CHAPTER ELEVEN

PROBLEMS WITH TRADITIONAL ASSUMPTIONS

The comparative method that I discussed in the Chapter 5 was developed mainly in the 1800s, largely by German scholars. This method may seem very straightforward if you carefully apply it, following the steps that I set out in that chapter. However, it can sometimes be difficult to apply the method in particular situations. In this chapter, I will look at some of the problems that linguists have come across in applying the method. I will begin by looking at the historical development of the comparative method, and its refinement by the neogrammarians of last century, along with some of the difficulties in the method that they recognised from the very beginning. I will then go on to look at some of the more fundamental objections that modern linguists have raised to a strict application of the comparative method.

11.1 THE NEOGRAMMARIANS

The comparative method that I described in Chapter 5 was first developed in Europe, mainly by German scholars, and it was first applied to the languages of the Indo-European language family. This family includes all of the languages that were first recognised by Sir William Jones in 1786 as being descended from a common ancestor. It was perhaps natural that European scholars should investigate the history of their own languages first, as these were languages with a very long history of writing. This made it possible to start their reconstructions further back in time than they could have done with languages that were unwritten, or which had only recently been written. A long history of writing also made it posssible to check on the accuracy of reconstructions that had been made from the present.

After the period of European voyaging and exploration between the 1400s

and the 1700s, scholars came into contact with a wide range of languages that were previously unknown in Europe. Word lists were compiled in 'exotic' languages for people to see the similarities and differences between them. Before the nineteenth century, a field of enquiry called *etymology* had become quite well established. This term is currently used to refer simply to the study of the history of words, though in earlier times the history of 'words' and the history of 'languages' were often confused.

Many of the early attempts at etymology would be regarded as childish by modern standards. One French scholar called Étienne Guichard in 1606 compiled a comparative word list in Hebrew, Chaldaic, Syrian, Greek, Latin, French, Italian, Spanish, German, Flemish, and English, in which he tried to show that all languages can be traced back to Hebrew! The kind of evidence that he presented to support his hypothesis was the existence of similarities between words such as Hebrew *dabar*, English *word* and Latin *verbum*. Some scholars who followed Guichard were more sceptical of these methods, and Voltaire, a famous French writer, described etymology as the science in which 'the vowels count for little and the consonants for nothing'. Unkind words, but true, at least as Guichard had applied it.

Sir William Jones's words in 1786 about Sanskrit and other Indo-European languages profoundly altered the perception of the nature of linguistic relationships among serious scholars. However, this did little to stop those less concerned with these more modern views from continuing in the path of earlier commentators — I hesitate now to use the word 'scholar' — in making random observations about similarities between languages as evidence of linguistic relationships. There were books published in the late 1800s which attempted to demonstrate the relationship between the languages of Vanuatu and those of the Middle East; this is a relationship that no modern linguist would take the slightest bit seriously, I should point out. Other scholars have taken random similarities in language and cultural artefacts as evidence that Hawaii was populated from Greenland; that parts of Polynesia were populated from South America; and that different peoples on earth were provided with aspects of their culture by beings from outer space. I wouldn't want to rule out these interpretations as impossible, but the linguistic evidence is certainly far from compelling, and modern linguists tend to assign these kinds of view to the lunatic fringe.

Sir William Jones also opened the eyes of European scholars to a whole new field of linguistic data by turning people's attention for the first time to Sanskrit and the languages of India, in addition to altering the perceptions that people had about the nature of language relationships. Jones emphasised that it was similarities in the *structure* of the Indo-European languages, rather than the individual similarities between words, that were important in determining language relationships. This observation led to a new intellectual climate in the study of language relationships, as scholars started looking instead for grammatical similarities between languages to determine whether or not they should be considered to be related. Lexical similarities, it was argued, were

poor evidence of genetic relationship, as similarities between practically any word in any two languages can be established with enough effort.

Rasmus Rask in 1818 investigated the history of the Icelandic language on the basis of its grammatical similarities to other Germanic languages (such as Norwegian, German, and English), and largely ignored the lexicon. Rask also argued, however, that while individual lexical similarities were not good evidence of linguistic relationship, repeated occurrences of sound correspondences between words could not be due to chance, so these were good evidence of genetic relationship. By recognising only repeated occurrences of sound correspondences as valid evidence in the study of language, it was possible to exclude chance lexical similarities such as those noted above by Guichard for Hebrew, English, and Latin.

In 1822, Jakob Grimm described a series of sound correspondences that he had noted between Sanskrit, Greek, Latin, and the Germanic languages (which also include the now extinct Gothic language, as well as English). For instance, he noticed that very often, where Sanskrit, Greek, and Latin had a /p/, the Germanic languages had an /f/; where Sanskrit, Greek, and Latin had a /b/, the Germanic languages had a /p/; and finally, where Sanskrit had a /bh/,[1] the Germanic languages had a /b/ — for example:

Sanskrit	Greek	Latin	Gothic	English	
pa:da	pous	pes	fotus	foot	
-	turbe:	turba	θaurp	thorp	('Thorp' is an old word in English for 'village', but now it only occurs in place names, such as Mablethorpe, Scunthorpe, etc.)
bhra:ta:	-	-	-	brother	

(You should note that we are considering only the sounds written in bold type at this point. The remaining sounds have far less obvious correspondences than these, so perhaps you can appreciate the advantage in having learned to apply the comparative method using the much more straightforward correspondences that are to be found in the Polynesian languages!) The full set of sound correspondences that Grimm noted are set out below, along with the reconstructed protophonemes:

Proto Indo-European	Sanskrit	Greek	Latin	Germanic
*p	p	p	p	f
*t	t	t	t	θ

[1] The sounds that are represented by the digraphs **bh, dh, gh** in Sanskrit and by **ph, th, kh** in Greek are voiced and voiceless aspirated stops respectively.

*k	c	k	k	x
*b	b	b	b	p
*d	d	d	d	t
*g	ɟ	g	g	k
*bh	bh	ph	f	b
*dh	dh	th	f	d
*gh	ɟh	kh	h	g

Germanic voiceless fricatives correspond mostly to voiceless stops in the other languages, and Germanic voiceless stops correspond to voiced stops. Germanic voiced stops have a more complicated set of correspondences, as they correspond to voiced aspirated stops in Sanskrit and voiceless aspirated stops in Greek (with the Latin correspondences being somewhat less predictable in this case).

According to the methodology that I set out in Chapter 5, the forms in the left-hand column can be reconstructed for the language from which all of these languages were descended. That is, we reconstruct in the protolanguage the form that is most widely distributed in the daughter languages, and we reconstruct original forms that involve 'natural' rather than 'unnatural' changes. You can see that of the four descendant languages, Sanskrit is clearly the most conservative as it has undergone fewer changes in these consonants from the protolanguage (though there are plenty more changes in other aspects of the language!). The Germanic languages are clearly the ones that have changed the most since Proto Indo-European with respect to these consonants.

No scholar at the time thought to distinguish between sound correspondences that were without exception and those which appeared to be sporadic (i.e. which applied in some words but not in others). In fact, while the correspondences that Grimm noted were found to be true for very many words, there were at the same time many words in which the correspondences did not hold, and other correspondences were apparent instead. There were, for example, many voiceless stops in Sanskrit, Greek, and Latin that corresponded to voiceless stops in Germanic instead of voiceless fricatives:

Latin	Gothic	
spuo	**speiwan**	'spit'
est	**ist**	'is'
noktis	naxts	'night'

The Gothic forms were not **/sfeiwan/, /isθ/,** and **/naxθs/** as we might expect if the correspondences noted by Grimm were to be completely general. However, it was soon realised that the correspondence of Sanskrit, Greek, and Latin voiceless stops to Germanic voiceless stops, and Sanskrit, Greek, and Latin voiceless stops to Germanic voiceless fricatives were in fact in complementary distribution.

