield (1928: 100) concluded:

> As an assumption, however, the postulate [of sound-change without exception] yields, as a matter of mere routine, predictions which otherwise would be impossible. In other words, the statement that *phonemes change* (sound-changes have no exceptions) is a tested hypothesis: in so far as one may speak of such a thing, it is a proved truth.

Mayan languages provide a somewhat clearer and more compelling case of the need to reconstruct distinct proto-sounds if the difference between two partially overlapping correspondence sets cannot be explained away. Consider the following two K'ichean (a subgroup of Mayan) sound correspondences:

	Tz'utujil	*Kaqchikel*	*K'iche'*	*Poqomam*	*Uspanteko*	*Q'eqchi'*
(1)	x	x	x	x	x	x
(2)	x	x	x	x	x-/-(V̀)x	h

For example, the correspondence set in (1) is illustrated by:

Kaqchikel	*K'iche'*	*Poqomam*	*Uspanteko*	*Q'eqchi'*	*gloss*
čax	čax	čax	čax	čax	'pine'
k'ax	k'ax	k'ax	k'ax	k'ax	'flour'
k'o:x	k'o:x	k'o:x	k'o:x	k'o:x	'mask'

The correspondence set in (2) is seen in:

Kaqchikel	*K'iche'*	*Poqomam*	*Uspanteko*	*Q'eqchi'*	*gloss*
ča:x	ča:x	ča:x	čà:x	čah	'ashes'
ka:x	ka:x	ka:x	kà:x	—	'sky'
o:x	o:x	o:x	ò:x	o:h	'avocado'
q'i:x	q'i:x	q'i:x	q'ì:x	-q'ih (in compounds)	'day, sun'

In (1), all the languages have *x* as the reflex, and we would naturally expect to reconstruct **x* for the Proto-K'ichean sound. However, (2) overlaps considerably with (1), where each language also has *x* except Q'eqchi', which has *h*; Uspanteko has *x* too; however, if there is a vowel preceding this *x*, it has falling tone (V̀), which is not the case for vowels preceding the *x* of correspondence set (1). Since no conditioning factor can be found to explain away the difference between the two sets in Q'eqchi' and Uspanteko, separate proto-sounds must be reconstructed. It has been proposed that correspondence set (2) represents a sound which is further forward than *x*, the sound of correspondence set (1), and thus **x̲* (a somewhat fronted velar fricative) has been proposed to represent correspondence set (2). While the reconstruction with **x* and **x̲* for these two sets is not phonetically ideal, nevertheless the decision to reconstruct something different for the two is confirmed when cognates are compared from other

branches of Mayan beyond K'ichean, which exhibit the following corresponding sounds:

	Yucatec	*Chol*	*Chuj*	*Jakalteko*	*Mam*	*K'ichean*
(3)	x	h	x	x	x	*x
(4)	n	n	ŋ	ŋ	x	*x̱

The correspondence set in (3) (which matches the K'ichean set in (1)) is exemplified by:

Yucatec	*Chol*	*Chuj*	*Jakalteko*	*Mam*	*K'ichean*	*gloss*
tax	tah	tax	tah	tsax	*čax	'pine'
k'ax	č'ah	k'ax	k'ah	k'ax	*k'ax	'flour, pinole'
k'o:x	k'o:h	k'o:x	k'oh	k'o:x	*k'o:x	'mask'

In Proto-Mayan these have **x*; they are, respectively: **ţax* 'pine', **k'ax* 'flour, pinole', and **k'o:x* 'mask' (where *ţ* represents a fronted dental or palatalized 't').

The correspondence set in (4) (which matches K'ichean set (2)) is seen in:

Yucatec	*Chol*	*Chuj*	*Jakalateko*	*Mam*	*K'ichean*	*gloss*
taʔan	tan-	taʔaŋ	taŋ	tsaʔx	*čax̱	'ashes'
kaʔan	čan	čaʔaŋ	kaŋ	kyaʔx	*kax̱	'sky'
ò:n	un	oŋ	oŋ	o:x	*o:x̱	'avocado'
k'ì:n	k'in	q'iŋ-	q'iŋ-	q'i:x	*q'i:x̱	'sun, day'

In Proto-Mayan these all have **ŋ*; they are, respectively: **ţaʔŋ* 'ashes', **kaʔŋ* 'sky', **o:ŋ* 'avocado', and **q'i:ŋ* 'sun, day'.

That is, the sounds of correspondence set (3) reflect Proto-Mayan **x*, whereas those of set (4) reflect Proto-Mayan **ŋ*. Since the two sounds are clearly distinguished in the other branches of the family and descend from distinct sounds in Proto-Mayan, the validity of the decision to reconstruct different sounds for Proto-K'ichean, one branch of Mayan, is confirmed. Perhaps also the phonetics of this reconstruction could be refined. Since the *x* of K'ichean (and several other Mayan) languages is phonetically [χ] (voiceless uvular fricative), it may seem appealing to reconstruct **χ* for set (3) in K'ichean and then let **x* (velar) represent set (4). Since K'ichean languages contrast uvular and velar stops, a similar contrast in the fricative series may make some sense (see step 5).

Step 5: Check the plausibility of the reconstructed sound from the perspective of the overall phonological inventory of the proto-language

Steps 5 and 6 are related. The rule of thumb in step 5 takes advantage of the fact that languages tend to be well behaved, that is, they tend to have symmetrical sound systems with congruent patterns. For example, in the reconstruction of sounds for the individual sound correspondences in step 3, we can reconstruct each sound of the proto-language with little regard for how these sounds may relate to

one another or how they may fit together to form a coherent system. Often in step 5 when we consider the broader view of these sounds in the context of the overall inventory, we refine and correct our earlier proposals. For example, if two related languages have the correspondence set Language$_1$ *d* : Language$_2$ *r*, we might initially reconstruct **r* and assume **r* > *d* in Language$_1$, since *r* > *d* is known to take place in languages, though the alternative of **d* with the assumption that Language$_2$ underwent the change **d* > *r* is just as plausible, since the change *d* > *r* is also found in languages. Suppose, however, that in step 5 we discover that we have reconstructed sounds based on other sound correspondences which would give the following phonological inventory for the proto-language:

*p	*t	*k
*b		*g
	*r	
	*l	

There is a gap in this inventory where **d* would be expected to complete the stop series, where the voiceless stops (**p*, **t*, **k*) would each be matched by a voiced counterpart (**b*, **d*, **g*), if a **d* existed, which would make the stop series symmetrical, the pattern congruent. The proto-language as tentatively reconstructed so far, with both **r* and **l* and **b* and **g*, but no **d*, would be unusual and unexpected. However, by revising our earlier tentative reconstruction of **r* for the *d* : *r* sound correspondence to the equally plausible **d* (assuming **d* > *r* in Language$_2$), we arrive at a much more coherent and likely set of sounds for the proto-inventory, where the two stop series are congruent:

*p	*t	*k
*b	*d	*g
	*l	

While this instance is presented as a hypothetical possibility, it is in fact encountered in a number of real language families, for example in branches of Austronesian. It is important, however, to keep in mind that while languages tend to be symmetrical and have pattern congruity, this is by no means always the case.

Let's consider one other hypothetical instance, also actually found in real language families. If in a family of two languages we encounter the correspondence set Language$_2$ *s* : Language$_2$ *ʃ*, either we could reconstruct **s* (assuming **s* > *ʃ* in Language$_2$) or we could postulate **ʃ* (and assume **ʃ* > *s* in Language$_1$). Both of these changes (**s* > *ʃ* and **ʃ* > *s*) are frequently found in other languages. Suppose, however, that in step 5 we discover that the other sound correspondences justify the reconstruction of several proto-sounds in the alveolar series, including **ts*, but no other palato-alveolar sound. This would give a proto-language with alveolar **ts* but palato-alveolar **ʃ* and no **s*, but this system would be asymmetrical and odd. However, a proto-language with **ts* and **s* but lacking **ʃ* would be normal and not at all unusual. Therefore, in step 5 we would revise the

preliminary reconstruction of Step 3 to make sure that we reconstructed **s* for the *s* : ʃ correspondence set (assuming **s* > ʃ in Language$_2$) to ensure a more plausible overall phonological inventory for the proto-language which we reconstruct. A real example which fits precisely this situation comes from Mixe-Zoquean (a family of languages from southern Mexico), where the languages of the Zoquean branch have *s* corresponding to ʃ of the Mixean languages. So, for Proto-Mixe-Zoquean, **s* is a better reconstruction for the *s* : ʃ correspondence set.

Of course, languages do not have to be symmetrical or fully natural, though they tend to be. Also, it is conceivable that a proto-language might have gaps (such as the missing **d* in the first example) and asymmetries (**ts* and *ʃ rather than **ts* and **s* in the second example); however, unless there is strong evidence to compel us to accept a less expected reconstruction, we are obliged to accept the ones motivated by pattern congruity, symmetry and naturalness. That is, languages in general have symmetrical (natural) systems much more often than not. Therefore, in the case of two possibilities, one with a more expected inventory and the other with a less expected, less normal inventory, the probability that the reconstruction with the symmetrical, natural system accurately reflects the structure of the formerly spoken proto-language is much higher than that the asymmetrical one does. Given the greater odds of the first being right, we choose it, not the second, which is less likely to have existed.

Step 6: Check the plausibility of the reconstructed sound from the perspective of linguistic universals and typological expectations

Certain inventories of sounds are found with frequency among the world's languages while some are not found at all and others only very rarely. When we check our postulated reconstructions for the sounds of a proto-language, we must make sure that we are not proposing a set of sounds which is never or only very rarely found in human languages. For example, we do not find any languages which have no vowels whatsoever. Therefore, a proposed reconstructed language lacking vowels would be ruled out by step 6. There are no languages with only glottalized consonants and no plain counterparts, and therefore a reconstruction which claimed that some proto-language had only glottalized consonants and no non-glottalized counterparts would be false. Languages do not have only nasalized vowels with no non-nasalized vowels, and so we never propose a reconstruction which would result in a proto-language in which there are only nasalized vowels.

Let us look at an actual case. The Nootkan family has the sound correspondences seen in Table 5.7. Since no other guidelines help here, we might be tempted, based on the majority-wins principle, to reconstruct voiced stops for Proto-Nootkan for these four correspondence sets and postulate that these changed to the nasal counterparts in Nootka. However, only a very few languages of the world lack nasal consonants; therefore, we do not expect a nasalless proto-language, and any postulated proto-language which lacks nasals altogether must be supported by very compelling evidence. In this case, Nitinat and Makah belong to the area of the Northwest Coast of North America where languages of several different families lack nasal consonants. The lack of nasals in these

languages is due to the influence of other nasalless languages in the linguistic area (see Chapter 12); Proto-Nootkan had nasals, as Nootka still does, but Makah and Nitinat lost nasality – their former nasals became corresponding voiced oral stops (**m* > *b*, **n* > *d*, **m̓* > *b'*, **n̓* > *d'*). The knowledge of universals and typological expectations in this case would direct us to reconstruct the proto-language with nasals and to assume a subsequent change in Makah and Nitinat.

Of course, in step 5, we also relied on general typological patterns in language and evaluated proposed proto-inventories on this basis; that is, steps 5 and 6 are not really distinct.

