1 The Indo-European language family

1.1 Introduction

Indo-European (IE) is the best-studied language family in the world. For much of the past 200 years more scholars have worked on the comparative philology of IE than on all the other areas of linguistics put together. We know more about the history and relationships of the IE languages than about any other group of languages. For some branches of IE – Greek, Sanskrit and Indic, Latin and Romance, Germanic, Celtic – we are fortunate to have records extending over two or more millennia, and excellent scholarly resources such as grammars, dictionaries and text editions that surpass those available for nearly all non-IE languages. The reconstruction of Proto-Indo-European (PIE) and the historical developments of the IE languages have consequently provided the framework for much research on other language families and on historical linguistics in general. Some of the leading figures in modern linguistics, including Saussure, Bloomfield, Trubetzkoy and Jakobson, were Indo-Europeanists by training, as were many of those who taught in newly founded university departments of linguistics in the second half of the twentieth century. Despite this pedigree, IE studies are now marginalised within most university linguistics courses and departments. In most US and European institutions, Indo-Europeanists with university posts do not teach in linguistics departments but in classics, oriental studies, celtic studies or the like. Historical linguistics courses may include a section on PIE, or Saussure's work on laryngeals as an example of internal reconstruction, but few students will engage in any current work on IE in any depth.

The intention of this book is not to convert general linguists to IE studies, or to restore the discipline to the central position in linguistics that it had a hundred years ago. Rather it aims to set forth some of the areas of debate in IE studies. In recent years a number of grammars and handbooks of PIE have been published in English (Gamkrelidze and Ivanov 1984 (English translation 1995), Sihler 1995, Beekes 1995, Szemerényi 1996, Meier-Brügger 2000 (English translation 2003), Fortson 2004). Most of these works are excellent, but sometimes the apodeictic style of the presentation leaves the reader uncertain about whether what is presented is actually hypothesis or 'fact'. One explanation for a historical change may be preferred over another, but the author may not make clear what is at stake in the choice between the alternatives. This book takes a different approach. It is

deliberately not intended to be a grammar of IE, or a survey of the developments that have taken place between PIE and the daughter IE languages, but rather to be a survey of some current debates and topics of more general interest in the reconstruction of PIE, and a guide to the ways in which some of these issues have been addressed. The material throughout the book is selective and illustrative, and the reader who wants to find out more will be advised to follow the further reading sections at the end of each chapter.

1.2 The IE languages

The IE language family is extensive in time and space. The earliest attested IE language, Hittite, is attested nearly 4,000 years ago, written on clay tablets in cuneiform script in central Anatolia from the early second millennium BC. We have extensive textual remains, including native-speaker accounts of three more IE languages from 2,000 years ago: Ancient Greek, Latin and Sanskrit. Also from the beginning of the Christian Era we have much more limited corpora of many more IE languages. The stock of recorded IE languages further increases as we move forward in time. In 2003, over 2.5 billion people spoke an IE language as their first language, and there were at least seventy codified varieties, each spoken by a million or more native speakers. Four hundred years ago nearly all speakers of IE lived in Europe, Iran, Turkey, Western Asia and the Indian subcontinent, but migrations have now spread speakers to every part of the world. The wealth of historical material makes IE the best-documented language family in the world.

What is it that makes an IE language IE? What does it mean to be classed as an IE language? It is usual at the opening of books on IE to repeat the famous words of Sir William Jones in 1786 which are traditionally taken to have inaugurated the discipline. Jones remarked on the similarity of Sanskrit to Latin and Greek, stating that they all bore 'a stronger affinity, both in the roots of verbs and in the forms of grammar, than could possibly have been produced by accident; so strong, indeed, that no philologer could examine them all three, without believing them to be sprung from some common source, which, perhaps, no longer exists'. Jones also noted that Gothic, Celtic and Persian could be added to the same family. Since 1786, a considerable methodology has been established to qualify and quantify Jones' notion of 'affinity' between the grammars and lexicons of the IE languages, and to work out a hypothetical model of the 'common source', PIE. But there has been no advance on Jones' criterion for relatedness between languages of the family: greater similarity in verbal roots and morphological paradigms than might be expected by chance. Languages which belong to the IE family do so either because the similarity between them and other IE languages is so strong as to be self-evident, or because they can be clearly related to languages which do obviously belong to the family. For a language which has textual remains sufficient

for the linguist to extract lexical and grammatical information, it is possible to apply the techniques of reconstruction, such as the comparative method, to build a picture of its development from PIE. However, the operation of the comparative method does not guarantee a language's place in the family; only the initial recognition that two or more languages are related can do that. (We shall return to examine the implications of this point more fully in section 1.6.)