In Chapter 5, you saw that when a conditioned sound change takes place in any of the daughter languages, the result is that the sound correspondence sets end up being in complementary distribution. So, once you have set out the full range of correspondence sets, you must check to see whether phonetically similar correspondence sets are in complementary or contrastive distribution. If it turns out that they are in complementary distribution, you need only reconstruct a single original phoneme that has undergone a conditioned sound change. The first of the two correspondences just mentioned was found only when Gothic had a preceding fricative, whereas the second correspondence was found when there was no preceding fricative. We can therefore reconstruct both correspondences as going back to a single voiceless stop series. This would make it necessary to reconstruct a conditioned sound change of the following form in the Germanic languages:

voiceless stop → { voiceless / fricative ____ ; voiceless / elsewhere }

More and more sound correspondences came to be recognised as being due to the influence of phonetic factors of some kind, such as the nature of the preceding or following sounds, the position of stress, or the position of the sound in the word (i.e. whether it occurred word initially, medially, or finally). By taking into account yet other phonetic factors, Herman Grassmann was able to account for a further set of consonant correspondences in these languages. Scholars had noted that some voiced stops in the Germanic languages corresponded to aspirated stops in Sanskrit and Greek (as covered by Grimm's statement, as you have just seen), but some voiced stops corresponded to unaspirated stops. Scholars were once again faced with a double set of correspondences.

Grassmann was able to show that these two sets of correspondences were also in complementary distribution, and that both Sanskrit and Greek had undergone conditioned sound changes. Note the following forms in these two languages:

Greek		Sanskrit	
do:so:	'I will give'	a-da:t	'he gave'
di-do:mi:	'I give'	da-da:mi	'I give'
the:so:	'I will put'	a-dha:t	'he put'
ti-the:mi:	'I put'	da-dha:mi	'I put'

The first pairs of forms in these two languages indicate that there is a regular

morphological process of partial reduplication involving the initial syllable of the verb. This process derives the present stem of the root of these verbs, which are seen more clearly in the Greek future and Sanskrit past tenses. When a syllable containing an initial aspirated stop is reduplicated, the reduplicated syllable contains an unaspirated stop. In Chapter 2, this kind of change was described as dissimilation at a distance.

Grassmann related this kind of morphological alternation in these two languages to the unpredictable correspondence between Germanic voiced stops and Sanskrit and Greek unaspirated stops, as illustrated by the example below:

Sanskrit	Greek	Gothic	
bo:dha	**p**ewtho	**b**ewda	'bid'

According to Grimm's earlier generalisation about sound correspondences, where Germanic languages such as Gothic have **/b/** we would have expected to find **/bh/** in Sanskrit and **/ph/** in Greek. Grassmann concluded that Sanskrit and Greek did in fact have these forms originally in words such as these but that the aspiration was subsequently lost under the influence of the aspiration of the stop in the following syllable. So, an earlier (and unrecorded) form of Sanskrit, for example, would have had **/*bho:dha/**, which would have corresponded regularly with Gothic **/bewda/**. However, with two adjacent syllables in Sanskrit containing aspirated stops, the first of these then lost its aspiration to become a plain stop. A parallel change was also suggested for Greek to explain the once apparently irregular correspondence for this language.

In 1875, Carl Verner was able to dispose of yet another set of apparently irregular forms according to Grimm's statement of sound correspondences in the Indo-European languages. If you compare Latin **/pater/** with Gothic **/fadar/**, both meaning 'father', you will see that there is a correspondence here between Latin **/t/** and Germanic **/d/**. However, you will remember from the statement of the corresondences that Grimm noted earlier that where Latin has **/t/**, we would normally have expected Germanic languages to have **/θ/**. Verner collected a full set of such irregular forms and showed that the correspondences of **t = d** and **t = θ** were in complementary distribution, with one correspondence showing up when the following vowel was stressed in Proto Indo-European, and the other correspondence showing up when the vowel was unstressed.

Grimm had stated earlier that:

> ... the sound shifts succeed in the main but work out completely only in individual words, while others remain unchanged.

He stated this because of the large number of forms which did not fit his generalisations. However, with the discoveries of Grassmann, Verner, and others, most of these irregularities were eventually eliminated. Towards the

end of the nineteenth century, scholars such as Brugmann and Leskien were stating that 'sound laws operate without exception'.

The sound correspondences that Grimm, Verner, and Grassmann had noted were restated as 'laws' to emphasise the fact that they could not be 'broken'. Newtonian physics gave Brugmann and Leskien a model of a closed system in which there could be no exceptions, just like the laws of gravity. Darwinian biology offered them a model of organisms developing according to unbendable laws of nature (i.e. the survival of the fittest). This was the birth of the *neogrammarian* school, often also referred to as the *Junggrammatiker*, using a word taken from German.

The neogrammarians argued that these phonetic laws operated without exception in a language, and they argued further that the only conditioning factors that could determine the course of a sound change were phonetic factors. They claimed that it was impossible for semantic or grammatical factors to be involved in the conditioning of sound changes. Thus, for example, it would be impossible for a particular change to affect all words referring to trees, but not words referring to birds as well, and it would be impossible for a change to operate in nouns without affecting verbs at the same time. The only factors which could condition a sound change were phonetic factors such as the nature of the preceding and following sounds, the position of the sound in the word, and so on.

This was a very significant innovation in thinking for historical linguists. Once it was acknowledged that sound change was a regular process which operated without exceptions, it became possible for the study of etymology, or the study of the history of words (and therefore also of languages) to become *scientific* (i.e. rigorous and open to proof). Scholars now had a way of arguing scientifically against proposals such as those of Étienne Guichard who tried to relate all languages to Hebrew, as you saw earlier in this chapter. A sound correspondence or a similarity between two languages is of no value for reconstruction or for determining linguistic relationships unless it is *systematic* or *regular*.

In reconstructing the history of languages, you therefore need to make the important distinction between a *systematic* (or *regular*) sound correspondence and an *isolated* (or *sporadic*) correspondence. This is a distinction that I did not make in Chapter 5 when I was talking about the comparative method, but it is very important. Between steps 2 and 3 of the comparative method as I summarised it at the end of Chapter 5, therefore, we need to add a further step which says the following:

Separate those correspondences which are systematic from those which are isolated (i.e. which occur in only one or two words) and ***ignore*** *the isolated correspondences.*

Let us look at an example of what I mean by this. In addition to the forms that I gave in Chapter 5 for Tongan, Samoan, Rarotongan, and Hawaiian, let us also add the cognate forms below:

Tongan	Samoan	Rarotongan	Hawaiian	
fonua	fanua	ʔenua	honua	'land'

If we were to set out the sound correspondences that are involved in that cognate set, we would have an initial correspondence of **f = f = ʔ = h**, followed by a correspondence of **o = a = e = o**, then **n = n = n = n**, then **u = u = u = u**, and finally **a = a = a = a**. There is nothing new in the correspondences involving the initial consonants, nor the final segments **/-nua/**, but correspondence involving the vowels of the first syllable *is* different from any other correspondence that you saw in Chapter 5.

According to what I said in Chapter 5, you should assume that each set of correspondences that is not in complementary distribution with any other correspondence should be reconstructed as going back to a separate original phoneme. If we were to reconstruct this new correspondence as going back to a separate protophoneme, however, you would end up reconstructing a new phoneme which occurs in just this single word. Rather than complicate the statement of the phonemes of the original language, what you do is simply *ignore* such isolated correspondences, and reconstruct *only* on the basis of the evidence provided by systematic sound correspondences. You should therefore reconstruct the word for 'land' on the basis of regular correspondences only. There is not enough data in these four languages to allow you to decide whether the original vowel was **/*e/**, **/*o/**, or **/*a/**. The occurrence of reflexes of ***o** in both Tongan and Hawaiian might suggest that **/*fonua/** was the original form, with Samoan having undergone a sporadic shift of the vowel to **/a/**, and Rarotongan having upredictably shifted the vowel to **/e/**. Comparing these languages with non-Polynesian languages which also have cognates of this word, such as Fijian **/vanua/**, we might be tempted to reconstruct Proto Polynesian as having had **/*fanua/** instead. But whatever the reconstruction, we are simply going to have to accept that there have been some completely unpredictable changes in the vowels of some of these languages.

