Note that there is a recorded version of two lectures based on the reading material below to help you get to grips with the basics of historical linguistics. You can access the recorded material using this link.

Linguistic Theory ANGD-A2 - Attila Starcevic

Data and Historical Linguistics

Ȝe [you] knowe ek [also] that in [the] fourme [form] of speche [speech] is chaunge [change]
With-inne a thousand ȝeer [years], and wordes tho [then]
That hadden pris [value] now wonder [wonderfully] nyce [stupid] and straunge [strange]
Us thenketh hem [we think of them], and ȝet [yet] thei spake [spoke] hem [them] so,
And spedde [managed] as wel in loue [love] as men now do.

(Geoffrey Chaucer [1340—1400],
Troilus and Criseyde, book II, line 22—26)

Rough translation
You (plural) also know that there is change in the form of speech
within a thousand years, and words that had value then now wonderfully
stupid and strange seem to us, and yet they spoke them so,
and managed equally well in love as men now do.

1. Introduction

History and all things historical have always fascinated us humans. Chaucer, who spoke Middle English and was not a historian by profession, not a linguist and certainly not a historical linguist, was not the first one to notice that languages change over time. In the short passage above Chaucer makes the right observation that the meaning of words changes over time, and back in the time when they were used they had a meaning that now may seem obsolete and strange to us. Yet, they served the same purpose that words have always done: communication. Communication never breaks up between successive generations of speakers of the same language, yet the language itself is continuously changing.

Of course, not all the changes are semantic in nature (cf. *nyce* ‘nice’ which has acquired a positive meaning compared to the original meaning of ‘stupid, simpleton’). Some words may completely disappear, known as lexical loss (*e.g., ek* ‘also’, and it is only through the work of etymologists and their etymological dictionaries that we know the word is still to be found in *nickname*, meaning ‘also-name’ literally, from *an ek name,* reanalysed morphologically as *a nick-name*). The word was replaced by *also,* which existed in Chaucer’s time but meant something along the lines of its German counterpart *also* ‘so, in this way’.

There is a gamut of further changes as well: lexical (*hem* was replaced by *them*, *ye* by *you*), morphological (in *hadden* the suffix *-en* shows plural inflection, which was subsequently lost; *wonder* as an adjective could also be used as an adverb, not requiring *-ly*), grammatical (i.e., syntactico-morphological change in e.g., *Us thenketh hem* 'they appear/seem to us' or 'we think of them as' in which both the plural subject and the plural object appear to be in the accusative case (*us, hem*) and the finite verb is marked for 3 person singular present agreement (*-eth*, which itself was later replaced by *-(e)s*)). There are also phonological changes that are not reflected by modern spelling (*so* as **sɔ**ː in Middle English), or changes in spelling that are irrelevant linguistically (e.g., *loue* 'love' in which the <u> could not have been anything else than **v**). There is also a case of analogical levelling (the vowel of *spoken* was levelled into the past tense *spake* resulting in *spoke*). Some changes are difficult to interpret given the fact that we only have spelling as a source of immediate information (e.g., *pris* which is now *price* or *chaunge* which is found as *change*, *ȝet* as *yet*). As always, some changes are less radical than changes in spelling will have us believe (*ȝet* was **jɛt** in Chaucer's time, showing that the spelling change from *ȝ* (known as 'yogh') to *y* is nothing more than a red herring: *yet* has had the same pronunciation for hundreds, if not thousands, of years).

The reverse is also true: some changes in spelling may not seem radical but do in fact disguise a change: we already know that the yogh is **j** in *ȝeer* 'year'. But what about the vowel? Analysis points in the direction of the vowel having been **ɛ**ː (a front mid-low vowel) in Chaucer's time. The vowel was later spelled as <ea>, showing that it was an *e*-type of sound (an 'open *e*', note the use of <a> in the <ea> digraph) different to that spelled <ee> (a 'close *e*', as in *deer*). Subsequent changes have resulted in **jɪ**ː) in (some varieties of) British English. So a mid-low vowel is now a mid-high vowel phonetically, still spelled as if it was a mid-low vowel (<ea>). If we had no speakers today, we could still be thinking this is an open-*e* sound. Spelling should always be taken with a note of caution.

All in all, this short passage shows some of the possible changes that a historical linguist is confronted with when analysing past stages of the same language or different stages of different languages (known as comparative linguistics). We may find all these changes intriguing, puzzling, strange, bewildering, even stupid in Chaucer's understanding of the notion, but all these offer a window on the past stages of English and have provided an engaging mental exercise for those who are into historical linguistics.

2. What Historical Linguistics is Not About

Before we continue to see what historical linguistics is about, we must see what it *isn't* about. Historical linguistics is not about the *history of linguistics* although it has played an important part in the development of linguistics. If you ever read a chapter on the history of linguistics, the comparative method (described in a later section of this chapter) is heralded as one of the most important achievements of historical linguistics.