When does a linguist decide that there is enough material to relate a language to the IE family? There is no absolute set of criteria beyond the general rule that the evidence must convince both the individual linguist and the majority of the scholarly community. A language which only survives in a very limited corpus may contain sufficient IE features to be generally agreed to be IE. As an example, take the case of Lusitanian. Lusitanian is known from a handful of inscriptions from the west of the Iberian peninsular, written in the Latin alphabet around the first century of the Christian Era. One of these inscriptions, from Lamas de Moledo in Portugal, reads as follows (the slash / signals the end of the line in the original inscription):

RVFINVS. ET
TIRO SCRIP/SERVNT
VEAMINICORI
DOENTI
ANGOM
LAMATICOM
CROVCEAIMAGA
REAICOI. PETRANIOI. T
ADOM. PORGOM IOVEAI
CAELOBRIGOI

The first four words are Latin: 'Rufus and Tiro wrote (this).' But the remainder of the inscription is not Latin. The inscription is taken to refer to the sacrifice of animals by a people called the *Veaminicori* to gods who are also addressed with their cult titles. Not all the words are understood, although the structure is clear: *Veaminicori* is nominative plural, *doenti* is a verb meaning 'they give'. The rest of the inscription has nouns in the accusative singular, denoting what is given: *angom lamaticom*, *tadom porgom*; and the names of the recipients in the dative singular: *petranioi*, *caelobrigoi*. This is not much, but enough that no Indo-Europeanist doubts that Lusitanian is a member of the IE family. Several of the word-forms are very similar to Latin. For example, the dedicated item *porgom* is very likely to mean 'pig' (Latin accusative singular *porcum* 'pig'), and *angom* to mean 'lamb' (Latin accusative singular *agnum* 'lamb'). The verb-form *doenti* 'they give' contains the root *do-* 'give', familiar from the equivalent forms in Greek (*dō-*), Latin (*da-*) and Sanskrit (*dā- / d-*). More importantly, it shows a third person plural ending *-enti* which is also found in these languages (dialectal Greek *-enti*, Archaic Latin *-nti* and Sanskrit *-anti*). Furthermore, the ending *-oi* coincides with a dative singular marker elsewhere (Greek *-ōi*, Archaic Latin *-oi*

and Sanskrit *-ai*), and the nominative plural ending *-i* accords with the nominative plural *-i* of one Latin noun declension. The interpretation of this inscription rests entirely on the identification of its language as IE, but most scholars have found it hard to believe all these similarities are entirely due to chance.

Compare with Lusitanian the case of Tartessian, another language from Ancient Spain which is known only from short inscriptions. Tartessian is better attested than Lusitanian, and from a period 600–800 years earlier. Unfortunately, we are not confident about our reading of the Tartessian script, and we do not have the helpful marks which are usually present in the Lusitanian inscriptions indicating where words begin or end. We consequently do not know a lot about the morphology of the language. However, some scholars have identified in Tartessian repeated patterns of (what they take to be) verbal endings. Consider the following inscription, reproduced in its entirety:

botieanakertorobatebarebanarkenti

The final nine letters, *narkenti*, occur elsewhere in the inscriptional corpus, as do the similar forms *narken*, *narkenii*, *narke*, *narkenai*. Here again we see a final element *-nti* that could represent the third person plural of a verbal ending in an IE language, just as in Lusitanian above. However, there is no obvious connection in the older IE languages to what would appear to be the verbal 'stem' *nark-*. Moreover, if we try to use what we know of IE morphology and vocabulary to interpret the rest of the inscription, we do not get very far. In Lusitanian, the assumption that the language was IE yielded vocabulary and morphology. In Tartessian, we have nothing more than the ending *-enti*. We do not even know enough about the morphological structure of the language to be confident that *narkenti* should be analysed as stem *nark-* + affix *-enti*. Accordingly, the general consensus is that Tartessian should not be included among the IE family.

The status of languages as IE or not may change in the light of an increase in our knowledge of the family. This is the case with the languages Lydian and Lycian, spoken in Anatolia in the first millennium BC, and known from inscriptions written in modified forms of the Greek alphabet. Before the discovery and accurate description of older IE languages in the Anatolian family, Hittite and Luwian, written in cuneiform and hieroglyphic scripts hundreds of years earlier, Lydian and Lycian could not be securely included in the IE family. However, their affinity to the earlier Anatolian languages is now patent, and since these show clear morphological and vocabulary similarities with the rest of IE, there is no doubt that Lydian and Lycian belong in the family as well. If we did not have any Anatolian languages other than Lydian and Lycian, we would not now be so certain of their ancestry. Indeed, we would not be able to make much sense of them at all, since it is only through the knowledge of how Anatolian languages are structured that headway has been made with the interpretation of the surviving inscriptions. It is, consequently, conceivable that a language such as Tartessian could come into the IE fold, if we were to have some intermediate steps to show the link between the rest of the family and the inscriptional remains that we have.