TABLE 5.7: Nootkan correspondences involving nasals

	Makah	*Nitinat*	*Nootka*
1.	b	b	m
2.	d	d	n
3.	b'	b'	m̓
4.	d'	d'	n̓

Step 7: Reconstruct individual morphemes

When we have reconstructed the proto-sounds from which we assume that the sounds in the sound correspondences descend, it is possible to reconstruct lexical items and grammatical morphemes. For example, from the cognate set for 'goat' in Table 5.1, the first sound (in SOUND CORRESPONDENCE 1) was reconstructed as **k* (based on the *k* : *k* : *k* : ʃ correspondence set); for the second sound in the cognates for 'goat', we reconstructed **a*, as in SOUND CORRESPONDENCE 3 (with *a* : *a* : *a* : ɛ); the third sound is represented by SOUND CORRESPONDENCE 2 (*p* : *b* : *b*: *v*), for which we reconstructed **p*; the next sound in cognate set 1, as represented by SOUND CORRESPONDENCE 4, reflects Proto-Romance **r* (based on the *r* : *r* : *r* : *r* correspondence set); and the last sound in the 'goat' cognates reflects SOUND CORRESPONDENCE 2 (or actually a modification of it involving final vowels in French) which was reconstructed as **a*. Putting these reconstructed sounds together following the order in which they appear in the cognates for 'goat' in set 1, we arrive at **kapra*. That is, we have reconstructed a word in Proto-Romance, **kapra* 'goat'. For cognate set 2 'dear' in Table 5.1, we would put together **k* (SOUND CORRESPONDENCE 1), **a* (SOUND CORRESPONDENCE 3), **r* (SOUND CORRESPONDENCE 4) – all seen already in the reconstruction of 'goat' – and **o* (SOUND CORRESPONDENCE 5, with *o* : *o* : *u* : *Ø*), giving us the Proto-Romance word **karo* 'dear'. For cognate set 3 'head', we have combinations of the same correspondence sets already seen in the reconstructions for 'goat' and 'dear', SOUND CORRESPONDENCEs 1, 3, 2 and 5, giving the Proto-Romance reconstructed word **kapo* 'head'. In this way, we can continue reconstructing Proto-Romance words for all the cognate sets based on the sequence of sound correspondences that they reflect, building a Proto-Romance lexicon.

The reconstruction of a sound, a word or large portions of a proto-language is,

in effect, a hypothesis (or better said, a set of interconnected hypotheses) concerning what those aspects of the proto-language must have been like. Aspects of the hypothesized reconstruction can be tested and proven wrong, or can be modified, based on new insights. These insights may involve new interpretations of the data already on hand, or new information that may come to light. The discovery of a heretofore unknown member of the family may provide new evidence, a different testimony of the historical events which transpired between the proto-language and its descendants, which could change how we view the structure and content of the proto-language. There are a number of well-known cases where this has happened which illustrate this point. Bloomfield's Swampy Cree case has already been mentioned. With the discovery and decipherment of Hittite (or better said, the languages of the Anatolian branch of Indo-European), the whole picture of Proto-Indo-European phonology changed; this included clearer evidence of several new proto-sounds (the laryngeals).

5.3 A Case Study

Let us apply the comparative method in a somewhat more complex example (though still simplified) which illustrates what we have until now been considering mainly through a simplified comparison of Romance languages. The forms in Table 5.8 are cognates found in Finnish, Hungarian and Udmurt (Votyak). These languages belong to the Finno-Ugric family, but since there are many other languages also in this family (see Figure 6.2 in Chapter 6), the data in this example are far from complete enough to offer a full perspective on the proto language – these three are compared here only for illustration's sake. These languages separated from one another a very long time ago, which explains why some of these cognates are not as immediately apparent based on mere superficial similarity. The languages have undergone many changes and are now quite different, and we would need much more information than presented here to reconstruct all the sounds of Proto-Finno-Ugric. Therefore, here we will be concerned only with the initial sounds in these data.

Step 1 is already done; the cognates have been assembled in Table 5.8. In step 2, we compare these cognates and set up sound correspondences. It is helpful to keep a good record of what we have looked at, either by noting with each sound correspondence the numbers which identify the cognate sets in which it is found, or if we do not use numbers, then the glosses. This is just a matter of bookkeeping – a means of being able to go back and check things without having to search back through all the data to find the cognates which exhibit the correspondence in question, particularly useful, for example, in steps 5 and 6.

Sound correspondences found in the cognates of Table 5.8 are:

(1) Finnish *p-* : Hungarian *f-* : Udmurt *p-* (in Set I, nos 1–11)
(2) Finnish *t-* : Hungarian *t-* : Udmurt *t-* (in Set II, nos 12–17)
(3) Finnish *k-* : Hungarian *h-* : Udmurt *k-* (in Set III, nos 18–26)
(4) Finnish *k-* : Hungarian *k-* : Udmurt *k-* (in Set IV, nos 27–35)
(5) Finnish *s-* : Hungarian *s-* : Udmurt *ɕ-* (in Set V, nos 36–41)
(6) Finnish *s-* : Hungarian Ø- : Udmurt *s-* (in Set VI, nos 42–48)

TABLE 5.8: Some Finno-Ugric cognate sets

	Finnish	*Hungarian*	*Udmurt (Votyak)*	*gloss*
Set I				
1.	pää [pæ:]	fej [fej]	pum, puŋ	‘head, end’
2.	pata [pata]	fazék [fɔze:k]	—-	‘pot’
3.	pato ‘dam, wall’	fal [fɔl] ‘wall’	—	
4.	pääsky- [pæ:sky]	fecske [fečke]	poçkɨ-	‘swallow’ (bird)
5.	pelkää- [pelkæ:-]	fél [fe:l]	pulɨ-	‘to fear’
6.	pesä [pesæ]	fészek [fe:sek]	puz-	‘nest’
7.	pii [pi:] ‘tooth of rake’	fog [fog]	pinʲ	‘tooth’
8.	pilvi [pilvi]	felhő [felhø:]	pilʲem	‘cloud’
9.	poika [poika]	fiú [fiu:]	pi	‘boy’
10.	puno- [puno-]	fon [fon]	pun-	‘spin, braid’
11.	puu [pu:]	fa [fɔ]	pu	‘tree’
Set II				
12.	tä- [tæ-]	té- [te:-] (cf. tétova ‘here and there’)	ta	‘this’
13.	täi [tæi]	tetű [tety:]	tei	‘louse’
14.	talvi [talvi]	tél [te:l]	tol	‘winter’
15.	täyte- [tæyte-]	tel- [tel-] (in derived forms)	—	‘full’
16.	tunte- [tunte-]	tud [tud]	tod	‘to know, sense’
17.	tyvi [tyvi]	tő [tø:]	[dinʲ]	‘base’
Set III				
18.	kala [kala]	hal [hɔl]	—	‘fish’
19.	kalime- [kalime-]	háló [ha:lo:]	[Komi *kulem*]	‘fishnet’
20.	kamara [kamara]	hám- [ha:m-]	kəm	‘peel’
21.	koi [koi]	haj- [hɔj-]	[Komi *kɨa*]	‘dawn’
22.	kolme [kolme]	három [ha:rom]	kuinʲm-	‘three’
23.	kota [kota]	ház [ha:z]	kwa-/-ko/-ka ‘summer hut’/ ‘house’	‘hut’
24.	kunta [kunta] ‘community, group, society’	had [hɔd] ‘army’	—	

TABLE 5.8: continued

	Finnish	*Hungarian*	*Udmurt (Votyak)*	*gloss*
Set III				
25.	kuole- [kuole-]	hal [hɔl]	kul-	‘to die’
26.	kusi [kusi]	húgy [hu:dʲ]	kɨʑ	‘urine’
Set IV				
27.	käte- [kæte-]	kéz [ke:z]	ki	‘hand’
28.	keri [keri]	kér [ke:r]	kur	‘(tree-) bark’
29.	kerjää- [kerjæ:]	kér [ke:r]	kur-	‘to beg’
30.	kii- [ki:-] ‘rut, mating’	kéj [ke:j] ‘(carnal) pleasure’	[*Komi* koj-] ‘to make mating call’	
31.	kivi [kivi]	kő [kø:]	kə ‘mill stone’	‘stone’
32.	kyynel [ky:nel]	könny [kønnj]	-kɨlʲi- (in *çin-kɨlʲi*; *çin(m)-* ‘eye’)	‘tear’ (noun)
33.	kytke- [kytke-]	köt [køt]	kɨtk-ɨ ‘to harness’	‘to tie’
34.	kyy [ky:] ‘adder’	kígyó [ki:dʲo:]	kɨj	‘snake’
35.	kyynär [ky:nær]	könyök [kønʲøk]	[gɨr-]	‘elbow’
Set V				
36.	salava [salava] ‘willow’	szil [sil] ‘elm’	—	
37.	sarvi [sarvi]	szarv [sɔrv]	çur, çɨr	‘horn’
38.	sata [sata]	száz [sa:z]	çu	‘hundred’
39.	silmä [silmæ]	szem [sem]	çinm-	‘eye’
40.	suu [su:]	szá(j) [sa:j]	çu- (?) (in compounds)	‘mouth’
41.	sydäme- [sydæme-]	szív [si:v]	çulem	‘heart’
Set VI				
42.	sappi [sappi]	epe [epe]	sep	‘gall’
43.	sää [sæ:] ‘weather’	ég [e:g] ‘sky’ <‘weather’	[*Komi* sɨnəd] ‘sunshine haze, mist’	
44.	säynä- [sæynæ-]	őn [ø:n]	son- (son-tçorɨg, tçorɨg ‘fish’)	‘fish (*Leuciscus idus*)’

TABLE 5.8: continued

	Finnish	*Hungarian*	*Udmurt (Votyak)*	*gloss*
Set III				
45.	sula-[sula]	olva- [olvɔ-]	sɨlm-	'to melt'
46.	suoni [suoni]	ín [i:n]	sɨn	'sinew'
47.	syksy [syksy]	ősz [ø:s]	sizʲɨl	'autumn'
48.	syli [syli]	öl [øl]	sul, sɨl	'lap, bosom'

Note that in a few cases where Udmurt has no cognate or the cognate is unknown, cognate forms from closely related Komi have been included for comparison.)

In Step 3 we attempt to reconstruct the proto-sound which we believe is reflected in each of these correspondence sets. For SOUND CORRESPONDENCE (1) (*p* : *f* : *p*) our choices are: [1] reconstruct **p* and assume Hungarian changed to *f*; [2] reconstruct **f* and assume Finnish and Udmurt changed this to *p*; or [3] reconstruct some third thing (say **p*^{h}) and assume that it changed in all three languages, that Hungarian changed in one way to give *f* while Finnish and Udmurt changed in another to give *p*. From directionality of change as a guideline, we conclude that possibilities [1] (**p*) and [3] (some third thing, like **p*^{h}) are plausible, but not [2] (**f*), since in sound changes familiar from languages around the world we see that voiceless bilabial stops (*p, p*^{h}) frequently become *f*, but extremely rarely do we find instances of *f* changing to *p* or *p*^{h}. In the majority-wins guideline, since Finnish and Udmurt both have *p*, against Hungarian alone with *f*, majority wins suggests **p* as a more likely reconstruction than **f*. In the guideline of factoring in features held in common, we may conclude from the sounds *p* and *f* in the sound correspondence that the proto-sound was voiceless and a labial of some kind, but this is consistent with all three of the possibilities [1]–[3]. In this case, then, factoring in the common features provides no basis for choosing among the alternatives. The guideline of economy also urges us towards [1] (**p*). With **p* (as in [1]), we would need to postulate only a single change, **p* > *f* in Hungarian; in choice [2] (**f*) we would have to assume the change of **f* > *p* twice, in Finnish and again in Udmurt. Choice [3] (**p*^{h}) would require us to postulate the change **p*^{h} > *p* twice, in Finnish and Udmurt, and another change, **p*^{h} > *f*, in Hungarian. Steps 4 and 5 can help us resolve which of these possibilities is the best reconstruction; however, we have sufficient reason now for selecting [1], with **p*, based on these considerations from directionality of change, majority wins, and economy.