Another area that historical linguistics is not concerned with is the *origin of human language*. This may seem controversial at first sight: if historical linguistics is about the history of languages, then the first stage of every language is its development arising from gestures, primate call systems, etc. This is a misunderstanding. Think of a similar situation coming from physics this time. Whatever we think we know about the universe around us is based on our understanding of the workings of the universe, cast in terms of rules, formulas and equations involving mass, speed of particles, etc. The rules of physics as we know them have existed from the time Big Bang happened a few billion years ago. The rules have remained immutable since this moment, the moment when the universe began expanding. However, the rules of physics must have been fundamentally different at a time *before* the expansion of mass occurred. There exist speculations about these rules, but this is not what physics in general is concerned with. The same situation holds for linguistics in general (and historical linguistics in particular): we can speculate about what a system of gestures and calls may have been like in the stage before 'coagulating into' languages, but this is beyond the reach of (historical) linguistics. Linguistics deals with languages that are fully formed, i.e. languages with a system of phonemes, allophones, morphemes, phrase structure rules, etc. As a matter of fact, no language has been found that is not fully formed, stuck between primate calls and a complete system of interacting modules making up the system of a human language. In other words, historical linguistics has nothing to say about the stage of languages before becoming a language. No matter how deep one digs in the history of languages, they will always be fully formed languages. No historical linguist is likely to say that 'mmmmm' (an inarticulate roar accompanied by a forceful movement of the fist aimed at the skies) developed into the 3 personal singular present morpheme *-b* in a given language.

Many lament over changes that they observe in their language. Some think that *What're doing guys?* said to a mixed company is borderline offensive because *guys* can only refer to the male members of the company. Those who think this seriously will go as far as to claim that changes in language in general are a result of laziness, lack of education and decorum,

slovenliness, etc. Changes are almost always seen as a sign of decay, corruption, even threat to national identity and security. In historical linguistics, as in any 'serious' science for that matter, questions related to (national) perception of change are both futile and meaningless and make as much sense as asking a biologist if a pigeon, a hawk or a parrot show signs of corruption and threat just because they happen to have a heritage linking them to the archyopterix (a feathered dinosaur) but are no longer dinosaurs. The question would be meaningless and ridiculous.

Needless to say, even biologists can be overheard to say that they hate pigeons and would have them banned from city centres because they despoil statues and monuments of their intended beauty by defecation. But this would not be said by a biologist-scientist, but by a biologist who decided to step outside the domain of science. The same is true of linguists. No serious linguist will ever say that change is to be despised or deprecated. Passing sweeping moral judgments on change is never part of the job description of a scientist. If you come to think about it and look at the changes we listed in Section 1, and still entertain the idea that bygone eras of language are pure, uncorrupted and close to the origin, we would be forced to say that we all speak a debased version of Chaucer's Middle English.

In conclusion, change is inevitable and because changes do not happen haphazardly, this makes them a worthwhile mental exercise (a science in other words).

3. What Historical Linguistics is About

Now it is time to see what historical linguistics *is* really about. Languages can be analysed in two (some would say complementary) ways. One approach is the *diachronic* (from Greek *dia* 'across' and *chronos* 'time'), the main concern lies with change in language or languages over time. What we have done in Section 1 shows some of the changes between Middle English and Present Day English. As we see, Middle English is very different from modern English.

Synchronic linguistics (from Greek *syn* 'together' and *chronos* 'time') deals with language (of a speech community) as it is (or was) spoken at a particular point in time with no regard to the history of the language. So we can have a synchronic description of Old English or Middle English in the same way as we may of modern English: we can have a grammar of any one of these periods of English with a complete (or almost complete) description of its system of phonemes (even sounds if we are lucky enough to know how to interpret the written records that survive from that particular period), phrase structure rules, etc. Synchronic linguistics is independent of diachronic linguistics: when we devise a grammar of Middle

English, reference to earlier or later periods of English can only serve the purpose of devising a better grammar of Middle English. Old English grammar is not part of Middle English grammar in the same way of Middle English is not part of modern English grammar: the fact that the Middle English 3 person singular present indicative suffix was *-th* is interesting in itself, but of little use in saying anything linguistically relevant about modern English *-s* (as in *sees, makes*, etc.), apart from the obvious fact that both Middle and modern English have a suffix for the 3 person singular present indicative finite form. To weigh the importance of modern English *-s* we need to look at how it compares with the rest of the grammar of modern English (not Middle English). Diachronic facts may enrich our understanding of any one period of any language, but they are not strictly speaking part of the synchronic state of that particular language.

However, in a diachronic grammar of English historical linguists will point out that Middle English *-eth* has nothing to do with modern English *-s* (the suffix *-eth* was lost in modern English and substituted with *-s*, a suffix that was used interchangeably with *-eth* in Early Modern English).