1.3 The branches of the IE tree

It follows from the remarks about Lydian and Lycian that the sub-families of IE are vitally important in determining the membership of the family. Whereas the affinity of the oldest IE languages declares itself as stronger than could be produced by chance (to most of those who study them), the affinity of languages attested more recently is sometimes only discernible through first relating them to sub-families of IE. Thus, to take an example of two languages at the far ends of the historical IE speech area, Modern Irish and Sinhala would not strike a linguist who was fluent in each, but unacquainted with their history, as *necessarily* related. It is only through relating Modern Irish to Old Irish, and Sinhala to Sanskrit, that the connection between the two languages becomes clear.

The majority of IE languages currently spoken belong to six large sub-groups of IE. Modern Irish and Old Irish are members of the Celtic sub-group, which also includes Welsh, Scots Gaelic, Breton, Cornish and Manx. Sinhala is part of the large Indic family, comprising most of the languages currently spoken in North India and Pakistan, Sanskrit and the Middle Indian Prakrits. English is a member of the Germanic branch; this includes Dutch, German and the Scandinavian languages among living languages, as well as earlier stages of these languages, such as Old English, Old High German and Old Norse, and other extinct varieties such as Gothic, once spoken in south-east Europe and southern Russia. The other large sub-groups are Romance and Slavic in Europe, and Iranian in Asia. All of these sub-groups of IE were themselves recognised as linguistic families before Jones' identification of the larger IE family cited above. The traditional criterion for grouping these languages was, in general terms, analogous to the criterion Jones used for IE. The members of a sub-group are so much more similar to each other than they are to other IE languages that the similarity cannot be put down to chance. Now, however, there are firmer criteria for membership of a sub-group. Two languages grouped together in a sub-group are assumed to have derived from a language, the 'sub-group parent language', which is chronologically earlier than either of the grouped languages, but which was spoken after PIE. The relationship can be represented diagrammatically as a family tree, with the historically prior languages situated at higher nodes in the tree. In figure 1.1, languages A and B constitute a sub-group, since they derive from a single language intermediate between them and the parent. Languages C and D do not constitute a sub-group between each other or with either A or B.

The family tree model has been very influential in IE studies, and we shall consider it in more detail below. In some cases, as in the Romance language sub-group of IE or the Indic sub-group, we have records of an early language variety which either can be identified with the sub-group parent, or which is very close to the sub-group parent (Latin and Sanskrit in the two cases respectively). But for some other sub-groups we do not have an attested parent, and it has to be reconstructed using the comparative method. It is now generally agreed among

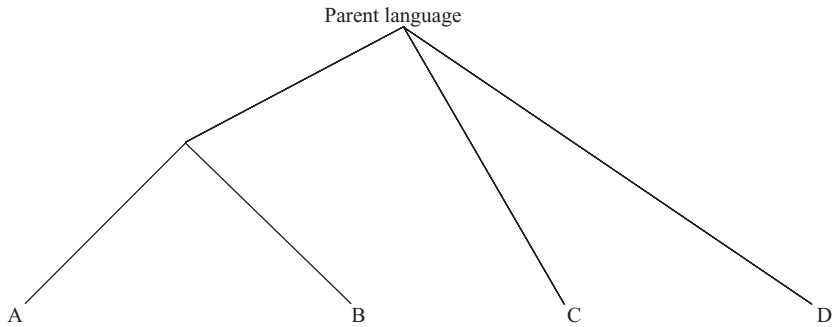

Figure 1.1 *A language family tree*

linguists that the most certain sub-groups are constructed on the basis of unique shared morphological innovations. That is, where there is no attested parent for a group of languages, they may be reckoned to belong to the same sub-group if they share a significant number of new developments in their morphology, particularly inflectional morphology. If, for example, two languages have constructed a new morphological category with a new morphological marker, and the marker is not found in other IE languages, this is reckoned to be a significant morphological innovation. It is only through morphological changes of this sort that we can be sure that there is a reconstructed sub-group parent: lexical and phonological developments are too easily shared through linguistic convergence, and we do not generally have enough information about reconstructed syntax to be certain that syntactic changes are innovations.