Another example to illustrate the same kind of problem involves the additional cogate set below:

Tongan	Samoan	Rarotongan	Hawaiian	
paaʔi	paʔi	paki	paʔi	'slap'

In this case, the medial correspondence of **aa = a = a = a** is not attested outside this cognate set, and the same is true of the correspondence of **ʔ = ʔ = k = ʔ**. The Samoan, Rarotongan and Hawaiian data is perfectly consistent with what you saw in Chapter 5, pointing to the original form having been **/*paki/**. If the Tongan form were to behave as predicted, it should have been **/paki/**, but instead we find **/paaʔi/**. We must note that there has been an unpredictable change in Tongan of **/*a/** to **/aa/**, and another unpredictable change of **/*k/** to **/ʔ/**.

According to the Neogrammarian Hypothesis that sound change is without exception, there *must* be some kind of explanation for irregularities such as this. What neogrammarians said was that instead of being irregular, such correspondences must involve some other factors. It could simply be a matter of 'undiscovered regularity' — there may in fact be a regular phonetic conditioning factor which nobody has yet been clever enough to uncover. In this case, the explanation is perhaps that the Tongan form /paaʔi/ has been incorrectly identified as cognate with the forms in the other languages. Despite the similarity in the phonological shape and the meaning, it could be that this word is in fact derived from the quite separate (and not cognate) root /paa/, and that the final syllable is a suffix /-ʔi/, which is added to many transitive verbs in Tongan.

The neogrammarians did find some ways of accounting for some irregular sound correspondences as well, and it is to these that I will turn my attention in the following sections.

11.2 ANALOGY

The term *analogy* is used in a non-technical sense to mean that we find similarities between things that are not ordinarily regarded as being similar. In presenting an argument, we often 'draw an analogy' as a way of illustrating a new concept, by taking a concept that we know our audience is familiar with and showing how it is similar to the new concept that we are talking about. For example, if you were trying to explain the unfamiliar concept of complementary distribution of the allophones of a phoneme to a beginning student of linguistics, you could use an analogy to help get your point across. You might say that complementary distribution can be compared to the relationship between formal and non-formal education. Formal education is carried out only in certain contexts and by certain people (i.e. by qualified teachers in approved schools). Non-formal education also takes place in particular sets of contexts, but different ones, and is generally carried out by different people as well (i.e. out of school; by our parents, community leaders, agricultural extension officers, village leaders in Pacific villages, and so on). Similarly, you could say that certain allophones of phonemes may occur only in certain phonetic contexts, and other allophones in other contexts. Although there is nothing else in common between phonemes and education, we can use the similarity that does exist to illustrate this particular difficult concept.

Analogies can be represented by using a formula of the following type:

A:B::C:D

This formula is to be read as follows:

A is to **B** as **C** is to **D**

Alternatively, it can be read as follows:

> *The relationship between **A** and **B** is the same as the relationship between **C** and **D**.*

Using this formula, we can represent the analogy that I just drew between phonemes and education as follows:

> *formal education: non-formal education:: one allophone: another allophone*

This can be read as follows:

> *The relationship between formal and non-formal education is the same as the relationship between two allophones of the same phoneme.*

Analogy was frequently invoked by the neogrammarians as a way of accounting for problematic sound correspondences in the languages that they were studying. I will now discuss analogical change in language under a number of headings.

(a) Analogical change by meaning

Analogy is a very powerful force in language change, and this fact was recognised by the neogrammarians. Speakers of a language often perceive a partial similarity between two forms on the basis of their meaning alone, even when there is no similarity in their actual forms. Speakers of languages sometimes even change the shape of a word to become more like that of another word to which it is related only by meaning. To do this is to change the phonetic shape of a word by analogy, and we can express this using the following formula:

$meaning_a$: $meaning_b$:: $form_a$: $form_b$

Given that the relationship between form and meaning in language is by and large arbitrary (as Saussure noted towards the beginning of this century), we would not ordinarily expect that two related meanings would be expressed by related forms. However, similarities in meaning sometimes *do* cause words to change their shape so that they end up being phonologically closer to each other than they would have been if they had been subject to all of the regular sound changes. Let us examine the history of the words for 'four' and 'five' in Latin:

		Latin	
*kwetwo:res	→	kwattwor	'four'
*penkwe	→	kwinkwe	'five'

If **/*penkwe/** had changed according to the regular rules in Latin, it should have ended up as **/pinkwe/** rather than as **/kwinkwe/**. Why, then, did **/*p/** irregularly change to become **/*kw/** in this single word in Latin? The answer is that on the basis of the similarity in meaning of the two words (i.e. both refer to numbers one after the other), this similarity is also extended to the shape of the words as well as their meaning. Speakers of Latin at some point in time changed one of these two forms so that it became a little more like the other. So, on the analogy of **/*kw-/** initially in the word for 'four', **/*p-/** shifted irregularly to **/kw-/** in the word for 'five'.

Presumably, of course, the analogy could have gone the other way, with the word for 'four' shifting unpredictably to become more like the word for 'five'. If that had happened, presumably the word for 'four' would have ended up being **/pattwor/** instead of **/kwattwor/** (and the modern French word for 'four' would presumably have been something like *patre* instead of *quatre*!). In fact, in Germanic languages, this is exactly what happened. That is why the English words 'four' and 'five' both have initial **/f-/**. If the English word 'four' had not been influenced by the initial consonant of the next numeral, our word for 'four' today would have been written *whour*!

As a further illustration of the point that analogy operates unpredictably, let us turn our attention to the words *deux* 'two', *trois* 'three' and *quatre* 'four' in some non-standard varieties of modern French. When the word *quatre* appears before a noun that is pronounced with an initial vowel, some speakers of French now add a final **/-z/** to the word *quatre*, making it *quatres*, on the analogy of the **/-z/** at the end of the words *deux* and *trois*. So, compare the following examples:

	Standard French	Non-Standard French	
deux articles	dœz aʁtikl	dœz aʁtikl	'two articles'
trois articles	tʁwaz aʁtikl	tʁwaz aʁtikl	'three articles'
quatre articles	katʁ aʁtikl	katz aʁtikl	'four articles'

(b) Analogical change by form

Analogy need not take just meaning as the basis for comparing two forms, as in the examples that we have just looked at. Analogical change can also operate when there is a perception of partial similarities between two *forms* without any consideration of meaning. For instance, earlier in the history of English there was a word *ewt* which referred to a creature that looks like a small lizard. In modern English, this word has become *newt*, having unpredictably added an initial **/n-/**. It was not a regular change in English for **/n-/** to be added to words that have initial vowels, so we need to find an explanation for this particular irregularity.

Once again, we can invoke analogy as the explanation. In English, we also have words like *name* which have always had an initial **/n-/**, and words like *apple*, which have always had an initial vowel. The indefinite article in English varies in shape between *a* and *an*, with *a* occurring when the following noun begins with a consonant, and *an* occurring when there is a vowel at the beginning of the noun. So, compare the following:

a name
an apple

The old word *ewt* began with a vowel, so according to this rule, the indefinite article should have taken the form *an* rather than *a*, i.e. *an ewt*. However, in saying *an ewt*, earlier speakers of English evidently stopped breaking up the words between *an* and *ewt* as they started to associate this phrase with phrases like *a name*, rather than with other phrases such as *an apple*. So, by analogy of one form with another, *an ewt* became *a newt*.