Some identify historical linguistics with *etymology* (from Greek *etymon* 'origin'), the study of the history of words (their shape and meaning). Etymology is never an end in itself; etymology is, so to speak, the by-product of historical analysis. Take, for example, *fine* 'adequate, acceptable' and *fine* 'amount of money paid to settle a dispute'. From a synchronic point of view the two words are homonymous: $fine_1$ and $fine_2$ have nothing to do with each other (the meanings do not seem to share a common core meaning, i.e., this is not a case of polysemy, as in the case of *bulb* which means both 'underground stem' and 'a light bulb'). If we take a look at an etymological dictionary, we will see that *fine*, too, was a polysemous word: it comes from Old French *fin* **fi**:**n**, which itself comes from Latin *finis* and means 'completion, settlement' and *finalis* 'final, complete'. We can now see that $fine_1$ and $fine_2$ was originally *fine*, a polysemous word: *fine* meant both 'complete, settled, good' and 'money paid to achieve settlement of a dispute'. In addition to this, the modern English pronunciation of *fine* **fɑjn** from Middle English **fi**:**n** is perfectly in line with all the other words that contained **i**: in Middle English: **i**: (ultimately) diphthongised to **ɑj** (cf. *nice, vice, mine*). So, the etymology of *fine* is not an end in and for itself. It is part of a bigger diachronic picture: polysemous words becoming homonymous is a natural consequence of linguistic change, so is the diphthongisation of **i**: to **ɑj** in the transition between Middle and (Early) Modern English. The etymological dictionary

entry for *fine* is not about *fine* in itself, the entry is about semantic change and phonological change exemplified using the word(s) *fine*.

So, what is historical linguistics all about? If a historical linguist compares languages and comes to the conclusion that language X is related to language Y this will mean that the two languages share a common ancestor, i.e. that they both descend from a common language (which is no longer spoken). In other words, languages X and Y are (historically viewed) dialects of this now extinct language: they are *genetically related*. This choice of words from biology (genetically related) may be unfortunate, but is deeply rooted in historical linguistics. Genetically related simply means that language X and Y share the same sounds (and their later descendants through phonological change), common vocabulary (basic words for parts of the body, bodily functions, lower numbers, etc.), common morphemes (e.g., plural suffixes, finite inflections, etc.). Language X and Y can cover any two (or by extension more) languages: Southern Standard British English, Northern English, Scottish English are all related through (one of the dialects) of Middle (and Old) English. Similarly related are French and Italian (they derive from a common language known as Vulgar Latin). As you can see by comparing languages we can find out about their 'pedigree', their descent from a common source. We will see what this means in more detail below.

4. Data and Evidence

In deciding what counts as evidence in historical linguistics we must tread carefully. The age old maxim is still valid here: not everything is gold that glitters. The opposite also seems to be true (and is probably even more intriguing for a linguist): there may be data that do not glitter but will be found worthy of analysis.

The limits of knowledge can be discussed at length. Some of the aspects are philosophical and need not concern us. The path to knowledge leads from gathering data to extracting evidence through a chosen method of analysis. The choice of comparative method that traditional historical linguistics embraces was heralded as one of the great intellectual achievements of the 19th century and its meticulous application to data (similarly to the natural sciences) finally gave linguistics a strong foothold and a chance to push the comparative method to its limits. What we know about language families and the relationship between languages today we owe it to the comparative method of reconstruction. Although it seems irrelevant and difficult to grasp at this stage, we must understand that this method gave linguists the confidence

to say Yes or No to a question: issues could (and still can) be decided not on a whim of fancy but by following a succession of steps leading from data to evidence to conclusion.

For all intents and purposes data extracted from living languages are abundant: a simple google search for any item of vocabulary of any language may return more than a billion hits. The question, of course, is what this shows. For those who speak English and French it will come as no surprise that out of 10 randomly picked words 7 or 8 will show up in both languages. If we only look at the spelling of these words, the differences tend to be minimal (cf. *relation*, spelt identically, as opposed to *connection* vs *connexion* with a minimal difference in spelling). We know that spelling as such is not of primary concern for linguistics. If we look at the pronunciation of French *relation* **rəlasjõ** and compare it to English *relation* **rɪjlɛ́jʃən** we will see more than just an accidental difference between the two languages: word-final <n> in French shows nasalisation of the preceding vowel (compared to this English still preserves the word-final consonant and as such seems more traditional), where French has the sequence **sj** English has **ʃ** (which is a palatal consonant and is usually the result of a change known as palatalization, the influence of **j** (or the palatal/front vowels) on the preceding consonant. In this respect, it is French that appears more conservative.), etc. These differences are not confined to these two words alone and typologically they are very common to languages. In addition to these correspondences there are many others that the spelling disguises. Luckily, the two languages are still spoken, so we have direct access to pronunciation: e.g., French *coin* **kwɛ̃** 'corner' and English *coin* **kɔjn** are spelt identically, yet the more substantial difference in pronunciation makes the chase for relatedness more intriguing: perhaps the differences between these two words show a deeper level of relatedness (a more ancient one given the greater difference in pronunciation - this would be a good example for linguistic gold that does not glitter). We may hypothesise that the word now spelt *coin* was part of both languages before they parted ways and the two words went they separate ways. We would effectively be claiming that English and French are two dialects of the same ancient language. This is theoretically possible, of course, but not likely in the case of English and French given all the rest of the evidence against this.

The two languages differ, but change in itself is of little interest to the linguist (languages *do* differ after all, and with time the differences increase). Up to this point we have only been able to assert that English and French differ systematically, but nothing more: change as such is part and parcel of every language. This fact about change in itself is unedifying. The question still is what these data show and whether there is evidence that English and French are related.