Using this methodology of sub-grouping it is possible to identify further sub-groups of Indo-European beyond the six large sub-groups identified above. Lithuanian and Latvian are only attested from the Early Modern period, and together with the now extinct Old Prussian they form the Baltic sub-group. Two sub-groups are no longer extant: Anatolian, mentioned in section 1.2 above, which was widespread in central and western Anatolia before the Christian Era, and Tocharian, known from the textual remains of two separate languages (now known as Tocharian A and Tocharian B) spoken in central Asia in the sixth to eighth centuries AD. Sub-grouping methodology also makes it clear that the Indic branch and the Iranian branch are more closely related to each other than to any other branch, and they are now recognised as an Indo-Iranian sub-group. Baltic and Slavic are usually also assumed to stem from a single Balto-Slavic branch, but in this case we cannot be so sure, since the languages are attested so much later.

A few IE varieties still spoken are not allocated to sub-groups, but are usually represented as separate ‘branches’ of the IE family tree. The languages in question are all spoken around the Eastern Mediterranean: Greek, Albanian and Armenian. Greek, as we have seen, has a long history, but the other two languages are more recent: Armenian dates from the middle of the first millennium, Albanian

from the second millennium of the Christian Era. Greek, Albanian and Armenian are thought by some scholars to comprise a 'Balkan IE' sub-group, but this hypothesis is disputed, since Albanian and Armenian have undergone so much linguistic change that their morphological developments are difficult to identify with confidence. Finally, there are varieties of IE no longer spoken which are not securely allocated to sub-groups. These are sometimes called 'fragmentary IE languages', since most are known from only a small corpus of material. Lusitanian, discussed in section 1.2 above, is an example of such a language.

It is a curious paradox of IE linguistics that the languages which are attested earliest are often the most difficult to assign to any sub-group. Of the IE languages spoken today, only Greek, Armenian and Albanian do not have close relatives in the same way that English compares to Dutch and German, or French to Italian and Spanish. Two thousand years ago, the linguistic map was different. Many of the languages spoken around the Mediterranean in 500 BC were superseded by Latin and its descendants following the Roman Conquest. As far as we can tell from the scanty textual remains of these languages, most were independent branches of IE, and not part of a sub-group. Lusitanian is one example of such a language, and Messapic provides another. Messapic is the name given to the language of around 300 short inscriptions from the heel of Italy, which were written in the Greek alphabet between the fifth and second century BC. Like Lusitanian, it is generally recognised to be IE, but it is not securely associated with any other IE language. The difficulty of assigning Messapic to any branch of IE is not just a problem of interpretation of a scanty corpus; the language shows significant divergences from the IE branches which are attested closest to it: Greek, Latin and the Sabellian languages of Italy, and Albanian. Other scantily attested Mediterranean languages which do not fit into a sub-group include: Phrygian, attested in central Asia Minor in two different varieties (Old Phrygian, from the eighth to the fourth century BC, and New Phrygian, from the second and third century AD); Venetic, attested in north-east Italy in nearly 300 short inscriptions from around the sixth to the second century BC; Thracian, the name given to the language of a text of sixty-one letters inscribed on a gold ring found at Ezerovo in Bulgaria and some short inscriptions on coins. Of the languages attested in the last 200 years, the only good candidates for a new branch of IE are the Nuristani languages spoken in remote valleys in eastern Afghanistan, which are thought to represent a third branch of the Indo-Iranian sub-group beside Indic and Iranian.

Table 1.1 is intended to illustrate the point about sub-groups; it shows first attestations of language and language groups by date and place, dividing the IE speech area into four different zones. Northern Europe comprises the area north of the Alps stretching from Ireland in the west to the Urals. The western Mediterranean comprises Spain, southern France and Italy. The eastern Mediterranean comprises Greece, Anatolia and the Black Sea area. The fourth zone includes Asia east of the Urals, the Indian sub-continent, and Iran and neighbouring countries to the east. The table gives the first appearance of languages in lower case and IE sub-groups or languages which represent independent branches of IE in

Table 1.1 *IE languages by date and place of first attestation.*

Date	Northern Europe	Western Mediterranean	Eastern Mediterranean	Iran / Central Asia / India
1800 BC			Old Hittite (ANATOLIAN)	
1400 BC			Mycenaean Greek (GREEK) Mittani (INDIC)	
500 BC		Latin (ROMANCE) South Picene (SABELLIAN) VENETIC Lepontic (CELTIC) MESSAPIC	PHRYGIAN THRACIAN MACEDONIAN	Old Persian (IRANIAN)
1 AD	LUSITANIAN			
500 AD	Rune inscriptions (GERMANIC)		ARMENIAN	
1000 AD	Old Church Slavonic (SLAVIC)			TOCHARIAN
1500 AD	Old Prussian (BALTIC)	ALBANIAN		
2000 AD	NURISTANI			

SMALL CAPS. The information in the table relies on dated texts, which means that the Indic family is attested first through the existence of some personal names and words relating to horse-training which occur in Hittite, Hurrian and Babylonian records from 1400 BC on, and not through the orally transmitted Vedic hymns. A similar problem surrounds the dating of the Iranian languages: Gathic Avestan, the language of the central portion of the sacred books of the Zoroastrians, certainly reflects an earlier stage of Iranian than the Old Persian inscriptions, but its transmission history does not allow us to date it securely. In the table, once one member of a sub-group is attested the sub-group is not recorded again, even when later representatives of the family occur in a different zone.