(c) Folk etymology

Another kind of analogy that we often find is referred to as *folk etymology* or *popular etymology*. Etymology, as you have already seen, is the study of the history of words. When we speak of *folk* or *popular* etymology, we mean that people who speak a language often make their own guesses about what the history of a word is on the basis of partial similarities to some other words (and in doing this they obviously have no interest in what the professional etymologist might have to say about the history of the word!). Speakers of the language may then actually change the word so that its pronunciation comes more into line with what they think is the origin of the word.

Folk etymology tends to take place in words that are relatively long and in some sense felt to be 'unusual' by speakers of the language. Speakers may then take part of this word, or all of it, and change it so that it looks more like a word that they already know. For instance, the word 'crayfish' in English was originally copied from an older French word *crévisse* (and it had nothing to do with fish at all). Ordinarily, such a word would probably have been copied into English as something like *creviss*. Although this word was a single morpheme in French, English speakers apparently felt that it was long or unusual enough in its sound that it must 'really' be two morphemes. They noted a partial similarity in meaning between French *crévisse* and English 'fish', as both are edible creatures that live in water, and they also noticed the partial similarity in shape between French *-visse* and English 'fish'. So, these earlier speakers of English changed the word to become 'crayfish' because they felt that was what the word should have been according to their own view of where it came from. Professional linguists, of course, would say that the word 'fish' originally had nothing to do with this word!

Folk etymology can be seen to be taking place when speakers make certain mistakes in pronunciation. A person who says *ashfelt* instead of *asphalt* is operating under this influence. Presumably they see the greyish-black colour of the asphalt (which is referred to as bitumen, tar, tar-seal, or tar macadam in other varieties of English) and equate it with the greyish-black ash from a fire, as well as the black colour of felt cloth, and rename it accordingly. A person who refers to *watercress* as *water grass* is doing the same thing, and so is somebody who says *sparrow grass* instead of *asparagus*.

(d) Hypercorrection

In Chapter 10, you saw how variability is involved as a factor in causing the spread of language change, and one of the concepts that you came across there was *hypercorrection*. Hypercorrection refers to the situation when a word may have two possible pronunciations, one of which is regarded as *prestigious* (i.e. looked up to, or having positive social value), while the other is *stigmatised* (i.e. looked down on, or having negative social value). In many varieties of English, for example, there are two different ways of pronouncing the word 'dance', i.e. **/dæns/** and **/da:ns/**. Of these, the second generally has higher social value than the first, and if you want to show people how educated you are, or you want to indicate that you are not from the working class, you might use the more 'posh' **/da:ns/** pronunciation. However, if somebody substitutes a variable sound in a word or in an environment where it is not appropriate, then that person is engaging in hypercorrection, or 'over-correcting'. For instance, if someone were to accidentally say **/ʌndəsta:nd/** instead of **/ʌndəstænd/**, this could be the reason.

Another example comes from Bahasa Malaysia. In the standard variety of this language there are words containing the phoneme **/r/**, and there are also words borrowed from Arabic that contain the voiced velar fricative **/ɣ/**. In the area of Malaysia known as Perak, there is a variety of the language that is known locally as Celaka Perak, which translates as 'the Perak misfortune'. You will no doubt guess from its name that people think that this dialect sounds 'funny', and that it is a stigmatised dialect. One of the features of Celaka Perak is that it merges the distinction between **/r/** and **/ɣ/**, and all words containing these sounds are pronounced in Celaka Perak with the velar fricative. The result is that we find the following regular correspondences between standard Bayasa Malaysia and Celaka Perak:

Standard Bahasa Malaysia	Celaka Perak	
ratus	ɣatuih	'hundred'
ribu	ɣibu	'thousand'
buruk	buɣuk	'rotten'
loɣat	loɣat	'accent'

When somebody from Perak is trying to speak the standard language, one thing that they have to remember to do is to substitute **/r/** for **/ɣ/** in order to avoid sounding like Perak bumpkins. Mostly people can do this without making mistakes, but as there are only very few words containing **/ɣ/** in the standard dialect, it is not too difficult to find people hypercorrecting in those few cases where there is *supposed* to be a velar fricative. So, if somebody from Perak pronounces **/lorat/** 'accent' instead of **/loɣat/**, they are producing an irregular sound correspondence (at least in their own speech) as a result of hypercorrection.

11.3 CONVERGENT LEXICAL DEVELOPMENT

When words undergo *convergent development* you will also find that sounds do not have reflexes that you would have predicted from the earlier forms. What happens when two words converge in this way is that words which are largely similar in form (but not identical) and which have very closely related meanings may end up combining their shapes and their meanings to produce a single word that incorporates features of the two original words. If somebody combines the words *dough* and *cash* into the previously non-existent word *dosh*, you can say that in the speech of this person there has been convergent development of these two lexical items. Another example of this kind of change is in Bislama (in Vanuatu) where the English words 'rough' and 'rob (him)' end up as **/ravem/**, and not **/rafem/** and **/robem/** as we might have expected. The mixed word **/ravem/** covers a wide range of meanings derived from the meanings of the two original words, i.e. 'rob, be rough to, do in a rough way, cheat, exploit'.

A similar development can be found when one language copies words from another language. What generally happens is that a language copies a single word from another language. However, there are cases when words in two different languages, which are partly similar in form and which are either the same or very similar in meaning, are copied at the same time into a third language. When such words are copied, they may take on a form and a meaning that have elements from both of the source languages. For instance, in New Zealand the English word *kit* (which also occurs in the compound *kit-bag*) seems to have taken on the meaning of the formally similar Māori word *kete* 'basket', and now Pākehā New Zealanders refer to traditional Māori baskets in English also as *kits*.

11.4 SPELLING PRONUNCIATION

Another factor that can interfere with the normal course of a sound change in literate societies is *spelling pronunciation*. Not all languages have spelling systems that accurately reflect their pronunciations, and English is a good example of such a language. We are all aware of the different pronunciations of *gh* in words like *rough*, *bough*, and *aghast*. It is possible for people to pronounce a word according to its spelling rather than pronouncing it as we would expect from its history. For instance, in English **/*sj/** sequences have regularly become **/ʃ/** by a process of phonological fusion, as shown by words such as the following:

		English	
*sjuə	→	ʃuə	'sure'
*sjʊgə	→	ʃʊgə	'sugar'

Earlier, words like *suit*, *consume*, *sue*, and so on were also (as we would have expected) pronounced with /ʃ/. However, most of us now pronounce these with /s/ because of the influence of the spelling system. This has therefore produced an irregular set of reflexes in English of earlier /*sj/ sequences.

Another example of the same kind of development can be found in the Bislama language of Vanuatu. By the normal changes of the language, we would have expected the English word *country* to have ended up in Bislama as /kantri/, and this is indeed the shape of the word that we find in the closely related dialect of Melanesian Pidgin that is spoken in Papua New Guinea (i.e. Tok Pisin). However, many people in Vanuatu now pronounce the word instead as /kauntri/. This appears to be because people are pronouncing the word on the basis of their knowledge of how the word is spelt in English (even though its spelling in English does not reflect its actual pronunciation).

11.5 LEXICAL COPYING

Most books on linguistics refer to *borrowing* when one language incorporates a word from one language and adapts it to fit the phonological structure of another language. In this book I follow the preference of the linguist William Thurston in speaking instead of lexical *copying*, as this more accurately reflects what happens.

Some linguists may find this practice a little contrived, preferring to continue to use the traditional term *borrowing*. Certainly, the notion of lexical copying should not be taken too far, or we would be forced to refer to *copy-right languages* instead of *source languages*, which is not my intention.

Lexical copying is another factor that can cause sound correspondences between two languages to show up as irregular or unpredictable. It is possible for a language to copy a cognate form from another language which has undergone different sound changes to its own words. If a sufficiently large number of words have been copied into a language, it sometimes becomes difficult to establish what the correct sound correspondences should be. Another result of lexical copying is that sometimes a single word in a protolanguage may appear to have two reflexes, both of which clearly derive from the same original form.