The suspicion is that these data show nothing of relevance apart from the fact that one of the languages has borrowed from the other. Indeed, there are very few languages that have not adopted a word or two from English, or French for that matter: cf. Hungarian *lízing* (from English *leasing*), *csencsel* 'wheel-deal' (from *change*), *garzon* 'bedsit' (from French *garçon* 'boy'), *szalon* (*salon*), *bonbon* (*bonbon*), only to mention a few (there are other, perhaps less 'canonised' words form the realm of chatting, texting, forums, speed dating, etc.: e.g., *lájkol* (English *like*), *becsekkol* (English *check in*)). If you inspect your own language, it will not be long before you find words that were borrowed at some time from English or French or from any other language in close geographical proximity to yours. Today English is one of the great 'exporters' of words for a number of reasons.

One objection you may raise is that words that English and French share outnumber those shared by English and Hungarian or English and Turkish, for example. This may be true statistically, but then the question remains what these figures show. They may still only show that English has borrowed extensively from French or that English and French are geographically closer than French and Turkish. What is important is to understand that borrowing in itself is no reason to suppose that languages are genetically related. That is, English and French, despite their superficial similarities, are simply too close to each other (both geographically and historically) for borrowing *not* to occur. This is where the similarities end. This relationship between languages based on geographical proximity (or factors other than geographical: e.g., perceived or superimposed cultural supremacy of a culture and its language) is known as *areal relatedness*. If two languages are areally related they need not be genetically related.

Let us see another example. In the Slavonic languages perfect aspect is expressed with verbal prefixes. Let us see a few examples from Croatian: *kriti* 'hide' vs. *pokriti* (po+kriti) 'cover' (literally = to have covered), *raditi* 'do, work' vs. *uraditi* (u+raditi) 'finish' (= to have done), etc. Some analyses have claimed that Hungarian (through close geographical proximity and possibly speakers who were bilingual) also developed perfect aspect with the use of verbal prefixes: e.g., *csinál* 'do' vs. *megcsinál* 'finish' (= to have done), *hív* 'call' vs. *lehív* 'call round, invite', *nevez* 'name' vs. *átnevez* 'rename', etc.

German has a similar system of verbal prefixes that are used for expressing verbal meanings of completion similar to the perfect aspect. Compared to the Slavonic verbal prefixes

that are fixed (and can't ever me removed from their verbal base), the separable prefixes are relatively free in German:

(1) Separable verbal prefixes in German
   a) Ich will es nicht *aufmachen*
      'I will not open it up' (*auf* + *machen* = auf 'up' + *machen* 'do')
   b) Ich habe es nich *aufgemacht*
      'I have not opened it up' (auf + ge 'perfect' + mach + t 'perfect')
   c) Es *aufzumachen* ist wichtig
      'It is important to open this up' (auf + zu 'infinitive' + mach + en 'infinitive')
   d) Er *machte* das jeden Morgen *auf*
      'He opened it up every morning'

All of these examples show that the prefix *auf* is relatively free from the verbal stem: there can be further prefixes after it before the verbal stem (1b and 1 c) and it can occur detached from the verbal stem (1d). Hungarian has a similarly free verbal prefix (see (2)), exemplified here with the verbal prefix *meg*, stem *csinál* 'do', infinitive suffix *-ni*.

(2) Separable verbal prefixes in Hungarian
   a) *Meg* akarta *csinálni*
      He wanted to do that
   b) Ő akarta *megcsinálni*
      It was him who wanted to do that
   c) Ő *csinálta* ezt tegnap *meg*
      It was him who did that yesterday

As you can see, the suffix *meg* can be found glued to the stem (2b), but it can be found detached from it as well (1a and 1 c). We can thus say that German and Hungarian are typologically similar (they behave in a similar way morpho-syntactically). Whether this feature of Hungarian developed under the influence of German (a language that is geographically close to Hungarian) or whether this feature developed independently must remain an open question at this point. All in all, this does not prove that Hungarian and German are genetically related, it merely shows that Hungarian and German could have been geographically (areally) close enough for Hungarian to develop this typological feature. This relationship between Hungarian and German is not one that stems from the pedigree of the two languages (= they are no genetically related), it stems form the fact that they are close enough to borrow morpho-syntactic features from each other (= they are areally related).

5. The Comparative Method

Now it is time to see what it means for languages to be genetically related. Let us see what it means for languages to have a pedigree linking them to another family of the same group. The notion of being related in languages is very similar to degrees of relationship within your (extended) family sharing a pool of common genes: the farther you get from your immediate relations, resemblance will be less perspicuous. In historical linguistics, the terms used are mother, daughter, sister (and sometimes even aunt). So we will see what it means for languages to be a mother-to-daughter and sister-to-sister relationship (this relationship can involve relationships two or even more steps removed). Having a pedigree also means having a network of relationships that can be most conveniently represented in the form of a family tree. We have to tread carefully when we try to establish degrees of relationships. We have to (1) assemble cognates (sets of words that descend from the same ancient - usually reconstructed - word), (2) establish sound correspondences (a set of cognate sounds), (3) reconstruct the proto (ancient) sound (we will see that not every sound can develop into just any other sound), (4) determine the status of similar sets of sound correspondences, (5) check the plausibility of the reconstructed sound from the perspective of the overall phonological system of the proto language, (6) check the plausibility of the reconstructed sound from the perspective of linguistic universals and typological expectations and (7) reconstruct individual morphemes (words, prefixes, suffixes). Now read about all this in Lyle Campbell's *Historical Linguistics* (An Introduction, MIT Press Edition, 1999), read pages 108—148 (excluding section 5.7).