The order of attestation of different languages is reliant on the transmission of scripts and literacy. Unfortunately, the social and cultural changes which brought about an increase in literacy in much of the area where IE varieties are spoken also led to the spread of a few dominant languages at the expense of others. Table 1.1 shows the effect this has on the attestation of different languages. In the western and eastern Mediterranean zones at the onset of literacy in the first millennium BC a number of different languages are attested. In the early centuries of the Christian Era most of these languages were replaced by Latin and Greek and their descendants. The spread of these languages, and of the other

large sub-groups, is not surprising. Most of the area where the IE languages are spoken are classic 'spread zones' in the terminology of Nichols (1992). That is to say, they are areas where large-scale population movement is possible, and where one social group may readily achieve dominance over its neighbours. The IE languages for which we have fairly extensive records from before 1000 AD – Latin, Greek, Germanic, Iranian and Indic – have been the carriers of cultures which have in time predominated over other indigenous groups, with resultant language shift. Populations which once spoke Messapic, Venetic and Lusitanian eventually shifted to speaking Latin, Phrygians adopted Greek and Thracian lost out to overlapping waves of Greek, Latin, Germanic (Gothic) and Slavic. In the Mediterranean area, the early adoption of literacy allows us to know of a range of IE varieties. In northern and eastern Europe, where the first written records appear considerably later, we do not know whether there was a similar diversity in the territories later occupied by speakers of Celtic, Germanic, Slavic and Baltic languages. We shall consider further the question of how we can assess the evidence for the early relationship of the IE family, considering what we have lost, in the next section.

1.4 Cladistics: constructing family trees

The family tree model of IE is over 150 years old. The model was first put forward in the nineteenth century, and the first tree diagram was produced by the German Indo-Europeanist August Schleicher (reproduced in figure 1.2). Schleicher's tree does not include Armenian, which was not then recognised as a separate branch of IE, nor Anatolian or Tocharian, which were not then known. As our understanding of the IE languages has increased and changed, so also the tree has changed. In Schleicher's tree, the first split is made between Germanic, Baltic and Slavic and the other language groups. This split reflects the fact that the three sub-groups spoken in the north of the IE area form dative-ablative and instrumental plural cases in some noun paradigms with a marker involving the original phoneme **m*, whereas the other languages use a marker with **b^h*, as shown in table 1.2, which gives the instrumental plural markers in various IE languages (note that all reconstructed, as opposed to attested, sounds, morphs and words are preceded by * throughout this book and in most works on PIE).

This divergence between the languages is still unexplained – it may be that the two plural cases which use **m* or **b^h*, the dative-ablative and instrumental, originally took separate markers, but some languages generalised **m* to both of them, others **b^h*. Modern scholars do not see the distinction between the use of **b^h* and **m* in these cases as sufficient evidence for a fundamental split between two parts of the IE language family. Furthermore, there are other features which unite the languages of the western IE zone: Celtic, Germanic and Latin and the Sabellian languages. In constructing a family tree, the shape of the tree depends on what the linguist sees as important.

Table 1.2 *Instrumental plural markers in various IE languages.*

PIE **-mis*	Germanic: Gothic *-m*
	Slavic: Old Church Slavonic *-mi*
	Baltic: Lithuanian *-mis*
PIE **-b^{h}is*	Indo-Iranian: Sanskrit *-bhis*
	Greek: Mycenaean Greek *-pi* (/-p^{h}i/)
	Celtic: Old Irish *-^{i}b*

In recent years, the advance of statistical techniques and the use of computers to process very large amounts of data have allowed the construction of family trees from a much wider data set and a resurgence in interest in drawing a family tree for IE. Since computer analysis allows for such a large amount of discrete data to be handled, trees can be constructed using hundreds of different features. The new technology brings with it a new terminology, and now linguists are beginning to talk not of family trees, but *phylogenies*, and to use the term *cladistics* for referring to the techniques of constructing family trees. Two recent phylogenies of PIE are given in figure 1.3 (the 'New Zealand' tree constructed by Gray and Atkinson 2003: 437) and figure 1.4 (the 'Pennsylvania tree' taken from Ringe, Warnow and Taylor 2002: 90). The two phylogenies use different features in order to rank languages against one another.