In English, for example, the regular reflex of /*sk/ is /ʃ/, but alongside words such as *ship* and *shirt* (which correctly reflect the original pronunciation) we also find words such as *skiff* and *skirt* which are derived from the same sources. It might be tempting to say that /*sk/ sporadically became /sk/ in English, while generally being reflected as /ʃ/. However, /*sk/ did in fact regularly become /ʃ/, and the /sk/ forms were reintroduced at a later date in words from Danish (which had not undergone the same change as English had by that stage). If you were trying to reconstruct the history of English phonology by applying the comparative method, you would therefore need to

exclude *skirt* and *skiff* when you drew up your list of sound correspondences. You should not let the fact that there is a **sk = sk** correspondence between English and Danish force you to reconstruct an additional contrast in the protolanguage, as it is only the **sk = ʃ** correspondence that goes directly back to a phoneme in the protolanguage.

Sometimes when there are several different sets of sound correspondences in a number of related languages, some of these correspondences may be the result of lexical copying, rather than being directly inherited forms. While repeated (rather than sporadic) correspondences are normally taken to point to separate original forms, as you saw in Chapter 5 (as long as they cannot be shown to be in complementary distribution with other correspondences), it is possible for large scale lexical copying at different points in history to show up as separate sound correspondences. One famous case involves the Rotuman language of Fiji. Rotuman is spoken on the island of Rotuma in what is politically part of Fiji, yet it is closely related to the Polynesian languages. In addition to words that are clearly derived directly from Proto Polynesian, there are separate sets of sound correspondences between Rotuman and other Polynesian languages which suggest that there have been two waves of other Polynesian words that have been copied on a large scale into the vocabulary of Rotuman since it diverged from its sister languages.

When words are copied from languages which are unrelated, or only distantly related, this causes very few problems in recognition, as there will normally be sufficient difference in shape between the kinds of words found in both languages to make their source obvious. However, it can become very difficult to distinguish copied forms from directly inherited forms when words from one dialect are copied into another closely related dialect (as often happens in some of the smaller languages of Melanesia, for example), as these are generally very similar to each other. Look at the following examples from the Sinaugoro and Motu languages of Central Province in Papua New Guinea:

Sinaugoro	Motu	
ɣita	ita	'see'
ɣutu	utu	'lice'
ɣate	ase	'liver'
ɣulita	urita	'octopus'
tuliɣa	turia	'bone'
ɣatoi	ɣatoi	'egg'
leɣi	rei	'long grass'

From this set of cognates, there are two sound correspondences involving the velar fricative in Sinaugoro. Firstly, there is a correspondence of Sinaugoro /ɣ/ to Motu /ø/, and secondly there is a correspondence of Sinaugoro /ɣ/ to Motu /ɣ/. Clearly, however, you should be suspicious of the **ɣ = ɣ** correspondence, as there is only one example in the data. If you had more data, you would be in a better position to judge whether there is a single

example of this correspondence, or whether there are more words in these two languages that correspond in the same way. If it turns out that this is in fact a sporadic correspondence in these two languages, its irregularity could easily be explained by saying that Motu copied the Sinaugoro word /ɣatoi/ for 'egg' instead of keeping its own original word /atoi/, which no longer exists in the language. However, there is no way of deciding just by looking at the Motu word /ɣatoi/, as it looks like a perfectly ordinary Motu word.

When dealing with copied vocabulary, things can get very complicated indeed when you come to carry out the reconstruction of linguistic history. Some languages have relatively little vocabulary that is of foreign origin, while other languages have incorporated huge numbers of words from other languages. Sometimes there has been so much vocabulary entering a language from outside sources that linguists are genuinely confused about what family the language belongs to. For instance, the Maisin language of Oro Province in Papua New Guinea has been variously described by linguists as being Austronesian with considerable non-Austronesian influence, non-Austronesian with considerable Austronesian influence, and finally as a truly mixed language. The confusion has arisen because whatever conclusion we come to, we must recognise that there has been massive copying of vocabulary from some outside source.

11.6 NON-PHONETIC CONDITIONING

Another criticism that has been made of the Neogrammarian Hypothesis in more recent decades relates to the structuralist belief in the 'strict separation of levels'. Structuralist linguists in the 1930s to the 1950s held that, when we analyse the phonological system of a language, the only facts that we should concern ourselves with are purely phonetic facts. Consideration from other levels of language such as grammar and semantics should be carefully excluded when we come to working out the phonemes of a language. This view of phonology in which there is a strict separation of levels in linguistic analysis is often referred to as *autonomous phonemics*, because phonemics is supposed to be completely autonomous, or independent of all kinds of facts except facts from the same 'level' of analysis. In insisting on this rigid dichotomy between different levels of analysis, the structuralists were little different from the neogrammarians, who also insisted that only phonetic conditioning factors could be involved in the statement of sound changes.

In more recent years, some linguists have questioned, and even denied, the need for the strict separation of levels that earlier linguists insisted upon. If we allow reference to grammatical facts, for instance, we are able to state the distribution of the allophones of phonemes in a much more straightforward manner, as this allows us to use terms like *morpheme boundary* or *word boundary*. As these are grammatical rather than phonetic concepts, structuralist phonemicists were of course unable to use terms such as these.

Although modern linguistics has now developed far beyond these methods and beliefs, it is still often argued that phonological changes over time should only be stated in terms of purely phonological conditioning factors, and that sound changes are never conditioned by grammatical or semantic factors. It is indeed difficult to imagine a sound change that operates in a language only in words referring to the names of trees, or which only applies to verbs involving motion away from the speaker, so we probably can say that sound changes cannot be conditioned by semantic features. However, it seems that some languages do, in fact, provide evidence that at least some sound changes apply only in certain *word classes* (or *parts of speech*) and not in others. Such a sound change clearly involves grammatical rather than purely phonological conditioning.

Paamese is an example of a language that has undergone a grammatically conditioned sound change. There is a correspondence of southern Paamese /l/ to northern Paamese /i/, /l/ or zero. The southern varieties directly reflect the original forms in Proto Paamese with respect to this particular feature, with the northern varieties having undergone the following fairly complex set of conditioned changes:

l →
- ø / # ___ non-high V
- ø / non-high V ___ e
- ø / e ___ non-high V
- l / _____ high V
- l / high V _____
- i / elsewhere

This rule states the following:

(a) The lateral /*l/ is lost word-initially before the non-high vowels /*e/, /*a/, and /*o/, and word-medially between /*e/ and any of these non-high vowels, for example:

		Northern Paamese	
*leiai	→	eiai	'bush'
*alete	→	aet	'flat area'
*gela	→	kea	'(s)he crawled'
*melau	→	meau	'megapode'

(b) The lateral was retained unchanged when it was preceded or followed by a high vowel (i.e. /*i/ or /*u/) in any position of the word, for example:

		Northern Paamese	
*asilati	→	asilat	'worm'
*haulue	→	houlu	'many'

*gilela	→	kilea	'(s)he knew'
*teilaŋi	→	teilaŋ	'sky'
*ahilu	→	ahil	'hair'
*tahule	→	tahul	'wave'

(c) In all other situations, /*l/ changed to /i/, for example:

		Northern Paamese	
*la:la	→	a:ia	'kind of bird'
*malou	→	maiou	'kava'
*meta:lo	→	meta:io	'European'
*to:lau	→	to:iau	'northeast wind'
*amalo	→	amai	'reef'
*avolo	→	avoi	'mushroom'

The interesting point is that none of the examples of word initial changes to /*l/ that I have just given involves a verb. Verbs, it seems, are completely immune in Paamese to any changes involving initial /*l/, though the same sound changes according to the regular rules in verbs in any other position in the word (as the examples above also show). Just so you can see that word-initial laterals in verbs are retained intact, examine the following changes:

		Northern Paamese	
*leheie	→	lehei	'(s)he pulled it'
*loho	→	loh	'(s)he ran'
*la:po	→	la:po	'(s)he fell'

If these forms had obeyed the rule that I have just presented, we would have predicted **/ehei/**, **/oh/** and **/a:po/** respectively. This is therefore a clear example of a sound change that does not involve purely phonological conditioning factors, but also involves grammatical conditioning.