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 descendant languages, 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.)

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.

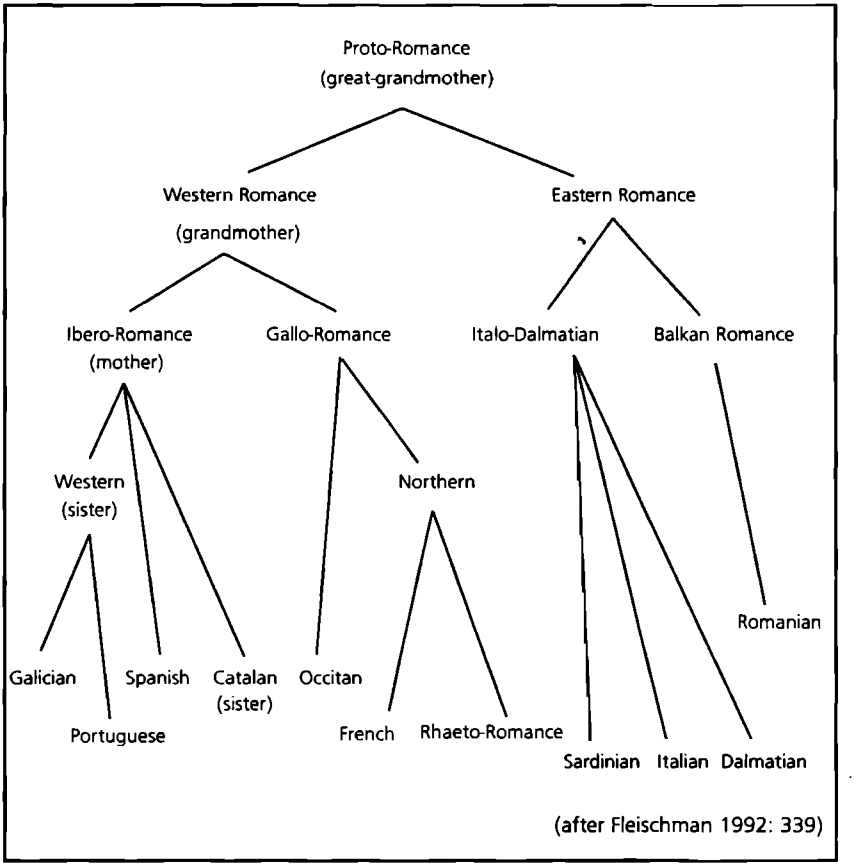

FIGURE 5.1: Proto-Romance family tree (and Spanish's genealogy)

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.)

TABLE 5.1: Some Romance cognate sets

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

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 daughter 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.)

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, where the

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

TABLE 5.2: Kaqchikel–English comparisons

English	*Kaqchikel*
man	ači
mouse	č'oy
moon	qati?t
mother	nan

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_1 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, nasalisation 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 (itself once a single daughter of the proto-language which subsequently split up) 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).

TABLE 5.3: Some additional Romance cognate sets

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

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, we have only two possibilities. One is to explain why the two sets are different. In this case, that would necessitate showing that while the other languages retained *k*, in French **k* had become *ʃ* in environments which must be specified so as to be able to determine when the postulated single sound, **k*, became *ʃ* and when it remained *k* in French. If we do not succeed in showing this, then we are forced to accept the other possibility, that there were two distinct proto-sounds which resulted in the two correspondence sets, where the two distinct sounds merged to *k* in all contexts in Italian, Spanish and Portuguese, in this example.

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*). Therefore, 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 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 6) 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 proto-sounds 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 /battere/	batir /batir/	bater /bater/	battre /batr/	battuere	to beat
11.	bolla /bolla/	bola /bola/	bola /bola/	boule /bul/	bulla	ball, bubble
12.	bontà /bonta/	bondad /bondad/	bondade /bõdaǰi/	bonté /bõte/	bonitāte	goodness
13.	bev- /bev-/	beber /beber/	beber /beber/	boire Old French beivre	bibere	to drink
14.	venire /venire/	venir /benir/	vir /vir/	venir /vənir/	venīre	to come
15.	valle /valle/	valle /baʎe/	vale /vale/	val /val/	valle	valley
16.	vestire /vestire/	vestir /bestir/	vestir /vestir/	vêtir /vetir/	vestīre	to dress

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*	
1.	ra:h	ja	jah	jah-	*ra:h	'hot, spicy'
2.	ri?x	jix	ji?ih	jeh-	*ri?ix	'old (man)'
3.	r-	j-	j-	—	*r-	'his/her/its'
4.	raʃ	jaʃ	ja?aʃ	yaʃ-	*ra?ʃ	'green'
5.	war	waj	waj	waj	*war	'to sleep'
6.	ja:x	jah	jah	ja?	*ja:h	'sick'
7.	jaʃ	jaʃ	—	—	*jaʃ	'crab, pincers'
8.	k'aj-	k'aj- ['sing']	k'aj- ['sing, sell']	č'aj- ['buy']	*k'aj-	'to sell'

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.