SOUND CORRESPONDENCE (2) (*t*- : *t*- : *t*-) appears to reflect **t*- (where none of the language has changed).

SOUND CORRESPONDENCES (3) (*k*- : *h*- : *k*-) and (4) (*k*- : *k*- : *k*-) could present more of a challenge. In (4) we reconstruct **k*-, since all three languages have *k*- and thus none of them appears to have changed. However, if (4) were not present

to complicate the picture, then (3) would also seem to be best reconstructed as **k-*. Directionality of change would support this possibility, since the change *k* > *h* (as would be required for Hungarian in this hypothesis) is very common and not unexpected, whereas a change *h-* > *k-* is all but unknown. Also the majority-wins criterion supports **k-*, with *k-* in two languages but *h-* in only one. We move to Step 4 to attempt to resolve the difficulty of the partially overlapping sound correspondences (3) and (4). If we can show that both sound correspondence sets reflect the same original sound because one of the languages has undergone a conditioned change where that sound changed in some environments but not in others, then we can reconstruct just a single sound, the same one for both correspondence sets. We would explain the difference between the two correspondences by pointing out the conditions under which one of the languages changed and thus resulted in two different outcomes from the single original sound. If we cannot explain the difference in this way, then we are obligated to reconstruct two distinct proto-sounds, one to represent each of the two sound correspondences, with the assumption that these two originally distinct sounds merged to *k-* in Finnish and Udmurt. This, then, requires us to take a closer look at the cognates in question (those of Sets III and IV). We notice that in the cognates of Set III the *h* of Hungarian appears only before back vowels (*u, o, a*), whereas in the cognates of Set IV Hungarian's *k* occurs only before front vowels. We conclude that Hungarian had a single original sound which changed to *h* before back vowels (as in Set III) but remained *k* before front vowels (as in Set IV). We reconstruct **k*. Someone might wonder whether the proto-language could not have had an **h* which then changed to *k* before front vowels in Hungarian and to *k* in all environments in Finnish and Udmurt. First, directionality argues against this possibility (since the change *h* > *k* is essentially unknown anywhere). Second, the criterion of economy also go against this alternative; it is more likely that only one change took place, **k* > *h* before back vowels in Hungarian, than that several independent changes occurred, one of **h* > *k* before front vowels in Hungarian and independently of the Hungarian development the changes of **h* > *k* in all contexts in Finnish and in Udmurt.

The SOUND CORRESPONDENCES (5) (*s-* : *s-* : *ɕ-*) and (6) (*s-* : *Ø-* : *s-*) present a similar problem of partially overlapping correspondence sets. However, the partial overlap in this instance is not like that seen in the sound correspondences (3) and (4), both of which come from a single original sound in different positions due to conditioned sound change. Both sound correspondences (5) and (6) in the cognates of Sets V and VI occur essentially in the same environments: both before the various vowels, front and back, and both before the same sorts of consonants in the following syllable (for example, *l*, of 36, 45, and 48), which would be clearer if we had more cognates in the data presented here. Careful scrutiny in this case eventually shows that it is not possible to explain the difference between the two sound correspondence sets as conditioned phonetic change in some environment, given that both occur in essentially the same environments. This being the case, we have no choice but to reconstruct a different proto-sound to represent each of these two sound correspondences. Let us see how the general guidelines for reconstruction fare in these partially overlapping but ultimately contrastive cases, first applied to (5), then to (6), with the results then compared.

By *directionality*, for (5) (*s*- : *s*- : *ɕ*-) we might assume either **s* which became *ɕ* in Udmurt, or **ɕ* which became *s* in Finnish and Hungarian. Both are known changes, though *s* > *ɕ* is not common without some conditioning environment, say before front vowels. Thus, while not compellingly strong, the directionality in this instance gives a slightly stronger vote for **ɕ* than for **s*, that is, for the change **ɕ* > *s* being the most likely. On the other hand, *majority wins* clearly votes for **s*, since two languages have *s* (Finnish and Hungarian) and only one has *ɕ* (Udmurt). The criterion of examining the *features held in common* avails little in this instance, since *s*- and *ɕ*- share all their features except palatalization, meaning the proto-sound presumably had all these same shared features – some kind of *s*-like sound. *Economy* would clearly favor **s*, since this would require only one change, **s* > *ɕ* in Udmurt; the postulation of **ɕ* would require the change of **ɕ* > *s* in Finnish and again in Hungarian. In sum, the guidelines do not all unanimously point in one direction, but appear to favor **s* for (5), which presumes the change **s* > *ɕ* in Udmurt.

However, the existence of sound correspondence (6) (*s*- : *Ø*- : *s*-) complicates this picture, since it, too, appears to point to **s* as the best probable reconstruction, and yet we were unable to combine the two as possibly coming from the same original sound with some conditioned changes in particular contexts. *Directionality* clearly favors **s* for (6), since *s* > *Ø* is a relatively frequent change (often through the intermediate stage of *s* > *h* > *Ø*), but *Ø* > *s* is unknown and there is no phonetic motivation for why such a change should take place. *Majority wins* also clearly favors **s*, given the two cases with *s* (Finnish and Udmurt) but only one with *Ø* (Hungarian). Similarly, the *features held in common* suggest **s*, since *s* is the sound in two of the languages, and the features of *Ø* do not contribute insight here. Finally, *economy* also supports **s* for (6), since this would require only the single change of **s* > *Ø* in Hungarian. Postulation of **Ø* , for example, because that is the reflex in Hungarian, would require the change of *Ø* to *s* in Finnish and again in Udmurt. Postulation of some third alternative, say **ʃ*, would require even more changes, **ʃ* > s in Finnish and in Udmurt, and , **ʃ* > *Ø* in Hungarian. In sum, then, the guidelines support **s* for (6), with the presumed change **s* > *Ø* in Hungarian.

However, this cannot be right. As already indicated, the two correspondences (5) and (6) occur in contrastive environments and apparently cannot be combined as separate outcomes from the same original sound due to conditioned change. This means that we cannot, then, reconstruct **s* both for (5) and for (6), since sound change is regular and such a reconstruction would afford no means of explaining why the proposed single original **s* behaves differently in the two different correspondence sets, why in Hungarian it is sometimes *s* (in Set V cognates) and sometimes *Ø* (in Set VI cognates), why in Udmurt sometimes *ɕ*, sometimes *s*, and so on. We must reconstruct a separate sound for each of these distinct correspondence sets. While the decision about what to reconstruct for each is not as straightforward as we might like, all the guidelines clearly suggest **s* for (6) (*s*- : *Ø* : *s*), where for (5) (*s*- : *s*- : *ɕ*-) there was not such agreement – directionality appeared to favour **ɕ*. Let us then propose these reconstructions:**ɕ* for (5) (postulating the sound changes **ɕ* > *s* in Finnish and in Hungarian), and **s* for (6) (with the changes **s* > *Ø* in Hungarian). In fact,

with the aid of much additional evidence from other Finno-Ugric languages, specialists reconstruct **ś* (IPA [sʲ] or aveolo-palatal [ɕ]) for the sound correspondence of (5) and **s* for that of (6) (Sammallahti 1988).

Let us return to SOUND CORRESPONDENCES (1), (2), (3) and (4) and apply Steps 5 and 6. Not enough cognate sets are given in the data here to reconstruct the full phonological inventory of Proto-Finno-Ugric, so that we are unable to apply Steps 5 and 6 fully. However, for now let us assume that we at least have available in the cognates of Table 5.8 the evidence for the voiceless stops and apply these steps to these to illustrate the procedures. Our tentative reconstructions to this point based on the sound correspondences were:

**p* (1) Finnish *p-* : Hungarian *f-* : Udmurt *p-* (in Set I, nos 1–11)
**t* (2) Finnish *t-* : Hungarian *t-* : Udmurt *t-* (in Set II, nos 12–17)
**k* (before back vowels) (3) Finnish *k-* : Hungarian *h-* : Udmurt *k-* (in Set III, nos 18–26)
**k* (before front vowels) (4) Finnish *k-* : Hungarian *k-* : Udmurt *k-* (in Set IV, nos 27–35).

We check these in Step 5 to see how plausible the resulting inventory of voiceless stops would be with these sounds in the proto-language. A language with the stops *p, t, k* would be quite normal, with an internally consistent pattern of voiceless stops. If we did attempt to reconstruct possibility [3] (some third thing from which to derive *p* and *f* plausibly, say **p*^{h}) for sound correspondence (1), we would no longer have a natural, symmetrical phonemic inventory of voiceless stops (**p*, **t*, **k*), but rather the unlikely **p*^{h} , **t*, **k*. In Step 5, we would see that this would result in a series of stops which is not internally consistent, where the presence of aspirated *p*^{h} (with no plain *p*) is incongruent with *t* and *k* (with no *t*^{h} and *k*^{h}). In Step 6, we would check this pattern to see how well it fits typologically with what we know of the sound systems of the world's languages. Here we would find that languages with only the stops *p*^{h} *, t, k* (but no *p* and no other aspirated stops) are very rare, while a large majority of languages have a stop series with *p, t, k*. For possibility [2] (which would reconstruct **f*), Step 5 tells us a language with *f, t, k* (but no *p*) is also internally not as consistent as one with *p, t, k*, and therefore not as good a reconstruction. Step 6 tells us the same thing; in looking at the sound systems of the world's languages, we find few with *f, t, k* (and no *p*), but hundreds with *p, t, k*. Putting these considerations of directionality, economy, internal consistency and typological realism together, we conclude that the reconstruction of **p* is the best of the alternatives for SOUND CORRESPONDENCE (1). In turn, we apply steps 5 and 6 to the reconstructions with **t* and **k* and we find these to be supported in similar fashion in these steps. We find that the possible alternative with **h* for SOUND CORRESPONDENCES (3) and (4) which might have been considered, would be inconsistent internally and typologically (leaving a system with *p, t, h*, but no *k*) not to mention being against economy, the known directionality of change, and the majority-wins guidelines.

5.4 Indo-European and the Regularity of Sound Change

The development of historical linguistics is closely associated with the study of Indo-European. *Grimm's Law*, *Grassmann's Law* and *Verner's Law* are major milestones in the history of Indo-European and thus also in historical linguistics, and traditionally all linguists have had to learn these laws – indeed, knowledge of them is helpful (some might say essential) for understanding the comparative method and the regularity hypothesis. (These laws have been considered in preliminary form in Chapter 2.) In this section, each is taken up individually and the development of the claim that sound change is regular based on these laws is considered.