11.7 THE WAVE MODEL AND LEXICAL DIFFUSION

The Neogrammarian Hypothesis upon which the comparative method rests has never been free from attack. Even when it was being formulated in its most rigid form in the 1870s by Brugmann and Leskien, there were people who claimed that their position was overstated. One of the points on which the neogrammarians were criticised related to their view of how languages diverge. In Chapter 8, I discussed the notion of *subgroups* of languages within larger families of related languages. This model of language change suggests that languages undergo sudden splits into two (or more) quite different daughter languages, and that once these splits have taken place there is no longer any contact between the new languages. Each new language, it is

assumed, then continues completely on its own, undergoing its own completely individual sets of changes.

However, many scholars have pointed out that this model of language change is nothing but an unrealistic, highly idealised picture of how languages actually do change. It has been pointed out that languages seldom split suddenly. Generally what happens is that a language develops two closely related dialects which only very gradually diverge into separate languages. While these languages are slowly becoming more and more different, there is usually some degree of contact between the two speech communities, often with some kind of mutual influence between the two dialects. Even when the two dialects finally end up as distinct languages (i.e. when speakers have to learn the other speech variety as a separate system in order to be able to understand it), there is often mutual influence.

The neogrammarian model would also suggest that there are quite discrete or separate areas of linguistic uniformity within language or dialect areas. In reality, this is hardly ever the case. Languages are, in fact, heterogeneous and there are often no distinct boundaries between languages or dialects at all. A detailed study of any language area (even very small ones) will generally reveal the existence of a number of *dialects*, or local varieties of the language. However, the dialect boundaries are also often very indistinct, and it is often impossible to say where one dialect begins and the other ends.

I will now look at a particular example to show you what I mean. On the island of Paama in Vanuatu, the people speak a single language, the Paamese language, of which there are about 4000 speakers. The island itself is quite small, being only about 10 kilometres from north to south, and 4 kilometres from east to west. There are 20 villages on the island. Even within this speech community, which is tiny by world standards, there is dialect variation. Speakers of the language themselves recognise two dialects, a northern and a southern variety. These two dialects differ in the following respects:

(a) Sequences of **/ei/** and **/ou/** in the north correspond to **/ai/** and **/au/** respectively in the south, for example:

Northern Paamese	Southern Paamese	
eim	aim	'house'
keil	kail	'they'
oul	aul	'maggot'
moul	maul	'alive'

(b) The south often has /l/ where the north has /i/ or zero (as determined by the rule that I presented earlier), for example:

Northern Paamese	Southern Paamese	
amai	amal	'reef'
a:i	a:l	'stinging tree'

tahe	tahel	'wave'
mea	mela	'get up'

(c) The south has initial /g/ and /d/ where the north has initial /k/ and /r/, for example:

Northern Paamese	Southern Paamese	
raho	daho	'(s)he is fat'
rei	dai	'(s)he chopped it'
kea	gela	'(s)he crawled'
keih	gaih	'(s)he is strong'

(d) The north often has /a/ when the following syllable contains an /a/ whereas the south has /e/ in the first syllable and /a/ in the second syllable, for example:

Northern Paamese	Southern Paamese	
atau	letau	'woman'
namatil	nematil	'I slept'

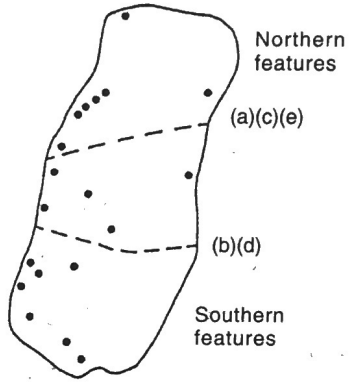
(e) The south has /m/ and /v/ when the north has the labio-velars /m^{w}/ and /v^{w}/, for example:

Northern Paamese	Southern Paamese	
m^{w}ail	mail	'left-hand side'
m^{w}eatin	meatin	'man'
v^{w}e:k	ve:k	'my sleeping place'
v^{w}akora	vakora	'coconut shell'

In addition to these phonological differences between the two dialects, Paamese speakers are also able to point to numerous lexical and morphological differences between the northern and southern varieties of the language (though I will not give examples of these as they are irrelevant to the point I want to talk about).

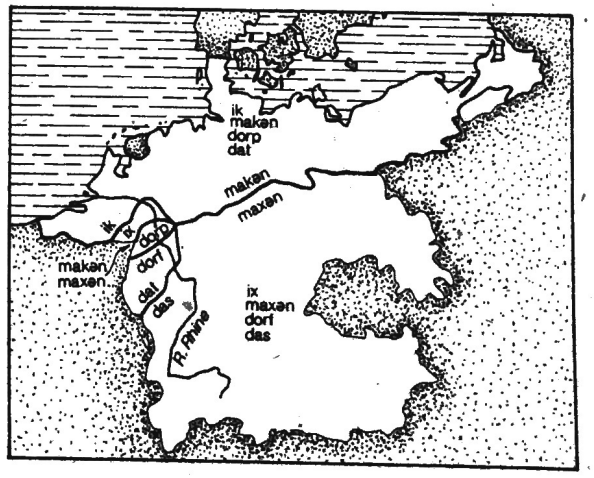
However, the picture is not nearly as simple as this. While the extreme north and the extreme south of this small island *do* differ in the ways that I have shown, it is in fact *impossible* to draw a single line that marks the boundary between the two dialects. To continue the discussion, I need to introduce the term *isogloss*. An isogloss is a line that is drawn on a map that marks two areas that differ in one particular linguistic feature. On the following map of Paama, each dot represents a single village. It is possible to draw isoglosses for each of these linguistic features. You will find that, while the northern and southern ends of the island have the features that I have indicated, the villages in the centre of the island share features from both the north and the south. So, for example, the isogloss dividing the features listed

under (a), (c), and (e) above and the isogloss dividing the features listed under (b) and (d) are located as shown in the following map.

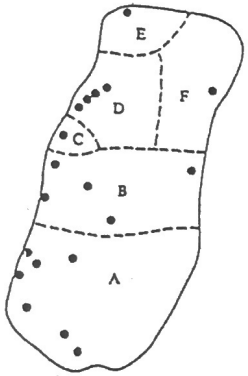


There is therefore clearly no single boundary that can be drawn between the northern and southern dialects of Paamese, as the isoglosses do not run together. This has been a very simple example because the island is so small and the number of linguistic features that I have given to illustrate the two dialects is also fairly small.

In a larger language, the situation can become much more complicated. In a language such as German, for example, there is a huge number of isoglosses criss-crossing the German-speaking area. While many of these do bunch together (to form an *isogloss bundle*), there are many other isoglosses that cross the bundle, and there are individual isoglosses that move away from the bundle in a direction all of their own, perhaps to rejoin the bundle at a later point, or perhaps to end up in a completely different part of the German-speaking area. The following map shows the Rhenish fan of isoglosses in the Dutch-German speaking area, which divides areas with fricative and stop pronunciations in words like *machen* 'make', *ich* 'I', *Dorf* 'village', and *das* 'the'.



Returning to the relatively simply example of Paamese, it turns out that even this discussion has been oversimplified, and that the real situation is more complicated. Even though I have set out a number of phonological correspondences between northern and southern Paamese, some words behave individually depending on whether they follow the stated correspondence or not. For instance, the correspondences between southern bilabial consonants and the northern labiovelar consonants (represented by /m^{w}/ and /v^{w}/) are grossly oversimplified. The reality of the situation is better shown by breaking these larger areas into much smaller areas, as set out in the following map.



These areas are characterised by the following facts:

Area A: There are no words containing labio-velar sounds, and all words contain plain labials.