Cognate sets 1–5 show SOUND CORRESPONDENCE 1:

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

Cognate sets 6–8 show SOUND CORRESPONDENCE 2:

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

Clearly, by our standard criteria, the best Proto-Mayan reconstruction for SOUND CORRESPONDENCE 2 would be **j* (preserved unchanged in

all the languages). However, all the languages except K'iche' also have *j* 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 **j* 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 *j* 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 **j* for SOUND CORRESPONDENCE 2, two distinct proto-sounds. From this, it follows that **r* merged with *j* in Tzeltal, Yucatec and Huastec, accounting for why they have *j* 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 *j* in cognates where K'iche' has *j* 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).

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.

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

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:

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

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 **x̟* (a somewhat fronted velar fricative) has been proposed to represent correspondence set (2). While the reconstruction with **x* and **x̟* 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, as in the following:

	Yucatec	*Chol*	*Chuj*	*Q'anjobal*	*Motocintlec*	*Mam*	*K'ichean*
(3)	x	h	x	x	x	x	*x
(4)	n	n	ŋ	ŋ	ŋ	x	*x̟

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$_1$ *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, and neither has *č*, only *ts*. 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 glottalised consonants and no plain counterparts, and therefore a reconstruction which claimed that some proto-language had only glottalised consonants and no non-glottalised counterparts would be false. Languages do not have only nasalised vowels with no non-nasalised vowels, and so we never propose a reconstruction which would result in a proto-language in which there are only nasalised 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

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

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 (*$m > b$, *$n > d$, *$\dot{m} > b'$, *$\dot{n} > d'$). 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.

Step 7: Reconstruct individual morphemes

When we have reconstructed the proto-sound 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 hypothesised 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 very simplified comparison of Romance languages. The forms in Table 5.8 are cognates between Finnish and Hungarian. These two languages belong to the Finno-Ugric family, but since there are many other languages also in this family, this example is far from complete enough to offer a full perspective on the proto-language – the two are compared here only for illustration's sake. Finnish and Hungarian 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 two 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 Finno-Ugric. Therefore, here we will be concerned only with

the initial sounds in Sets I–IV and with the medial consonants of Sets V and VI.

TABLE 5.8: Some Finnish–Hungarian cognate sets

	Finnish	*Hungarian*	
Set I:	1. puu	fɔ	tree
	2. pitæ-	fy:	keep
	3. poika	fiu:	boy
	4. pesæ	fø:	nest
	5. puhu- speak, blow	fu:(j)-	blow
	6. purki	forr	snow flurry
Set II:	7. tuo-	toj	take
	8. tutka-	tø:(l)d^{j}	tip, point
	9. tunte-	tud	know
	10. tyvi	tø:	base
	11. talvi	te:l	winter
Set III:	12. kota	ha:z	house, hut
	13. kuole-	hɔl	die
	14. kamara	ha:mlik	skin
	15. kala	hɔl	fish
	16. koi	hɔj-	dawn
	17. kolme	ha:rom	three
	18. kalin	ha:lo:	net
	19. kusi	hu:d^{j}	urinate
Set IV:	20. kivi	kø:	stone
	21. keri	ke:reg	bark
	22. kyynel	kønnj	tear (noun)
	23. kæte-	ke:z	hand
	24. kii- rut, mating	ke:j	(carnal) pleasure
Set V:	25. pato dam, wall	fɔl	wall
	26. ete-	el	before
	27. piti-	fel	long
	28. tæytæ-	tel	fill
	29. løytæ-	lel	find
Set VI:	30. kuole-	hɔl	die
	31. nuoli	nji:l	arrow
	32. kala	hɔl	fish
	33. liemi	leve-	broth
	34. lintu bird	lu:d	goose

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-* (in Set I, nos 1–6)
(2) Finnish *t-* : Hungarian *t-* (in Set II, nos 7–11)
(3) Finnish *k-* : Hungarian *h-* (in Set III, nos 12–19)
(4) Finnish *k-* : Hungarian *k-* (in Set IV, nos 20–24)
(5) Finnish *-t-* : Hungarian *-l-* (in Set V, nos 25–29)
(6) Finnish *-l-* : Hungarian *-l-* (in Set VI, nos 30–34)

In step 3, we attempt to reconstruct the proto-sounds which we believe are reflected by each of these correspondence sets. For SOUND CORRESPONDENCE (1) (*p* : *f*) our choices are: [1] reconstruct **p* and assume that Hungarian has changed to *f*; [2] reconstruct **f* and assume that Finnish has changed this to *p*; or [3] reconstruct some third thing (say **p*^{h}) and assume that both changed, that Hungarian changed in one way to give *f* and Finnish in another to give *p*. In looking at 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}. Since in this comparison only two languages are involved, we will not be able to make use of the majority-wins principle to help us in reconstruction. In the guideline of factoring in features held in common, we may conclude from *p* and *f* 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. Steps 4 and 5 will help us resolve which of these possibilities is the best reconstruction, which for now we will take to be [1], with **p*, based on directionality of change and on economy. Economy urges us to postulate only one change, **p* > *f* in Hungarian, whereas **p*^{h} would require the postulation of two changes, **p*^{h} > *p* in Finnish and **p*^{h} > *f* in Hungarian.