5.4.1 Grimm's Law

The forms of Table 5.9 illustrate Grimm's Law, a series of changes in the stops from Proto-Indo-European to Proto-Germanic:

voiceless stops > voiceless fricatives:

*p	>	f
*t	>	θ
*k, *$\hat{k}$	>	h (x)
*k^w	>	hw

voiced stops > voiceless stops

*b	>	p
*d	>	t
*g, *$\hat{g}$	>	k
*g^w	>	kw

voiced aspirated (murmured) stops > plain voiced stops

*bh	>	b
*dh	>	d
*gh, *$\hat{g}$h	>	g
*g^wh	>	gw, w

Note here that many scholars believe that the voiced aspirates did not become plain voiced stops directly, but rather went through an intermediate stage of becoming voiced fricatives, which then later hardened to voiced stops (or became *w* in the case of *$*g^wh$*): *bh > β > b, *dh > ð > d, *gh > γ > g, *$\hat{g}$h > γ^j > ĝ, *g^wh > γ^w > w. (The sounds *$\hat{k}$, *$\hat{g}$ and *$\hat{g}h$ represent the 'palatal' series in Indo-European.)

(Not all the stops are included in Table 5.9.) In Table 5.9, the Gothic and English forms show the results of these changes in Germanic, while the Sanskrit, Greek and Latin forms for the most part reflect the Indo-European stops unchanged; that is, they did not undergo Grimm's Law as the Germanic forms did.

Grimm's Law embodies systematic correspondences between Germanic and non-Germanic languages, the results of regular sound changes in Germanic. So, for example, as a result of the change *$p > f$ in the examples in Set Ia of Table 5.9, Gothic and English (the Germanic languages) have the reflex *f* corresponding to

TABLE 5.9: Indo-European cognates reflecting Grimm's Law

Sanskrit	*Greek*	*Latin*	*Gothic*	*English*
Set Ia: *p > f				
pad-	pod-	ped-	fōtus	foot
páńča [páɲča]	pénte	[quinque] [kʷinkʷe]	fimf	five
pra-	pro-	pro-	fra-	fro
pū-'make clear, bright'	pur	pūrus 'pure'	[OE fȳr]	fire
pitár-	patḗr	pater	fadar [faðar]	father [OE fæder]
nápāt-'descendant'		nepōs 'nephew, grandson'	[OHG nefo]	nephew [OE nefa]
Set Ib: *t > θ				
trī-/tráyas	treīs/tría	trēs	þrija	three
tv-am	tū (Doric)	tu	þu	thou
-ti-	-ti-	-tis/-sis		-th 'nominalizer'
gátis	básis	mor-tis		[health, birth, death]
'gait'	'going'	'death'		
Set Ic: *k, *k̂ > h (or [x])				
śván- [ʃvən-]	kúōn	canis [kanis]	hunds	hound 'dog'
śatám [ʃətə́m]	(he-)katón	centum [kentum]	hunda (pl.)	hundred
kravís 'raw flesh'	kré(w)as 'flesh, meat'	cruor 'raw, blood, thick'		raw [OE hrāw] 'corpse'
dáśa [də́ʃə]	déka	decem [dekem]	taíhun [tɛxun]	ten
Set IIa: *b > p (**b* was very rare in Proto-Indo-European, and many doubt that it was part of the sound system; some Lithuanian forms are given in the absence of cognates in the other languages)				
		(*Lithuanian*) dubùs	diups	deep [OE dēop]
	kánnabis	(*Lithuanian*) kanapẽs]		hemp (borrowing?)
		Latin lūbricus	sliupan	slip

TABLE 5.9: continued

Sanskrit	*Greek*	*Latin*	*Gothic*	*English*
Set IIb: *d > t				
d(u)vā́-	dúo/dúō	duo	twái [twai]	two
dánt-	odónt-	dent-	tunþus	tooth
dáśa [də́ʃə]	déka	decem [dekem]	taíhun [tɛxun]	ten
pad-	pod-	ped-	fōtus	foot
ad-	édō	edō		eat [OE etan]
‘eat’	‘I eat’	‘I eat’		
véda	woīda	videō	wáit	wit ‘to know’
‘I know’	‘I know’	‘I know’	[wait] ‘I know’	
Set IIc: *g, *ĝ > k				
janás	génos	genus	kun-i ‘race, tribe’	kin
jánu-	gónu	genū	kniu	knee
jnātá	gnōtós	(g)nōtus	kunnan ‘to know’	known
ájˇra- ‘country’	agrós	ager	akrs	acre ‘field’
mr̥j- ‘to milk’	(a-)mélgō ‘to squeeze out’	mulgeō ‘I milk’	miluk-s ‘milk’	milk
Set IIIa: *bh > b				
bhar-	phér-	fer-	baír-an [bɛran] ‘to bear’	bear
bhrā́tar	phrā́tēr	frā́ter	brōþar	brother
a-bhū-t	é-phū	fu-it	bau-an [bɔ̄-an] to dwell’	be
‘he was’	‘he grew, sprang up’	‘he was’		
Set IIIb: *dh > d				
dhā- ‘put’	ti-thē-mi ‘I put’	fē-cī ‘I made’		do [OE dō-n]
dhr̥ṣnóti ‘he dares’	thrasús ‘bold’	(fest-)	(ga-)dars ‘he dares’	dare
dvār-	thúr-a	for-ēs	daúr- [dor-]	door

TABLE 5.9: continued

Sanskrit	*Greek*	*Latin*	*Gothic*	*English*
vidhávā	ē-wíthewos 'unmarried youth'	vidua	widuwo	widow
mádhu	méthu			mead
madhya-	mésos	medius	midjis	mid
Set IIIc: *gh, *ĝh > g				
haṁs-á- [hə̃sə́] 'swan, goose'	khēn	āns-er	Gans [*German*]	goose
stigh- 'stride'	steíkhō 'I pace'		steigan [stīgan] 'to climb'	
vah- 'carry'	wókh-os 'chariot'	veh-ō 'I carry'	ga-wig-an 'to move, shake'	weigh/wain

p in Sanskrit, Greek and Latin (the non-Germanic languages), all from Proto-Indo-European **p*. While Grimm's Law accounts for the systematic correspondences seen in Table 5.9, nevertheless these are not entirely without exceptions. However, as we will see, these exceptions all have satisfactory explanations. One set of forms which seem to be exceptions to Grimm's Law involves stops in consonant clusters, and examples of these are given in Table 5.10. (An Old

TABLE 5.10: Exceptions to Grimm's Law in consonant clusters

	Sanskrit	*Greek*	*Latin*	*Gothic*	*English*
1.	páś-	[skep-]	spec-	[OHG speh-]	spy (?) 'to see'
2.	(ṣṭhiv-)	pū	spu-	speiw-an [spīw-an]	spew 'to spit'
3.	aṣṭáu [əṣtáu]	oktō	octō [oktō]	ahtau [axtau]	eight
4.	nákt-	nukt-	noct- [nokt-]	nahts [naxts]	night
5.			capt(īvus)	(haft)	[OE hæft] 'prisoner'
6.	-ti- gátis 'gait'	-ti- básis 'going'	-tis/-sis mor-tis 'death'		-t 'nominalizer' thrift, draught, thirst, flight, drift
7.			piscis [piskis]	fisks	[OE fisc] 'fish'

High German (OHG) form is sometimes substituted when no Gothic cognate is available; OE = Old English.)

In these forms, by Grimm's Law, corresponding to the *p* in (1) and (2) of Sanskrit, Greek and Latin we should expect to find *f* in Gothic and English, not the *p* seen in these forms. (And given the *p* of Gothic and English, the Germanic languages, we expect the correspondence in Sanskrit, Greek and Latin to be *b*, not the *p* that actually occurs.) In (3–6) we expect Gothic and English to have /θ/ (not the actually occurring *t*) corresponding to the *t* of Sanskrit, Greek and Latin. And in (7), we would expect Latin *k* to correspond to Germanic *x*, not to the *k* of the Gothic and English words in this cognate set. These exceptions are explained by the fact that Grimm's Law was actually a conditioned change; it did not take place after fricatives (**sp* > *sp*, not ✘*sf*) or after stops (**kt* > *xt*, not ✘*xθ*; the **k*, the first member of the cluster, does change to *x* as expected by Grimm's Law, but the **t*, the second member, does not change). In the case of (6), the difference between *thrift*, *draught*, *thirst*, *flight*, *drift* of Table 5.10 and the *health*, *birth*, *death* of Table 5.9 is explained in the same way. The /θ/ forms (as in Table 5.9) underwent Grimm's Law (**t* > *θ*); the forms with *-t* (in Table 5.10) are exempt from Grimm's Law because this **t* comes after a fricative in English (the <gh> of *draught* and *fight* was formerly [x], which was later lost; see Chapter 15). Thus, when Grimm's Law is correctly formulated – written to exclude stops after fricatives and other stops in consonant clusters, since that environment did not enter the change – the stops in clusters are not, in fact, exceptions to the sound change.

5.4.2 Grassmann's Law

Another set of forms which earlier had seemed to be exceptions to Grimm's Law is explained by Grassmann's Law (seen already in Chapter 2). In Greek and Sanskrit, Grassmann's Law regularly dissimilated the first of two aspirated stops within a word so that the first lost its aspiration, as in the change from Proto-Indo-European **dhi-dhē-mi* (**dhi-dheh*$_i$*-mi*) 'I put, place' (with reduplication of root *dhē-* (**dheh*$_i$*-*)) to Sanskrit *da-dhā-mi* and Greek *ti-thē-mi*. As a result of Grassmann's Law, some sound correspondences between Sanskrit, Greek and Germanic languages do not match the expectations from Grimm's Law, as, for example, in the following cognates:

Sanskrit	*Greek*	*Gothic*	*English*
bōdha	peutha	biudan	bid 'to wake, become aware'
bandha		bindan	bind 'to bind'.

The first is from Proto-Indo-European **bheudha-*, the second from **bhendh-*; both have undergone dissimilation of the first **bh* due to the presence of a second aspirated stop in the word (**dh* in this case). This gives the SOUND CORRESPONDENCE in (1):

(1) Sanskrit *b* : Greek *p* : Gothic *b* : English *b.*

By Grimm's Law, we expect the *b* of Sanskrit to correspond to *p* in Germanic (Gothic and English in this case), and we expect Germanic *b* to correspond to

Sanskrit *bh* and Greek *ph*. So SOUND CORRESPONDENCE (1) in these cognate sets appears to be an exception to Grimm's Law. The cognate sets with correspondence (1) (and others for the originally aspirated stops at other points of articulation), then, are not real exceptions to Grimm's Law; rather, their reflexes in Germanic are correct for Grimm's Law, and the Sanskrit and Greek reflexes are not those expected by Grimm's Law only because Grassmann's Law regularly deaspirated the first aspirated stop when it occurred before another aspirated stop in the word in these languages. That is, SOUND CORRESPONDENCE (1) (and the others like it at other points of articulation) is the result of regular changes, Grimm's Law in Germanic, and Grassmann's Law in Sanskrit and Greek.

5.4.3 Verner's Law

A final set of what earlier had seemed to be exceptions to Grimm's Law is explained by Verner's Law (called *grammatical alternation* in older sources; see Chapter 2). Some forms which illustrate Verner's Law are seen in the cognate sets of Table 5.11 (OE = Old English; OHG = Old High German).