The New Zealand tree, figure 1.3, relies upon vocabulary items only, following in a long tradition of language surveys which rely upon word lists or 'basic

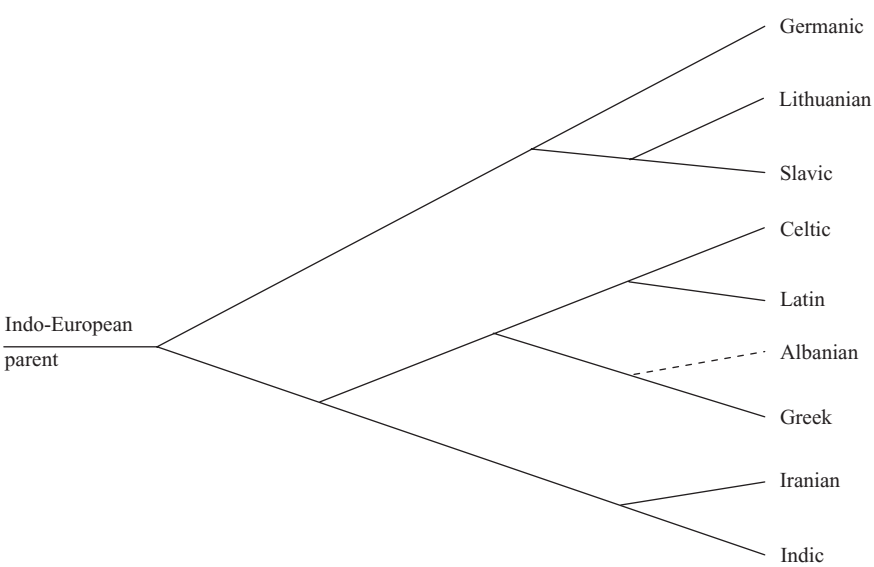

Figure 1.2 *Schleicher's Indo-European family tree*

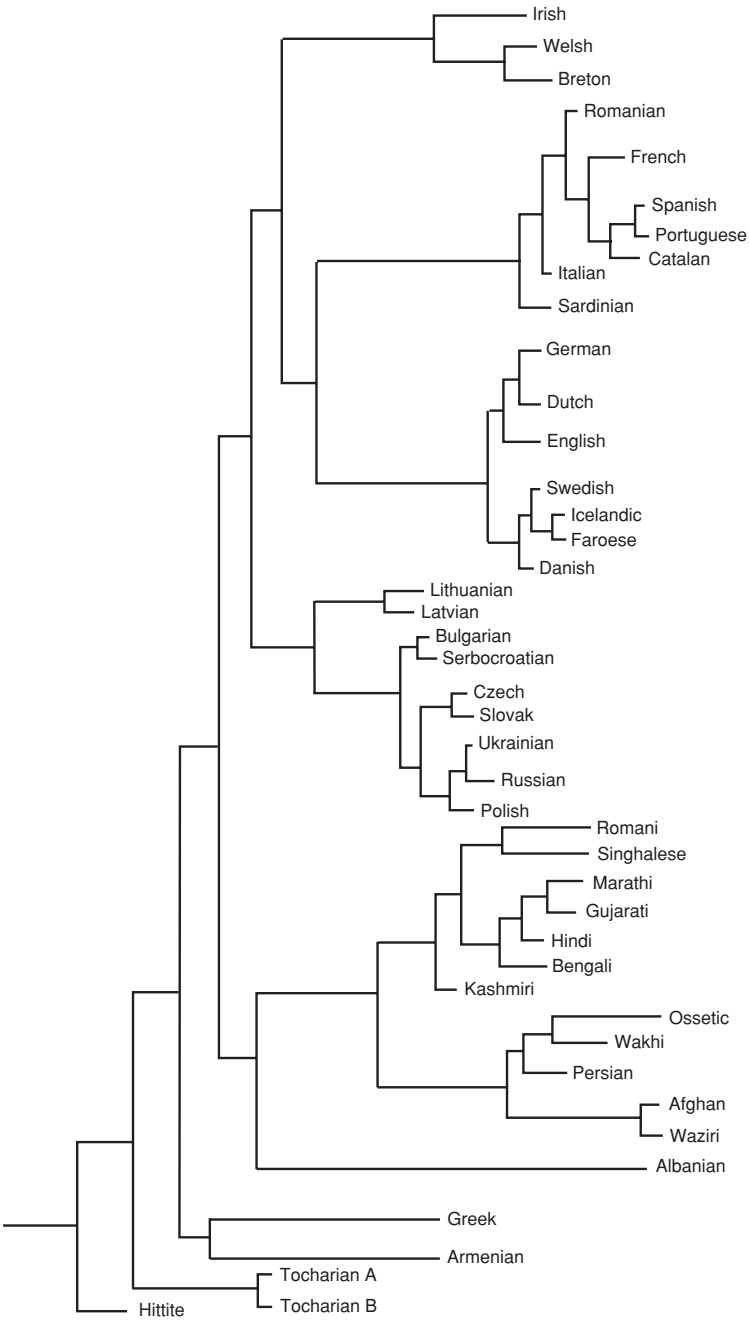

Figure 1.3 *The New Zealand family tree*
Adapted by permission from Macmillan Publishers Ltd: *Nature* 426 (2003)

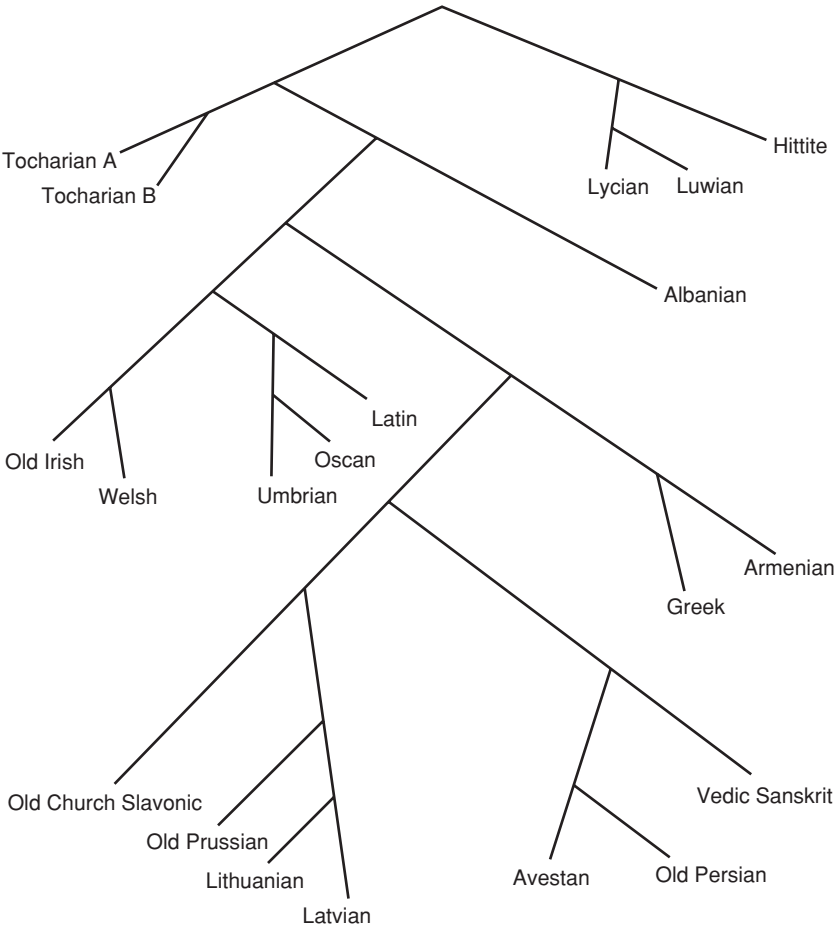

Figure 1.4 *The Pennsylvania family tree*
Reprinted by permission of Blackwell Publishing Ltd: *Transactions of the Philological Society* 100 (2002)

vocabulary' lists. As we have discussed in section 1.2 above, vocabulary alone has not generally been considered as sufficient for constructing sub-groups of languages. However, the ease with which the vocabulary data can be recovered and used in computer analysis makes it a very amenable data set. The use of the 'basic vocabulary list' originally compiled by Swadesh without the IE language family specifically in mind means that the selection of data is neutral: the New Zealand team cannot be accused of pre-selecting the data with a certain outcome in mind. Gray and Atkinson's tree also applies techniques first used by geneticists working on the cladistics of DNA sequences to assign dates to the divisions in the family. Their dating is still controversial, and we shall discuss the dating issue in more detail in the next section.

The 'Pennsylvania tree', figure 1.4, is constructed on the basis of compared features over a much larger range. Languages are compared not just through

vocabulary items, but also phonological and morphological features. While the New Zealand tree takes lexical data principally from the modern spoken forms of the languages, the Pennsylvania tree uses the earliest attested languages for information. Ringe has used his own extensive knowledge of the IE family to select items which have already been reckoned to be diagnostic for sub-grouping, and features can be marked as innovations in the computer sorting process. Furthermore, some features, such as morphological innovations, can be given extra weight, whereas others, such as lexical agreements, carry less weight in the tree. The Pennsylvania tree is made utlising the scholarship on the IE languages, the New Zealand tree is not.