SOUND CORRESPONDENCE (2) (*t-* : *t-*) appears to reflect **t-* (where neither language changed).
SOUND CORRESPONDENCEs (3) (*k-* : *h-*) and (4) (*k-* : *k-*) may present a challenge.

In (4) we reconstruct **k-*, since neither language changed. However, (3) would also seem to be best reconstructed as **k-* based on directionality of change, since the change *k* > *h* is very common and not unexpected, whereas a change *h-* > *k-* is all but unknown. We move to step 4 to attempt to resolve the difficulty of the partially overlapping SOUND CORRESPONDENCEs (3) and (4). This means that if we can show that both reflect the same original sound because one of the languages has undergone a conditioned change where that sound changed in some environments but not others, then we reconstruct only a single sound, the same for both sets, explaining the difference between them by writing out the conditions under which the one language changed so that it has two different outcomes from the single original sound. If we cannot explain the difference in this way, then we are obliged to reconstruct two distinct proto-sounds, one to represent each of the two sound correspondences, with the assumption that the two merged to *k* in Finnish. This, then, requires us to take a closer look at the cognate sets in question (those of Sets III and IV). We notice that in the cognates of Set III Hungarian has *h-* which appears only before back vowels (*u*, *o*, *a*), whereas in the cognates of Set IV Hungarian has *k* and it 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) and remained *k* before front vowels (as in Set IV); we reconstruct **k*. We might wonder whether the proto-language might not have had **h* which then changed to *k* before front vowels in Hungarian and to *k* in all environments in Finnish. First, directionality argues against this possibility (since the change *h* > *k* is essentially unknown anywhere). Second, the criterion of economy also goes against this alternative; it is more plausible to assume that only one change took place, **k* > *h* before back vowels in Hungarian, than to need to suppose that two independent changes occurred, one of **h* > *k* before front vowels in Hungarian and another independent one of **h* > *k* in all contexts in Finnish.

The medial sounds in SOUND CORRESPONDENCEs (5) and (6) present a similar problem. Since Hungarian has *-l-* in both these while Finnish has *-t-* in (5) but *-l-* in (6), in step 4 we must determine whether it is necessary to reconstruct two distinct sounds or whether these two can be put together as different outcomes from the same original sound due

to some conditioned sound change in Finnish which resulted in the difference. To make the long story shorter, which would be clearer if more cognate sets were presented, we search in vain for any conditioning factor by which we might assume that an original **-l-* became *-t-* in Finnish in some environments but remained *-l-* in others. Both *t* and *l* occur in all positions (initial, medial, final) and both before and after all vowels in the Finnish cognates. Therefore, we have no choice but to reconstruct two distinct sounds, and we choose **t* for (5) and **l* for (6). This requires us to assume that medial **-t-* and **-l-* merged to *-l-* in Hungarian.

Let us return to SOUND CORRESPONDENCE (1) (*p-* : *f-*) and apply steps 5 and 6. For this, let us assume that we have available in Table 5.8 all the evidence for possible stops in Finnish–Hungarian comparisons. Our tentative reconstructions based on the sound correspondences to this point give us:

**p* (1) Finnish *p-* : Hungarian *f-* (in Set I, nos 1–6)
**t* (2) Finnish *t-* : Hungarian *t-* (in Set II, nos 7–11)
**-t-* (5) Finnish *-t-* : Hungarian *-l-* (in Set V, nos 25–29)
**k* (before back vowels) (3) Finnish *k-* : Hungarian *h-* (in Set III, nos 12–19)
**k* (before front vowels) (4) Finnish *k-* : Hungarian *k-* (in Set IV, nos 20–24)
**l* (6) Finnish *-l-* : Hungarian *-l-* (in Set VI, nos 30–34)

We check these in step 5 to see how plausible the resulting phonemic inventory (sound system) would be if we keep these sounds. A language with the stops *p, t, k* would be quite normal. If we did attempt to reconstruct possibility [3] (some third thing from which to derive *p* and *f* naturally and plausibly, say **p^h*) for correspondence set (1), we would no longer have a natural, symmetrical phonemic inventory (**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*. 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* 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 that 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 very few with *f, t, k* (and no *p*), but hundreds with *p,*

t, *k*. Putting all these considerations together, directionality, economy, internal consistency and typological realism, we conclude that the reconstruction of **p* is the best of the alternatives for SOUND CORRESPONDENCE (1). In turn, we would apply steps 5 and 6 to the other reconstructions, **t* and **k*; we would find these to be supported. We would find that the possible alternative with **h* for SOUND CORRESPONDENCEs (3) and (4) which could have been considered would be inconsistent internally and typologically, not to mention being against economy and the known directionality of change.

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 (p, t, k) > voiceless fricatives (f, θ, h(x))
> voiced stops (b, d, g) > voiceless stops (p, t, k)
> voiced aspirated stops (bh, dh, gh) > voiced plain stops (b, d, g).

(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.