TABLE 5.11: Examples illustrating Verner's Law

	Sanskrit	*Greek*	*Latin*	*Gothic*	*English*
(1)	saptá	heptá	septem	sibun [siβun]	seven
(2)	pitár-	patā́r	pater	fadar [faðar]	OE fæder 'father'
(3)	śatám [ʃətəm]	(he-)katón	centum [kentum]	hunda (pl.)	hundred
(4)	śrutás 'heard'	klutós 'heard'			OE hlud 'loud'
(5)		makrós 'long, slender'	macer [maker]	[OHG magar]	meagre

In cognate set (1), by Grimm's Law we expect the *p* of Sanskrit, Greek and Latin to correspond to *f* in Germanic (Gothic and English), but instead we have Gothic *b* ([β]) and English *v*; given Gothic *b*, we expect the correspondence in Sanskrit to be *bh* and in Greek to be *ph*. Similarly, in cognate sets (2–4) we have the correspondence of Sanskrit, Greek and Latin *t* to Germanic *d*, not the *θ* expected by Grimm's Law in Germanic (and not the Sanskrit *dh* and Greek *th* we would expect, given Germanic *d*). These apparent exceptions to Grimm's Law are explained by Verner's Law. Verner's Law affects medial consonants; when the Proto-Indo-European accent followed, medial fricatives in a root – both original ones and those resulting from Grimm's Law – became voiced in Germanic, (*)f > β, (*)θ > ð, (*)x > γ, and *s > z. Since later in Proto-Germanic the accent shifted to the root-initial syllable, the earlier placement of the accent can only be seen when the cognates from the non-Germanic languages are compared. Thus, in the cognate sets of Table 5.11, we see in the Sanskrit and Greek cognates that the accent is not on the initial syllable but is on a later syllable, after the sound that changed, and that the Germanic forms do not match expectations

from Grimm's Law in these instances. In (1), we would not expect Gothic *sibun*, but rather something like *sifun*, given the *p* of Sanskrit *saptá* and Greek *heptá*; however, since the accent is on the last syllable in the Sanskrit and Greek forms, Verner's Law gives Gothic *β* (spelled *b*) in this case. The forms of Table 5.12 show how the forms with the accent later in the word (which undergo Verner's Law, symbolized as . . . C . . . ′) contrast with forms with the accent before the sound in question (indicated as ′. . . C. . .), cases which have undergone Grimm's Law), but where Verner's Law does not apply because they do not fit the environment for it.

TABLE 5.12: Examples showing the effects of Grimm's Law and further effects of Verner's Law on medial consonants in different contexts

Grimm's Law	*Verner's Law*
′. . . C. . .	. . . C . . . ′
*p > f	*p > f > β
(1a) OE hēafod 'head'	(1b) Gothic sibun [siβun] 'seven'
Latin cáput [káput]	Sanskrit saptá-
*t > θ	*t > θ > ð
(2a) Gothic brōþar [brōθar] 'brother'	(2b) OE fæder 'father'
Sanskrit bhrā́tar-	Sanskrit pitár-
*k > x	*k > x > ɣ
(3a) Gothic taíhun 'ten'	(3b) Gothic tigus 'decade'
Greek déka	Greek dekás

It is easy to see why Verner's Law was also often called 'grammatical alternation' (*grammatischer Wechsel* in German). The accent in Proto-Indo-European fell on different syllables in certain grammatically related forms, as seen in the forms compared in Table 5.13 (PIE = Proto-Indo-European; P-Germ = Proto-Germanic). As a result, Germanic languages have different allomorphs in grammatical paradigms which depend upon whether or not Verner's Law applied, and these grammatical alternations further support Verner's Law and its correlation with the place of the accent in the proto-language.

Just as expected by Grimm's Law, the Old English forms in the first two columns have /θ/ (spelled <þ>), where the accent in Proto-Indo-European preceded the original **t* (as illustrated by the Sanskrit forms). However, in the last two columns, Old English does not have the /θ/ expected by Grimm's Law, but the /d/ of Verner's law because the accent came after this medial **t* in Proto-Indo-European, again as shown by the Sanskrit forms. The Old High German forms subsequently underwent other sound changes of their own, but the difference between those with /d/ and those with /t/ has its origin in Verner's Law just as the alternations seen in the Old English cognates. The allomorphic variation which resulted, as for example that seen in the verb paradigm in Table 5.13, illustrates the 'grammatical alternation' that comes from Verner's Law.

TABLE 5.13: Verner's Law in grammatical alternations

	'I become'	*'I became'*	*'we became'*	*'became [participle]'*
PIE	*wértō	*(we)wórta	*(we)wr̥təmé	*wr̥tonós
Sanskrit	vártāmi	va-várta	vavr̥timá	vr̥tānáh
	'I turn'	'I have turned'	'we have turned'	'turned'
P-Germ	*werθō	*warθa	*wurðum(i)	*wurðan(a)z
OE	weorþe	warþ	wurdon	worden
OHG	wirdu	ward	wurtum	wortan

So, the Verner's Law cases (as in Tables 5.11, 5.12 and 5.13), which originally appeared to be exceptions to Grimm's Law, turn out also to be explained by regular sound change – by Verner's Law, a conditioned change having to do with the earlier location of the accent.

5.4.4 Indo-European sound laws and regularity of sound change

The laws just considered played an important role in the history of Indo-European studies and as a consequence in the overall history of historical linguistics. Grimm's Law, which was published first (in 1822), was quite general and accounted for the majority of sound correspondences involving the stop series between Germanic and non-Germanic languages. However, as initially formulated, it did appear to have exceptions. When Hermann Grassmann discovered his law (in 1862), a large block of these 'exceptions' was explained, and then Karl Verner through Verner's Law (in 1877) explained most of the remaining exceptions. This success in accounting for what had originally appeared to be exceptions led the Neogrammarians to the confidence that sound change was regular and exceptionless (see Chapter 2). This is one of the most significant conclusions in the history of linguistics.

5.5 Basic Assumptions of the Comparative Method

What textbooks call the 'basic assumptions' of the comparative method might better be viewed as the consequences of how we reconstruct and of our views of sound change. The following four basic assumptions are usually listed.

(1) The proto-language was uniform, with no dialect (or social) variation. Clearly this 'assumption' is counterfactual, since all known languages have regional or social variation, different styles, and so on. It is not so much that the comparative method 'assumes' no variation; rather, it is just that there is nothing built into the comparative method which would allow it to address variation directly. This means that what is reconstructed will not recover the once-spoken proto-language in its entirety. Still, rather than stressing what is missing, we can be happy that the method provides the means for recovering so much of the original language. This assumption of uniformity is a reason-

able idealization; it does no more damage to the understanding of the language than, say, modern reference grammars do which concentrate on a language's general structure, typically leaving out consideration of regional, social and stylistic variation. Moreover, dialect differences are not always left out of comparative considerations and reconstructions, since in some cases scholars do reconstruct dialect differences to the proto-language based on differences in daughter languages which are not easily reconciled with a single uniform starting point. This, however, has not been common practice outside of Indo-European studies.

Assumptions (2) and (3) are interrelated, so that it is best to discuss them together.

(2) Language splits are sudden.

(3) After the split-up of the proto-language, there is no subsequent contact among the related languages.

These 'assumptions' are a consequence of the fact that the comparative method addresses directly only material in the related languages which is inherited from the proto-language and has no means of its own for dealing with borrowings, the results of subsequent contact after diversification into related languages. Borrowing and the effects of subsequent language contact are, however, by no means neglected in reconstruction. Rather, we must resort to other techniques which are not formally part of the comparative method for dealing with borrowing and the results of language contact (see Chapters 3, 7 and 12). It is true that the comparative method contains no means for addressing whether the language of some speech community gradually diverged over a long period of time before ultimately distinct but related languages emerged, or whether a sudden division took place with a migration of a part of the community so far away that there was no subsequent contact between the two parts of the original community, resulting in a sharp split and no subsequent contacts between the groups. (Assumptions (2) and (3) are better seen as the consequence of the family-tree model for classifying related languages, dealt with in Chapters 6 and 7, since the tree diagram depicts a parent language splitting up sharply into its daughters.)

(4) Sound change is regular. The assumption of regularity is extremely valuable to the application of the comparative method. Knowing that a sound changes in a regular fashion gives us the confidence to reconstruct what the sound was like in the parent language from which it comes. If a sound could change in unconstrained, unpredictable ways, we would not be able to determine from a given sound in a daughter language what sound or sounds it may have come from in the parent language, or, looking at a particular sound in the parent language, we could not determine what its reflexes in its daughter languages would be. That is, if, for example, an original **p* of the proto-language could arbitrarily for no particular reason become *f* in some words, *y* in others, *q'* in others, and so on, in exactly the same phonetic and other linguistic circumstances, then it would not be possible to reconstruct. In such a situation, comparing, say a *p* of one language with a *p* of another related language would be of no avail, if the *p* in each could have come in an unpredictable manner from a number of different sounds.

5.6 How Realistic are Reconstructed Proto-languages?

The success of any given reconstruction depends on the material at hand to work with and the ability of the comparative linguist to figure out what happened in the history of the languages being compared. In cases where the daughter languages preserve clear evidence of what the parent language had, a reconstruction can be very successful, matching closely the actual spoken ancestral language from which the compared daughters descend. However, there are many cases in which all the daughter languages lose or merge formerly contrasting sounds or eliminate earlier alternations through analogy, or lose morphological categories due to changes of various sorts. We cannot recover things about the proto-language via the comparative method if the daughters simply do not preserve evidence of them. In cases where the evidence is severely limited or unclear, we often make mistakes. We make the best inferences we can based on the evidence available and on everything we know about the nature of human languages and linguistic change. We do the best we can with what we have to work with. Often the results are very good; sometimes they are less complete. In general, the longer in the past the proto-language split up, the more linguistic changes will have accumulated and the more difficult it becomes to reconstruct with full success.

A comparison of reconstructed Proto-Romance with attested Latin provides a telling example in this case. We do successfully recover a great deal of the formerly spoken language via the comparative method. However, the modern Romance languages for the most part preserve little of the former noun cases and complex tense–aspect verbal morphology which Latin had. Subsequent changes have obscured this inflectional morphology so much that much of it is not reconstructible by the comparative method.

5.7 Exercises

Exercise 5.1 Aimaran

Consider the following data from the two major branches of the Aimaran language family (Peru and Bolivia). Focus your attention on the sibilant fricatives (*s* and *š*) only (ignore *x* and χ for this exercise). What will you reconstruct? How many sibilant fricatives do you postulate for Proto-Aimaran? State your evidence.

NOTE: *š* = IPA [ʃ], č = IPA [tʃ]; χ = voiceless uvular fricative; *C'* = glottalized [ejective] consonants.

	Central Aimara	*Southern Aimara*	*gloss*
1.	saxu	sawu-	'to weave'
2.	sa(wi)	sa(ta)	'to plant'
3.	asa	asa-	'to carry flat things'
4.	usu	usu-	'to become sick'
5.	nasa	nasa	'nose'
6.	aski	hisk^{h}i	'to ask'
7.	muxsa	muχsa	'sweet'

	Central Aimara	*Southern Aimara*	*gloss*
8.	suniqi	sunaqi	'small spring'
9.	šanq'a	sanqa	'to snuffle'
10.	waša	wasa	'silent place'
11.	iši	isi	'dress'
12.	muši	musi	'to take care (of)'
13.	puši	pusi	'four'
14.	išt'a	hist'a-	'to close'
15.	išapa	isapa-	'to hear, listen'

(Cerrón-Palomino 2000: 145–6)

Exercise 5.2 Tulu

Tulu is a Dravidian language (of India) which has several varieties. Consider the following data from two principal varieties. Focus your attention only on the nasals. What will you reconstruct for these? How many nasals do you postulate for Proto-Tulu? State your evidence.
NOTE: *j* = [ǰ], IPA [dʒ]; *ṉ* = IPA [ɳ].