Area B: There are some words containing /m^{w}/ but none with /v^{w}/. Only a few words are consistently pronounced with the labiovelar nasal, including the following: /m^{w}eatin/ 'man', /m^{w}eahos/ 'male'.

Area C: There are some words containing /m^{w}/ and a few words with /v^{w}/. These words include those listed for Area B, and also the following: /am^{w}e/ 'married man', /ti:m^{w}e/ 'friend', /v^{w}e:k/ 'my sleeping place'.

Area D: There are some more words with /m^{w}/ and several more with /v^{w}/, including the following: /m^{w}eas/ 'dust', /rom^{w}eite/ 'top', /um^{w}e:n/ 'work', /v^{w}eave/ 'cottonwood/, and /v^{w}aila/ 'footprints'.

Area E: More words contain each of these two sounds rather than plain labials: /m^{w}ail/ 'left-hand side', /v^{w}alia/ 'spider', /v^{w}eihat/ 'coastal rocks', /v^{w}aiteh/ 'door'.

Area F: Yet more words contain labio-velars rather than plain labials: /m^{w}ai/ 'he straightened it', /v^{w}akora/ 'coconut shell', /av^{w}e/ 'bell'.

The simple isoglosses that I drew earlier to separate the areas that *have* labio-velars from the areas that *do not* represent a gross oversimplification.

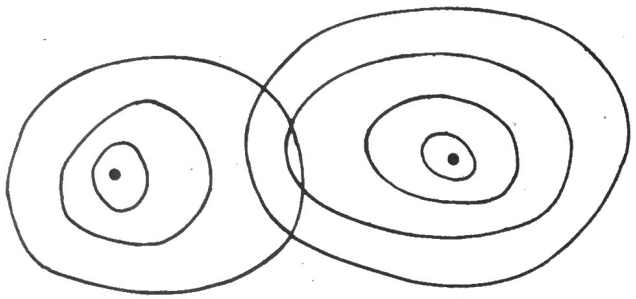
You can see that the labio-velars are more prevalent in Area F, and decreasingly prevalent until we get to Area A where there are no labio-velars at all. Which words will have labiovelars in any particular area seems to be quite unpredictable. Each word, in fact, seems to have its own behaviour. If the comparative method were strictly applied to this data, the facts that I have just described would need to be represented by recognising six 'dialects' in Paamese, with the following lexical correspondences between them:

A	B	C	D	E	F
meatin	m^{w}eatin	m^{w}eatin	m^{w}eatin	m^{w}eatin	m^{w}eatin
ame	ame	am^{w}e	am^{w}e	am^{w}e	am^{w}e
meas	meas	meas	m^{w}eas	m^{w}eas	m^{w}eas
mail	mail	mail	mail	m^{w}ail	m^{w}ail
mai	mai	mai	mai	mai	m^{w}ai

On the basis of the earlier statement that there was a northern dialect with labio-velars corresponding to a southern dialect with plain labials contrasting with correspondences between both dialects involving plain labials, we would probably want to reconstruct for Proto Paamese a contrast between labio-velars and plain labials. However, if we were strictly to apply the comparative method as I described it in Chapter 5 to the data that I have just set out, we would be forced to reconstruct six separate nasal protophonemes as there are six different sets of correspondences involving the nasals /m/ and /m^{w}/.

This brings us to the point where I should mention the French dialectologist Gilliéron. A *dialectologist* is a linguist whose speciality is the distribution of dialect features in a language. Gilliéron was a nineteenth century scholar who opposed the view of the neogrammarians, who were his contemporaries, when he made the famous statement that 'every word has its own history'. What he meant was that sound changes are not rigidly determined by purely phonetic factors, as the neogrammarians had so forthrightly stated. Instead, he said that only some words undergo a particular change, while others do not. Which words undergo a particular change can, in fact, be quite arbitrary, as you have just seen with the Paamese example. Gilliéron's view is totally incompatible with a strict application of the comparative method.

Gilliéron's view of linguistic change is consistent with what is referred to today as the *wave model*, and it contrasts sharply with the *family tree model* of change upon with the comparative method rests. The wave model implies that instead of sharp linguistic splits, changes take place like waves spreading outward from the place where a stone is dropped into water, travelling different distances with different stones, and crossing with waves caused by other stones.



Despite the success of the comparative method in reconstructing a large number of different protolanguages, the wave model of linguistic change has gained respectability in modern linguistics through recent work on *lexical diffusion*. This refers to the fact that sound changes do not operate simultaneously on every word in a language which meets the conditions for the application of a particular change. For example, if a language undergoes the devoicing of word final voiced stops, what will often happen is that final voiced stops in just *some* words will lose their voicing first, and this change will then gradually spread throughout the lexicon to other words that are of basically the same phonological shape. That is exactly what seems to be happening in Paamese. The original distinction between /m^{w}/ and /m/ is being lost, with /m/ coming to replace the labio-velar in the south. However, the change is only gradually moving through the lexicon, having affected all words in the far south, and just some words in villages further north. Over time, we can predict that increasing numbers of words in the central villages will undergo this change such that eventually the dialects of these villages will ressemble those of the far south.

11.8 DIALECT CHAINS AND NON-DISCRETE SUBGROUPS

In the previous section I indicated that dialects cannot usually be separated by single lines of a map, and that what you will find instead is that different linguistic features need to be mapped individually by means of isoglosses. While isoglosses do tend to bunch together in bundles, individual isoglosses frequently stray, making it impossible in many situations to draw a family tree diagram showing dialect relationships.

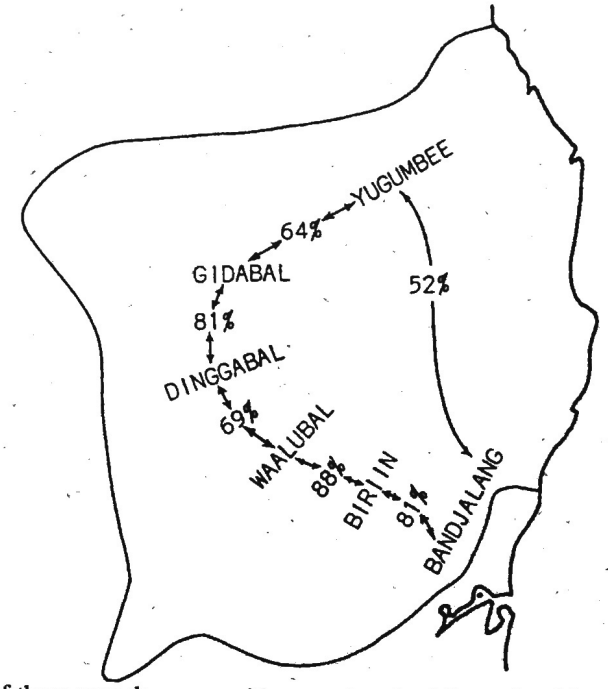
In situations where isoglosses do not bundle together closely, a different kind of case can arise, which again demonstrates a fundamental weakness of the comparative method. With dialect differences such as these, it is possible for there to be no clearly recognisable boundaries at all between one dialect and another, with dialects only gradually merging into each other.

You will note in the map of isoglosses in the previous section that the entire German and Dutch language areas were included on a single map. The

reason for this is that it is not possible to draw a single line on a map that separates the two languages. The Dutch-German political border represents a language boundary only in the sense that people on each side of this line have mutually unintelligible standard varieties. However, the local dialects of Dutch and German that are spoken on either side of the political border are little different from each other and people can readily understand each other.

What I am talking about in the case of Dutch and German is a *dialect chain* situation. Here, immediately neighbouring dialects exhibit only slight differences from each other, but as geographical distance between dialects increases, so too does the extent of difference between dialects. Eventually the point will be reached in a dialect chain where two different varieties will be mutually unintelligible, even though all of the neighbouring dialects in between are mutually intelligible.