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

Sanskrit	*Greek*	*Latin*	*Gothic*	*English*
Set Ia: *p > f				
pad-	pod-	ped-	fōtus	foot

Sanskrit	*Greek*	*Latin*	*Gothic*	*English*
páñča [páɲča]	pénte	[quinque] [k^{w}ink^{w}e]	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	treĩs/tría	trēs	þrija	three
tv-am	tū (Doric)	tv-am	þu	thou
-ti- gátis 'gait'	-ti- mor-tis 'death'	-tis/-sis básis 'going'		-th 'nominaliser' health, truth, birth, death

Set Ic: *k > 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 [dáʃə]	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)

Sanskrit	*Greek*	*Latin*	*Gothic*	*English*
		(*Lithuanian*) dubùs	diups	deep [OE dēop]
	kánnabis	(*Lithuanian*) kanapēs]		hemp (borrowing?)
		Latin lūbricus	sliupan	slip

Set IIb: *d > t

d(u)vā́-	dúo/dúō	duo	twái [twɛ̄]	two
dánt-	odónt-	dent-	tunþus	tooth
dáśa [dáʃə]	déka	decem [dekem]	taíhun [tɛxun]	ten
pad-	pod-	ped-	fōtus	foot
ad- 'eat'	édō 'I eat'	edō 'I eat'		eat [OE etan]
véda 'I know'	woīda 'I know'	videō 'I know'	wáit [wɛ̄t] 'I know'	wit 'to know'

Set IIb: *g > k

ǰánás	génos	genus	kun-i 'race, tribe'	kin
ǰánu-	gónu	genū	kniu	knee
ǰnātá	gnōtós	(g)nōtos	kunnan 'to know'	known
ájra- 'country'	agrós	ager	akrs	acre 'field'
mr̥ǰ- 'to milk'	(a-)mélgō 'to squeeze out'	mulgeō 'I milk'	miluk-s 'milk'	milk

Sanskrit	*Greek*	*Latin*	*Gothic*	*English*
Set IIIa: *bh > b				
bhar-	phér-	fer-	baír-an [bɛran] 'to bear'	bear
bhrā̆tar	phrā̆tēr	frā̆ter	brōþar	brother
a-bhū-t 'he was'	é-phu 'I brought forth'	fu-it 'he was'	bau-an [bō-an] 'to dwell'	be
Set IIIb: *dh > d				
dhā- 'put'	ti-thē-mi 'I put'	fē-cī 'I made'		do [OE dō-n]
dhṛaṣṇóti 'he dares'	thrasús 'bold'	(fest-)	(ga-)dars 'he dares'	dare [OE dear(r)] 'he dares'
dvār-	thŭr-a	for-ēs	daúr- [dor-]	door
vidhávā	ē-wíthewos 'unmarried youth'	vidua	widuwo	widow
mádhu	méthu			mead
madhya-	mésos	medius	midjis	mid
Set IIIc: *gh > g				
haṁs-á- '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

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

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.	aṣṭā̆u [əṣṭə̆u]	oktō	octō [oktō]	ahtáu [axtau]	eight
4.	nákt-	nukt-	noct- [nokt-]	nahts [naxts]	night
5.			capt(īvus)	(haft)	[OE hæft] 'prisoner'
6.	-ti- gátis 'gait'	-ti- mor-tis 'death'	-tis/-sis básis 'going'		-t 'nominaliser' thrift, draught, thirst, flight, drift
7.			piscis [piskis]	fisks	[OE fisc] 'fish'

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*, *truth*, *birth*, *death* of Table 5.9 is explained in the same way. The /θ/ forms (as in Table 5.10) underwent Grimm's Law (**t* > *θ*); the forms with *-t* (in Table 5.9) 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 14). 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* 'I put, place' (with reduplication of root *dhē-*) 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-	patḗr	pater	fadar [faðar]	OE fæder 'father'
(3)	śatám [śətəm]	(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 (plain) voiceless stops and fricatives in a root became voiced in Germanic; otherwise (when the accent preceded the sound or when the sound was root-initial) Grimm's Law applied. 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 *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, symbolised as . . . C . . . ´) contrast with forms with the accent before the sound in question (indicated as ´. . . C. . . , cases which undergo Grimm's Law).

TABLE 5.12: Examples contrasting the effects of Grimm's Law and Verner's Law on medial consonants

Grimm's Law	*Verner's Law*
´. . . C. . .	. . . C . . . ´
*p > f	*p > b [β]
(1a) OE hēafod 'head'	(1b) Gothic sibun [siβun] 'seven'
Latin cáput [káput]	Sanskrit saptá-
*t > θ	*t > d [ð]
(2a) Gothic brōþar [brōθar] 'brother'	(2b) OE fæder 'father'
Sanskrit bhrā́tar-	Sanskrit pitár-
*k > x	*k > g [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.

TABLE 5.13: Verner's Law in grammatical alternations

	'I become'	*'I became'*	*'we became'*	*'became [participle]'*
PIE	*wértō	*(we)wórta	*(we)wr̥təmé	*wr̥tomós
Sanskrit	vártāmi	va-várta	vavr̥timá	vr̥tānáḥ
	'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

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.

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 reasonable idealisation; 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 it may have been 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 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 glottalised 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.