	Shivalli	*Sapaliga*	*gloss*
1.	a:ṉɨ	a:nɨ	'male'
2.	uṉɨ	u:nɨ	'dine'
3.	maṉṉɨ	mannɨ	'soil'
4.	ko:ṉɛ	ko:nɛ	'room'
5.	e:ṉɨ	ya:nɨ	'I'
6.	ninɛ	ninɛ	'wick'
7.	ja:nɛ	da:nɛ	'what'
8.	sanɛ	tanɛ	'conceiving'

(Bhat 2001: 11)

Exercise 5.3 Polynesian

The Polynesian languages of the Pacific form a subgroup of the Oceanic branch of the Austronesian family of languages. (1) What are the sound correspondences found in these data? What sound do you reconstruct for the proto-language to represent each sound correspondence set? (2) What sound change or changes have taken place in each of these languages? (3) What is the best reconstruction (proto-form) for 6, 16, 20 and 32? Show how your postulated sound changes apply to each of these to produce the modern forms. Note that not all sounds of the proto language are represented in these cognate sets with their sound correspondences. For example, in one not represented clearly here, Tongan has Ø corresponding to *l* or *r* of the other languages (reflecting what is usually reconstructed as **r* of Proto-Polynesian), distinct from the set in which Tongan has *l* corresponding to *l* or *r* in these sister languages (reflecting Proto-Polynesian **l*). This distinction may not be clearly visible in the data presented in this exercise.
NOTE: <'> = [ʔ].

	Māori	*Tongan*	*Samoan*	*Rarotongan*	*Hawai'ian*	*gloss*
1.	tapu	tapu	tapu	tapu	kapu	'forbidden', 'taboo'
2.	pito	pito	—	pito	piko	'navel'
3.	puhi	puhi	—	pu'i	puhi	'blow'
4.	taha	tafa 'edge'	tafa	ta'a	kaha	'side'
5.	tae 'trash'	ta'e	tae	tae	kae	'excrement'
6.	taŋata	taŋata	taŋata	taŋata	kanaka	'man, person'
7.	tai	tahi	tai	tai	kai	'sea'
8a.	kaha	kafa	'afa	ka'a	'aha	'strong'
8b.	ma:rohi-	ma:lohi	ma:losi	ma:ro'i	—	'strong'
9.	karo	kalo	'alo	karo	'alo	'dodge'
10.	aka	aka	a'a	aka	a'a	'root'
11.	au	'ahu	au	au	au	'gall'
12.	uru 'tip of weapon'	'ulu	ulu	uru	ulu 'centre'	'head'
13.	uhi	ufi	ufi	u'i	uhi	'yam'
14.	ahi	afi	afi	a'i	ahi	'fire'
15.	ɸa:	fa:	fa:	'a:	ha:	'four'
16.	ɸeke	feke	fe'e	'eke	he'e	'octopus'
17.	ika	ika	i'a	ika	i'a	'fish'
18.	ihu	ihu	isu	puta-i'u 'nostril' (puta 'hole')	ihu	'nose'
19.	hau 'wind' (hauku: 'dew' [-ku: 'showery weather'])	hau	sau	'au	hau	'dew'
20.	hika	—	si'a	'ika	hi'a	'firemaking'
21.	hiku 'fishtale'	hiku	si'u	'iku	hi'u	'tail'
22.	ake	hake	a'e	ake	a'e	'up'
23.	uru	—	ulu	uru	ulu	'enter'
24.	maŋa	maŋa	maŋa	maŋa	mana	'branch'
25.	mau 'fixed'	ma'u	mau	mau	mau	'constant'
26.	mara 'marinated'	—	mala	mara	mala	'fermented food'
27.	noho	nofo	nofo	no'o	noho	'sit'
28.	ŋaru	ŋalu	ŋalu	ŋaru	nalu	'wave'
29.	ŋutu	ŋutu	ŋutu	ŋutu	nuku	'mouth'
30.	waka	vaka	va'a	vaka	wa'a	'canoe'
31.	wae	va'e	vae	vae	wae	'leg'
32.	raho 'testicle'	laho	laso	ra'o	laho	'scrotum'
33.	rou 'long forked stick'	lohu	lou	rou	lou	'fruit-picking pale'

	Māori	*Tongan*	*Samoan*	*Rarotongan*	*Hawai'ian*	*gloss*
34.	rua	ua	lua	rua	lua	'two'

Exercise 5.4 Orokolo-Toaripi

Orokolo and Toaripi are two closely related Eleman languages (usually assigned to the Trans-New Guinea grouping, though this is as yet uncertain). Compare the data presented here and reconstruct Proto-Orokolo-Toaripi. (1) List the sound correspondences you find. (2) Give the proto-sounds you reconstruct to represent these. (3) Present the sound changes which you postulate that each language has undergone. (4) If there is any relative chronology involved among these changes, state what it is and the evidence for it. (5) Give your reconstruction of 12, 25 and 35 together with how the individual sound changes apply to these to produce the modern forms.

Many find this problem very difficult. When looking for conditioning factors that may distinguish overlapping correspondence sets, consider TOGETHER AT THE SAME TIME the environments BOTH before and after the sounds in question.

NOTE: for this problem, consider Orokolo *r* and *l* the same sound. Do not struggle over the difference between *ae* and *ai* in no. 38.

	Toaripi	*Orokolo*	*gloss*
1.	uti	uki	'bone'
2.	ete	eke	'vagina'
3.	tete	keke	'fish scales'
4.	tao	kao	'tooth'
5.	toare	koare	'senior'
6.	tola	kora	'tree'
7.	tolotolo	korokoro	'leaves'
8.	tapare	kapare	'grease'
9.	torea	korea	'theft'
10.	turuturu	kurukuru	'thundering'
11.	aite	aire	'after'
12.	kite	kile	'mat'
13.	lauta	laura	'flame tree'
14.	ita	ila	'pig'
15.	puta	pura	'cloth'
16.	uta	ura	'hole'
17.	fi	hi	'cry'
18.	firu	hiru	'portion'
19.	fe	he	'penis'

	Toaripi	*Orokolo*	*gloss*
20.	fere	here	'betel nut'
21.	fapai	hapa	'open'
22.	fave	have	'stone'
23.	forerai	horera	'appear'
24.	furi	huri	'pus'
25.	afutae	ahurae	'ashes'
26.	sisia	hihia	'sour'
27.	siri	hiri	'mildew'
28.	ase	ahe	'sugarcane'
29.	seseroro	heheroro	'thin'
30.	sare	hare	'sun, day'
31.	sarea	harea	'sorcery'
32.	soa	hoa	'time'
33.	sua	hua	'pigeon'
34.	susu	huhu	'plank'
35.	farisa	harita	'arrow'
36.	marisa	marita	'girl'
37.	taisa	kaita	'paddle'
38.	saesa	haita	'dish'
39.[=12]	kite	kile	'mat'
40.	kiva	kiva	'care'
41.	koko	koko	'narrow'
42.	ekaka	ekaka	'fish'

Exercise 5.5 Lencan

Compare the cognates from the two Lencan languages (both of which have recently become extinct: Chilanga was spoken in El Salvador; Honduran Lenca was spoken in Honduras). Work only with the consonants in this problem (the changes involving the vowels are too complex to solve with these data alone). (1) Set up the correspondence sets; (2) reconstruct the sounds of Proto-Lencan; (3) find and list the sound changes which took place in each language; and (4) determine what the relative chronology may have been in any cases where more than one change took place in either individual language, if there is evidence which shows this.

NOTE: *t'*, *k'* and *ts'* are glottalized consonants. Also, these data do not provide enough information for you to recover all the consonants of the proto-language, so that it will be difficult to apply steps 5 and 6 here.

	Honduran Lenca	*Chilanga*	*gloss*
1.	pe	pe	'two'
2.	lepa	lepa	'jaguar'
3.	puki	puka	'big'
4.	ta	ta	'cornfield'
5.	tem	tem	'louse'
6.	ke	ke	'stone'
7.	kuma	kumam	'fingernail, claw'
8.	katu	katu	'spider'
9.	waktik	watih	'sandals'
10.	kakma	k'ama	'gourd'
11.	siksik	sisih	'shrimp'
12.	nek	neh	'tooth'
13.	insek	ints'eh	'beak'
14.	taw	t'aw	'house'
15.	tutu	t'ut'u	'flea'
16.	kin	k'in	'road'
17.	kunan	k'ula	'who'
18.	kelkin	k'elkin	'tortilla griddle'
19.	sewe	ts'ewe	'monkey'
20.	saj	ts'aj	'five'
21.	musu	muts'u	'liver'
22.	sak-	ts'ih-	'to wash'
23.	lawa	lawa	'three'
24.	liwa-	liwa-	'to buy'
25.	tal-	tal-	'to drink'
26.	wala	wala	'raccoon'
27.	was	wal	'water'
28.	asa	alah	'head'
29.	wasan	wila	'urine'
30.	wara	wara	'river'
31.	siri	sirih	'star'
32.	sili	sili	'iron tree' (tree species)
33.	suri-sur	ʃurih	'squirrel'

[NOTE: *suri-sur* involves reduplications; just compare the *suri-* segment of it]

	Honduran Lenca	*Chilanga*	*gloss*
34.	saj-	ʃej-	‘to want’
35.	so	ʃo	‘rain’
36.	suna	ʃila	‘flower’
37.	soko	ʃoko	‘white’
38.	sak	ʃah	‘firewood’
39.	wewe	wewe	‘baby’
40.	jet-	jete-	‘to laugh’
41.	juku	juku	‘coyol palm’ (palm tree species)
42.	sa	ʃam	‘good’

Exercise 5.6 Uto-Aztecan

1. State the sound correspondences.
2. Present the sound that you reconstruct for Proto-Uto-Aztecan for each sound correspondence.
3. List the sound changes that you observe in the various languages.

Ignore vowel length and do not attempt to reconstruct the vowels for this exercise. There are not sufficient examples in the data given here to be able to reconstruct the full set of Proto-Uto-Aztecan sounds. Assume that correspondences found in only a single cognate set would recur if more data were present. Attempt to reconstruct the consonants in initial and medial position and state the sound changes you postulate to get from your reconstructed consonants to the actual forms in the various daughter languages. Given the paucity of forms cited here, you may have to postulate some sound changes on the basis of poor evidence, in hopes of confirming (or disconfirming) them when more data are brought into the picture. It is often the case that the historical linguist must work with incomplete or imperfect data, so the challenge here of attempting to reconstruct with less than complete information is a realistic experience.

NOTE: a few examples have been regularized, slightly modified, in order to avoid complications for the reconstruction.