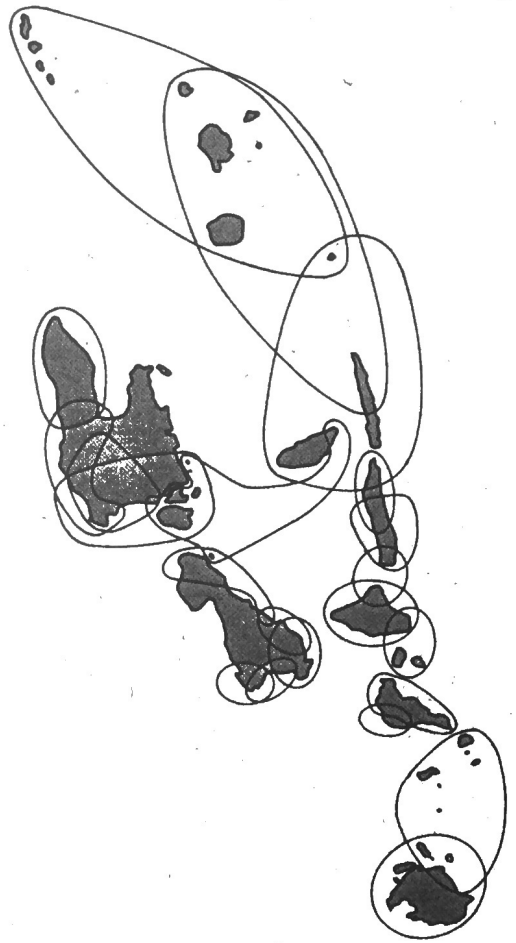
Even the languages spoken by relatively few people in Aboriginal Australia and in Melanesia commonly exhibit dialect chain features. There is an area on the border between Queensland and New South Wales where cognate counts in the basic vocabulary of a number of neighbouring speech communities are relatively high and where the two varieties are mutually intelligible. However, when we compare the basic vocabularies of the speech communities at the extreme ends of this chain, the cognate percentage drops to a level at which mutual intelligibility is not conceivable.



All of these speech communities are sharply differentiated from languages spoken outside the clearly definable area that is marked on the map, and

cognate sharing between areas on either side of this boundary is very low. Because there is mutual intelligibility between neighbouring speech communities within this bloc, as well as a sharp contrast with speech communities that clearly do not belong to the bloc, some linguists have proposed the term *family-like language* to refer to such situations.

The same principle that is involved in the phenomenon of dialect chains can extend to more distant levels of relationship as well. A lexicostatistical comparison of the languages of central and northern Vanuatu has revealed that sometimes a particular language, or a number of languages, may satisfy the criteria for membership in more than one subgroup at a time. That is, not only can we have dialect chains, but perhaps even *language chains* as well. The lines around the areas in the following map of part of Vanuatu indicate which languages appear to belong together in lexicostatistically determined subgroups, and you will see that some of the areas overlap. This means that the languages in those areas appear to belong to two different subgroups at once.



Of course, we should not place too much reliance on lexicostatistics as a method of determining subgroups, as I pointed out in Chapter 8. It may be that if we were to take into account the phonological and grammatical histories of these languages, the problem of languages that appear to belong to more than one subgroup at once might resolve itself. However, in the next chapter I discuss the fact that not only vocabulary diffuses from one language to another, but also structural features. This means that the problem of non-discrete subgroups may well be one that historical linguistics will have to learn to deal with.

READING GUIDE QUESTIONS

1. What is the basic difference between the study of etymology before the neogrammarians and in the present day?
2. What was the importance of Sir William Jones's statement in 1786 for the study of the history of languages?
3. What important contribution did Jakob Grimm make to the study of the history of languages?
4. What was the importance of Verner's and Grassmann's discoveries in the history of the Germanic languages?
5. What was the Neogrammarian Hypothesis? How did the neogrammarian view of language change differ from that proposed by Grimm?
6. How does the existence of sporadic sound correspondences affect the way that we apply the comparative method?
7. What is analogical sound change? How can it affect the way we apply the comparative method?
8. In what way can semantic or grammatical factors influence the direction of a sound change?
9. What is folk etymology?
10. What is spelling pronunciation?
11. What is meant by lexical copying? How can this cause sound correspondences between languages to become unpredictable?
12. How does the wave model of linguistic change differ from the family tree model?
13. What is lexical diffusion and how does this affect the application of the comparative method?
14. What is an isogloss? What is significant about the fact that isoglosses do not always coincide (and sometimes cross over each other)?
15. What is autonomous phonemics and what impact does the acceptance of this point of view have on the way that linguists view language change?
16. What is a dialect chain?
17. What is meant by non-discrete subgroups, and why is this a problem for the application of the comparative method?

EXERCISES

1. Papua New Guineans using English as a second language occasionally make errors such as the following in their speech:

Standard English	Papua New Guinea English
hibiscus	*hibiscuit*
pandanus	*panda nuts*
lingua franca	*lingo franco*

(Another example of the same thing, but involving only a spelling change rather than a change in pronunciation, is the change from *surname* to *sir name*.) What factor is responsible for these unpredictable phonetic changes in the English of those people who might say these things?

2. People for whom English is their first language normally pronounce the word 'gibberish' as /ʤɪbəɹɪʃ/ and 'gesture' as /ʤɛstʃə/. What factors might be responsible for the very common pronunciation of these two words by Papua New Guineans as /gɪbəɹɪʃ/ and /gɛstʃə/ respectively?
3. The English word *ambassador*, when copied into Tok Pisin, would normally have become **/embesada/**. Some speakers actually say **/embesirep/** instead. Can you say why?

FURTHER READING

1. John Samuel Kenyon 'Spelling Pronunciation', in Anderson and Stageberg (eds) *Introductory Readings in Language*, pp. 248–54.
2. Eugene Nida 'Analogical Change', in Anderson and Stageberg (eds) *Introductory Readings in Language*, pp. 86–92.
3. Leonard Bloomfield *Language*, Chapter 23 'Analogic Change', pp. 404–24.
4. Theodora Bynon *Historical Linguistics*, Chapter 4 'The Neogrammarian Postulates and Dialect Geography', pp. 173–97.
5. Otto Jespersen *Language: Its Nature, Development and Origin*, Chapters 1 to 4 'History of Linguistic Science', pp. 19–102.
6. Hans Henrich Hock *Principles of Historical Linguistics*, Chapter 15 'Linguistic Contact: Dialectology', pp. 426–71.

CHAPTER TWELVE

LANGUAGE CONTACT

There are many bilingual and multilingual societies in the world. Canada is officially bilingual, with both English and French functioning at the national level. Switzerland is officially quadrilingual, functioning in German, French, Italian, and Romansh. Other nations are more complex in their linguistic make-up, such as the former Soviet Union, India, or Indonesia, where there are hundreds of separate languages spoken. The most complex nations in the world in terms of their linguistic composition are the small Melanesian countries. Papua New Guinea boasts over 800 distinct languages, spoken by a population slightly larger than that of New Zealand (i.e. about three and a half million people). Nearby Vanuatu has only a hundred or so languages, but its population is much smaller, with the total number of people scarcely reaching 140,000!

However, just because a society is multilingual or bilingual does not necessarily mean that there is a great deal of language contact, as we can speak of language contact only when there are significant numbers of individual members of the society who are bilingual or multilingual. While Belgium recognises both Flemish and French as official languages, there is relatively little language contact as 85 per cent of the population is monolingual in either Flemish or French, and does not speak the language of the other group.

For genuine language contact to occur, there must be significant numbers of people who operate in two (or more) languages. But in world terms, monolingualism is relatively rare. This may come as a surprise to some people, especially to people from Western industrialised societies. There is a standard joke among migrants to Australia that goes like this:

Q. What is a person who speaks three languages?
A. Trilingual.
Q. What is a person who speaks two languages?
A. Bilingual.