	Cupeño	*Hopi*	*Comache*	*Tohono O’odham*	*Huichol*	*Nahuatl*	
1.	paqa	pa:qa-	paka ‘arrow’	wa:-pka	haka	a:ka	‘reed’
2.	pa-	pa:hɨ	pa:	wa-	ha:	a:	‘water’
3.	pah	pa:y-	pahi-	—	hai-	e:y	‘three’
4.	puš	po:si	pui	wuhi	hɨši	i:š	‘eye’
5.	pi-	pöhɨ	puʔe	wo:g	hu:ye:	oʔ	‘road’
6.	—	pɨtɨ	pɨhtɨ	we:č	he:te	ete-	‘heavy’
7.	tama	tama	ta:ma	ta:-tami	tame	tla:n	‘tooth’
8.	—	teni	—	čini	teni	te:n	‘mouth’
9.	tuu-, tula	tö:vɨ	(kuh-)tu:bi	ču:d	tɨ:	ti:l	‘charcoal’

	Cupeño	*Hopi*	*Comache*	*Tohono O'odham*	*Huichol*	*Nahuatl*	
10.	—	tös-	tus(oyuni) 'grindstone'	čuhi	tɨsi	tiš	'grind, flour'
11.	—	qa:si	—	kahio	—	ikši	'leg, thigh'
12.	qʷaše	kʷasi	kʷasɨ-pɨ	bahi	kʷaši	ikʷši	'cooked, ripe'
13.	qʷaš	—	kʷasi	bahi	kʷaši	—	'tail'
14.	qʷeʔ-	kʷɨʔɨ	—	baʔa	-kʷaʔa	kʷa	'eat'
15.	—	kʷita	kʷita-	bi:t	kʷita	kʷitla	'excrement'
16.	maqa	maqa	maka	ma:k	—	maka	'give'
17.	—	mo:ki	—	mu:ki	mɨki	miki	'kill, die'
18.	mala	mata	—	mač ˇud	ma:ta:	matla	'grindstone'
19.	naqa	na:qa	na:ki	na:k	naka	nakas	'ear'
20.	nema	nɨ:ma	nɨ:ma	nem	nema	—	'liver'
21.	waxe	la:ki	—	gaki	-waki	wa:ki	'dry'
22.	wexi-	lökö	woko	—	huku	oko	'pine'

(Based on Stubbs 2011)

Exercise 5.7 Jicaquean

Jicaquean is a family of two languages in Honduras. Jicaque (Jicaque of El Palmar) is extinct; Tol (Jicaque of La Montaña de la Flor) is still spoken by a few hundred people, but has become extinct or nearly so everywhere except in the village of La Montaña de la Flor. Reconstruct Proto-Jicaquean; state the sound correspondences which you encounter in the following cognate sets, and reconstruct a proto-sound for each. State the sound changes that have taken place in each language.

HINT: your reconstruction should include the following sounds:

p	t	ts	k	ʔ	i	ɨ	u
pʰ	tʰ	tsʰ	kʰ		e		o
p'	t'	ts'	k'			a	
		s					
		l					
m	n						
w		j		h			

What happens to each of the proto-sounds which you reconstruct in initial and in final position in these two languages? Can you make guesses about an appropriate reconstruction and sound changes to account for sounds in medial positions?

NOTE: the correspondences involving affricates and sibilants are quite complex, and you will need to pay special attention to the possibilities for combining some of the initial correspondence sets with some of the medial ones as reflecting the

same proto-sound. The consonants *p'*, *t'*, *ts'*, *k'* are glottalized. The accent mark on a vowel (for example *á*) means that it is stressed; this is not relevant to the sound changes. In a few cases, a non-initial *h* does not match well in the two languages; ignore this, since it is due to changes for which you do not have enough evidence in these data. The hyphen (-) before some words, as in 9 (*-rɨk*), means that these occur with some other morpheme before them which is not relevant and so is not presented here.

	Jicaque	*Tol*	*gloss*
1.	pe	pe	'stone'
2.	pit	pis	'meat'
3.	pɨné	pɨné	'big'
4.	pɨga-	pɨʔa-	'jaguar'
5.	pen	pel	'flea'
6.	kamba	kampa	'far, long'
7.	arba-	alpa	'above'
8.	to-bwe	to-pwe	'to burn'
9.	-rɨk	-lɨp	'lip'
10.	kek	kep	'woman'
11.	ik	hip	'you'
12.	huruk	hulup	'grain' (of corn)
13.	huk	hup	'he, that'
14.	nak	nap	'I'
15.	-kuk	-kup	'we'
16.	te	te	'black'
17.	tek	tek	'leg'
18.	tebé	tepé	'he died'
19.	tɨt	tɨt'	'louse'
20.	mandɨ	mantɨ	'vulture'
21.	n-gon	n-kol	'my belly'
22.	harek	halek	'arrow'
23.	mak	mak	'foreigner'
24.	n-abuk	n-ajp^{h}uk	'my head'
25.	kon	kom	'liver'
26.[=6]	kamba	kampa	'far, long'
27.	pɨrɨk	pɨlɨk	'much'
28.	keré	kelé	'nephew'
29.	mik	mik	'nose'
30.	korok	kolok	'spider'
31.	p^{h}e	p^{h}e	'white'

	Jicaque	*Tol*	*gloss*
32.	pʰen	pʰel	‘arm, shoulder’
33.	-pʰa	-pʰa	‘dry’
34.	pʰɨja	pʰɨja	‘tobacco’
35.	m-bat	m-pʰats’	‘my ear’
36.	lɨbɨ-	lɨpʰɨ	‘wind’
37.	pʰɨbɨh	pʰɨpʰɨh	‘ashes’
38.	urubana	(j)ulupʰana	‘four’
39.	ten	tʰem	‘boa constrictor’
40.	tut	tʰutʰ	‘spit’
41.	peten	petʰel	‘wasp’
42.	kun	kʰul	‘fish’
43.	ke-ke	(kʰ)ekʰe	‘agouti’

[NOTE: *keke* is a reduplicated form and should be treated as the root *ke-* repeated, rather than as having an intervocalic *-k-*]

44.	kan	kʰan	‘bed’
45.	kere	kʰele	‘bone’
46.	to-gon-	to-kʰol	‘to grind’
47.	kujuh	kʰujuh	‘parrot’
48.	pɨt	p’ɨs	‘deer’
49.	m-bɨj	m-p’ɨj	‘my body’
50.	pɨčá	p’ɨsá	‘macaw’
51.	-te	-t’e	‘to cut’
52.[=19]	tɨt	tɨt’	‘louse’
53.	-tja	-t’ja	‘to be late’
54.	mata	mat’a	‘two’
55.	kat	ʔas	‘blood’
56.	kot	ʔos	‘I sit, am’
57.	kaw-	ʔaw-a	‘fire’
58.	kona	ʔona	‘sour’
59.	kan	ʔan	‘zapote’ (fruit)
60.[=4]	pɨga-	pɨʔa-	‘jaguar’
61.	te-ga	te-ʔa	‘to give’
62.	čok	sok’	‘tail’
63.	čorin	tsolin	‘salt’
64.	ču(h)	tsu	‘blue’
65.	čiwiri	-tsiwil-	‘to lie’
66.	čigin-	tsikin	‘summer’
67.	čoʔ-	tsoʔ-	‘to nurse’

	Jicaque	*Tol*	*gloss*
68.	čuba	tsupa	'to tie'
69.	nočot	notsots	'fly'
70.	ʃeme	ts^{h}eme	'horn'
71.	ʃijó	ts^{h}ijó	'dog'
72.	ʃe(w)	ts^{h}ew	'scorpion'
73.	čin	ts'il	'hair, root'
74.	-čun	ts'ul	'intestines'
75.	čoron	ts'olol	'oak'
76.	čih	ts'ih-	'caterpillar'
77.	te-neče	te-nets'e	'to sing'
78.	ločak	lots'ak	'sun'
79.	m-bat	m-p^{h}ats'	'my ear'
80.	čot	sots'	'owl'
81.	-čɨ	-sɨ	'water'
82.	čok	sok'	'tail'
83.[=2]	pit	pis	'meat'
84.	-mut	mus	'smoke'
85.	hoč(uruk)	hos-	'his heart'
86.[=50]	pɨčá	p'ɨsá	'macaw'
87.	mon	mol	'cloud'
88.[=25]	kon	kom	'liver'
89.	ma	ma	'land'
90.	wa	wa	'house'
91.	wara	wala	'forehead'
92.	jo	jo	'tree'
93.	he	he	'red'
94.[=22]	harek	halek	'arrow'

(Data from Campbell and Oltrogge 1980)

Exercise 5.8 K'ichean languages

K'ichean is a subgroup of the Mayan family. Compare these cognate forms and set up the sound correspondences; propose the most appropriate reconstruction for the sound in the proto-language for each, and write the sound changes which account for the developments in the daughter languages. Are any instances found in any of the individual languages in which it is necessary to state what the relative chronology of changes was?

NOTE: ɓ = voiced imploded bilabial stop; t', ts', č', k', q', m', w' = glottalized consonants. In Uspanteko, the accent mark over the vowel, as in *ò:x* 'avocado', indicates falling tone. Although the correspondence set in which Q'eqchi' *h*

corresponds to *x* of the other languages is not found in these data before *u*, ignore this – this correspondence occurs in general with no restrictions that have anything to do with *u*.

NOTE: *y* = [j], *š* = IPA [ʃ], č = IPA [tʃ] *C'* = glottalized [ejective] consonants.

	Kaqchikel	*Tz'utujil*	*K'iche'*	*Poqomam*	*Uspanteko*	*Q'eqchi'*	*gloss*
1.	pak	pak	pak	pak	pak	pak	'custard apple'
2.	pur	pur	pur	pur	pur	pur	'snail'
3.	pim	pim	pim	pim	pim	pim	'thick'
4.	toʔ	toʔ	toʔ	toʔ	toʔ	toʔ	'to help'
5.	tox	tox	tox	tox	tox	tox	'to pay'
6.	kiʔ	kiʔ	kiʔ	kiʔ	kiʔ	kiʔ	'sweet'
7.	ka:ʔ	ka:ʔ	ka:ʔ	ka:ʔ	ka:ʔ	ka:ʔ	'quern'(metate)
8.	k'el	k'el	k'el	k'el	k'el	(k'el)	'parrot'
9.	qa-	qa-	qa-	qa-	qa-	qa-	'our'
10.	qul	qul	qul	—	qul	—	'neck'
11.	q'o:l	q'ol	q'o:l	q'o:l	q'o:l	q'o:l	'resin, pitch'
12.	q'an	q'an	q'an	q'an	q'an	q'an	'yellow'
13.	si:p	si:p	si:p	si:p	si:p	si:p	'tick'
14.	saq	saq	saq	saq	saq	saq	'white'
15.	tsuy	tsuy	tsuh	suh	tsuh	suh	'water gourd'
16.	uts	uts	uts	us	uts	us	'good'
17.	tsats	tsats	tsats	sas	tsats	sas	'thick'
18.	ts'iʔ	ts'iʔ	ts'iʔ	ts'iʔ	ts'iʔ	'ts'iʔ'	'dog'
19.	če:ʔ	če:ʔ	če:ʔ	če:ʔ	če:ʔ	če:ʔ	'tree, wood'
20.	ču:n	ču:n	ču:n	ču:n	ču:n	ču:n	'lime'
21.	č'o:p	č'o:p	č'o:p	č'o:p	č'o:p	č'o:p	'pineapple'
22.	xul	xul	xul	xul	xul	xul	'hole, cave'
23.	winaq	winaq	winaq	winaq	winaq	kwinq	'person'
24.	we:š	we:š	we:š	we:š	—	kwe:š	'trousers'
25.	ya:x	ya:x	ya:x	ya:x	ya:x	ya:x	'genitals, shame'
26.	mu:x	mu:x	mu:x	mu:x	mù:x	mu:h	'shade'
27.	o:x	o:x	o:x	o:x	ò:x	o:h	'avocado'
28.	ča:x	ča:x	ča:x	ča:x	čà:x	ča:h	'ashes'
29.	tu:x	tu:x	tu:x	tu:x	tù:x	tu:h	'steambath'
30.	q'i:x	q'i:x	q'i:x	q'i:x	q'ì:x	(-q'ih)	'day, sun'
31.	ka:x	ka:x	ka:x	ka:x	kà:x	—	'sky'
32.	čax	čax	čax	čax	čax	čax	'pine'
33.	k'ax	k'ax	k'ax	k'ax	k'ax	k'ax	'flour'
34.	k'o:x	k'o:x	k'o:x	k'o:x	k'o:x	k'o:x	'mask'
35.	ɓa:y	ɓa:y	ɓa:h	w'a:y	ɓa:h	ɓa:h	'gopher'

	Kaqchikel	*Tz'utujil*	*K'iche'*	*Poqomam*	*Uspanteko*	*Q'eqchi'*	*gloss*
36.	ɓa:q	ɓa:q	ɓa:q	w'a:q	ɓaq	ɓaq	'bone'
37.	ɓe:y	ɓe:y	ɓe:h	w'e:h	ɓe:h	ɓe:h	'road'
38.	siɓ	siɓ	siɓ	sim'	siɓ	siɓ	'smoke'
39.	xaɓ	xaɓ	xaɓ	xam'	xaŋ	haŋ	'rain'
40.	xuku:ʔ	xuku:ʔ	xuku:ɓ	xuku:m'	xuku:ɓ	xukuɓ	'canoe, trough'
41.	a:q'aʔ	a:q'aʔ	a:q'aɓ	a:q'am'	a:q'aɓ	(a:q'ɓ)	'night'
42.	xal	xal	xal	xal	xal	hal	'ear of corn'
43.	xe:y	xe:y	xe:h	xe:h	xe:h	he:h	'tail'
44.	č'o:y	č'o:y	č'o:h	č'o:h	č'o:h	č'o:h	'mouse, rat'
45.	k'yaq	k'yaq	k'yaq	k'aq	k'aq	k'aq	'flea'
46.	kyaq	kyaq	kyaq	kaq	kaq	kaq	'red'
47.	(i)kyaq'	(i)kyaq'	kyaq'	kaq'	—	—	'guava'
48.	išk'yaq	šk'yaq	išk'yaq	išk'aq	išk'aq	—	'fingernail'
49.	winaq	winaq	winaq	winaq	winaq	kwinq	'person'
50.	šikin	šikin	šikin	šikin	šikin	(šikn)	'ear'
51.	išoq	išoq	išoq	išoq	—	išq	'woman'
52.	nimaq	nimaq	nimaq	nimaq	nimaq	ninq	'big' (plural)
53.	sanik	sanik	sanik	(sanik)	sanik	sank	'ant'
54.	suʔt	suʔt	suʔt	suʔt	sù:t'	(suʔut)	'cloth, kerchief'
55.	poʔt	poʔt	poʔt	poʔt	pò:t'	poʔot	'blouse'
56.	piʔq	piʔq	piʔq	piʔq	pì:q'	—	'corncob'
57.	atiʔt	atiʔt	atiʔt	atiʔt	atì:t'	atiʔt	'grandmother'
58.	k'ax	k'ax	k'ax	k'ax	k'ax	k'ax	'flour'
59.	k'ay	k'ay	k'ah	k'ah	k'ah	k'ah	'bitter'
60.	k'ay	k'ay	k'ay	k'ay	k'ay	k'ay	'to sell'
61.	mo:y	mo:y	mo:y	mo:y	mo:y	mo:y	'blind' (dark)
62.	ča:x	ča:x	ča:x	ča:x	čà:x	ča:h	'ashes'
63.	čax	čax	čax	čax	čax	čax	'pine'
64.	č'ax	č'ax	č'ax	—	č'ax	č'ax	'to wash'
65.	č'ay	č'ay	č'ay	č'ay	—	—	'to hit'

Exercise 5.9 Quechuan

Quechuan is a family of several languages spoken in the Andes region of South America, with varieties found in Columbia, Ecuador, Peru, Bolivia and Argentina.

Compare the cognates from the languages listed here. Set up the correspondence sets; reconstruct the sounds of Proto-Quechuan; find and list the sound changes which took place in each language (variety); determine what the relative chronology may have been in any cases where more than one change took place in an individual language (variety), if there is evidence which shows this. What do you think the inventory of Proto-Quechuan sounds was? (Note that there is

some controversy about the historical status of glottalized consonants (p', t', č', k', q') and aspirated consonants (p^h, t^h, $č^h$, k^h, q^h) in Quechuan. For the purposes of this exercise do not try to reconstruct them, but rather treat those few which occur (in the Cuzco variety) as though they were equal to the plain counterparts.) (NOTE: [ɳ]= uvular nasal; *y* = IPA [j]; l^y = IPA [l^j]; n^y = IPA [n^j].)

	Ancash	*Junín*	*Cajamarca*	*Amazonas*	*Ecuador*	*Ayacucho*	*Cuzco*	*gloss*
1.	paka-	paka-	paka-	paka-	paka-	paka-	paka-	'begin'
2.	apa-	apa-	apa-	apa-	apa-	apa-	apa-	'wash'
3.	rapra	lapla	rapra	rapra	—	rapra	raɸra	'leaf, wing'
4.	pampa	pampa	pamba	pamba	pamba	pampa	pampa	'plains'
5.	tapu-	tapu-	tapu-	tapu-	tapu-	tapu-	tapu-	'ask'
6.	wata-	wata-	wata-	wata-	wata-	wata-	wata-	'tie'
7.	utka	utka	utka	utka	—	utka	usk^{h}a	'cotton'
8.	inti	inti	indi	indi	indi	inti	inti	'sun'
9.	kimsa	kimsa	kimsa	kimsa	kimsa	kimsa	kimsa	'three'
10.	puka	puka	—	puka	puka	puka	puka	'red'
11.	haksa-	saksa-	saksa-	saxsa-	saxsa-	saksa-	saxsa-	'be full, fed up'
12.	kuŋka	kuŋka	kuŋga	kuŋga	kuŋga	kuŋka	kuŋka	'neck'
13.	qam	am	qam	kam	kaŋ	χam	qaŋ	'you' (sg.)
14.	qoha	usa	qosa	kusa	kusa	χosa	qosa	'husband'
15.	waʁa-	waʔa-	waʁa-	waka-	waka-	waχa-	waqa-	'cry'
16.	hoχta	suʔta	soχta	sukta	suxta	soχta	soχta	'six'
17.	heŋʁa	siŋʔa	seŋʁa	siŋga	siŋga	seŋχa	seŋqa	'nose'
18.	tsaki	čaki	čaki	čaki	čaki	čaki	č'aki	'dry'
19.	mutsa-	muča-	muča-	muča-	muča-	muča-	muč'a-	'kiss'
20.	mantsa-	manča-	manča-	manča-	manča-	manča-	manča-	'fear, be afraid'
21.	putska-	pučka-	pučka-	pučka-	puʃka-	pučka-	puska-	'to thread'
22.	e:tsa	ayča	ayča	e:ča	ayča	ayča	ayča	'meat'
23.	čaki	čaki	čaki	čaki	čaki	čaki	čaki	'foot'
24.	kača-	kača-	kača-	kača-	kača-	kača-	kača-	'send'
25.	učpa	učpa	učpa	učpa	uʃpa	učpa	usp^{h}a	'ashes'
26.	kički	kički	kički	kički	kiʃki	kički	k'iski	'narrow'
27.	haru-	salu-	saru-	saru-	saru-	saru-	saru-	'to step on'
28.	hara	sala	sara	sara	sara	sara	sara	'maize, corn'
29.	qaha	asa	qasa	kasa	kasa	χasa	qasa	'ice'

	Ancash	*Junín*	*Cajamarca*	*Amazonas*	*Ecuador*	*Ayacucho*	*Cuzco*	*gloss*
30.	isqoŋ	isʔuŋ	esqoŋ	iʃkuŋ	iʃkuŋ	isχoŋ	esqoŋ	‘nine’
31.	—	aysa-	aysa-	e:sa-	aysa-	aysa-	aysa-	‘pull’
32.	waʃa	waʃa	waʃa	waʃa	waʃa	wasa	wasa	‘behind’
33.	iʃke:	iʃkay	iʃkay	iʃke:	iʃkay	iskay	iskay	‘two’
34.	hatuŋ	hatuŋ	atuŋ	atuŋ	hatuŋ	hatuŋ	hatuŋ	‘big’
55.	hutsa	huča	učа	učа	huča	huča	huča	‘fault’
36.	humpi	humpi	—	umbi	humbi	humpi	hump’i	‘sweat’
37.	laki	lʸaki	ʒaki	ǰaki	ʒaki	lʸaki	lʸaki	‘pain, trouble’
38.	kila	kilʸa	kiʒa	kiǰa	kiʒa	kilʸa	kilʸa	‘moon’
39.	alba	alʸpa	aʃpa	aǰpa	aʒpa	alʸpa	halʸp’a	‘land’
40.	aylu	aylʸu	ayʒu	e:ǰu	ayʒu	aylʸju	aylʸu	‘family’
41	rima-	lima-	rima-	rima-	rima-	rima-	rima-	‘to speak’
42.	karu	kalu	karu	karu	karu	karu	karu	‘far’
43.	warmi	walmi	warmi	warmi	warmi	warmi	warmi	‘woman’
44.	waχra	waʔla	waχra	wakra	—	waχra	waχra	‘horn’
45.	nina	nina	nina	nina	nina	nina	nina	‘fire’
46.	yana	yana	yana	yana	yana	yana	yana	‘black’
47.	wayna	wayna	wayna	wayna	wayna	wayna	wayna	‘young man’
48.	aŋya-	aŋya-	aŋya-	aŋya-	aŋya-	aŋya	aŋya-	‘to reprove’
49.	nawi	nʸawi	nʸawi	nʸawi	nʸawi	nʸawi	nʸawi	‘eye’
50.	wanu-	wanʸu-	wanʸu-	wanʸu-	wanʸu-	wanʸu-	wanʸu-	‘to die’
51.	qepa	ipa	qepa	kipa	kipa	χepa	qʰepa	‘behind’
52.	weɢe	wiʔi	—	wiki	wiki	weχe	weqe	‘tear (drop)’ (noun)
53.	qeʃpi-	iʃpi-	—	kiʃpi-	kiʃpi-	χespi-	qespi-	‘to escape’
54.	qo-	u-	qo-	—	ku-	χo-	qo-	‘to give’
55.	qoɳɢa-	uŋʔa-	qoɳɢa-	kuŋga-	kuŋga-	χoɳχa-	qoŋqa-	‘to forget’