If we compare the New Zealand tree of IE with the Pennsylvania tree, we see that they share some fundamentals on the interrelationship of the IE languages. In both models, the first split in the tree is between the Anatolian group of languages and all the others, and the second is between Tocharian and the rest of the family. This is in accordance with the views of the majority of Indo-Europeanists at present. Anatolian is radically different from the rest of the family in many respects, and much of the rest of this book will be concerned with looking for an explanation of these differences. However, once we proceed beyond the first splits in the tree, there appear to be striking differences between the trees; note in particular the position of Albanian and the Germanic group. In the New Zealand tree, Albanian is grouped closest to the Indo-Iranian languages, forming a separate branch with them. Germanic forms a branch with the Romance languages, with Celtic as an outlier group. The Pennsylvania tree sees no close connection between Albanian and Indo-Iranian. Germanic is omitted from the tree, since there is no best-fit tree with Germanic, but Italic and Celtic are closely linked on the tree.

The existence of these discrepancies is not in itself reason to cast doubt on the exercise of drawing up trees through phylogenetic techniques. The New Zealand tree and the Pennsylvania tree actually assess different things: the first is a measure of the affinity of lexicons, not of grammatical systems, and the second tracks a range of innovations and changes across families. It is perfectly possible for both phylogenies to be correct. The relationship of vocabularies may not be the same as the participation in linguistic changes. The modern Germanic languages have vocabularies which have been influenced by and which have influenced the modern Romance languages, and this is reflected in the closeness of the languages on the New Zealand tree. The difficulties that arose in the construction of the trees are themselves instructive. As already mentioned, for the construction of the Pennsylvania tree, Germanic was removed from the analysis in order to provide a ‘best fit’. This may point to unusual changes in the prehistory of Germanic, and possible convergence with other IE varieties. Comparing the cladistics in this way may alert researchers to potentially problematic or interesting areas of linguistic development and allow the construction of hypotheses to account for why the divergences exist.

However, we have seen in the preceding section that many of the scantily attested IE varieties which appear early in the historical record, languages such as

Lusitanian and Messapic, do not fit well into the language sub-group and family tree model. Often we do not know enough about these languages to include them in the phylogenetic analyses. We cannot draw up a basic vocabulary list or know for certain which innovations they have or have not undergone. Yet these fragmentary languages may be representative of a much wider range of IE varieties which have been lost beneath the spread of the big IE sub-groups. Is there any way we can include these languages in the model of PIE?

Recent work by Andrew Garrett (Garrett 1999 and 2006) has used the fragmentary languages of the Mediterranean area to rethink the validity of the family tree model for the early stages of the IE language family. Furthermore, Garrett's work calls into question the reconstruction of sub-groups in the west of the IE speech area. Take the example of Celtic. We have so far discussed the Celtic languages as a unit, with Lepontic, recorded in inscriptions in northern Italy from around 600 BC, as the earliest branch of the language to be attested. Other members of the sub-group attested before the Christian Era include Gaulish, known from texts in France, and Celtiberian, known from a few inscriptions in central Spain written in a modified form of the Iberian script. These languages share some characteristic lexical features with the other members of the Celtic group, principally Irish and Welsh, and the loss of the sound **p*, generally assumed for all the Celtic languages. However, the number of morphological innovations which are shared by all the Celtic languages is extremely small, and if we use the strictest criteria for reconstructing sub-groups, the Celtic languages do not qualify. Even the loss of **p* seems only to be underway in the Lepontic inscriptions and may have spread across the whole language area from language to language, rather than being a feature of a sub-group parent from which they all descended. It is therefore possible that other sub-group parents as such do not exist, but that sub-groups actually represent later convergence of closely related languages.

Garrett further underlines his doubt of the existence of sub-group parents by detailed consideration of the history of the Greek language. Greek is first attested from 1400 BC in the Linear B texts, in a phase of the language termed 'Mycenaean Greek'. After the collapse of Mycenaean society at around 1200 BC, there is no direct evidence for the Greek language until around 800 BC, when the first inscriptions written in the Greek alphabet appear (this stage is sometimes called 'alphabetic Greek'). The Greek we have preserved from Mycenaean times is largely uniform, but the later Greek texts show considerable dialectal variety by region and also by literary genre. Although we do not have clear evidence for different dialects attested within Mycenaean Greek, we know that there must have been other dialects spoken, although not written down. This is because Mycenaean Greek has undergone some changes, such as the phonological change of original **-ti* to *-si*, which are shared with some later Greek dialects but which did not take place in others.

By the family tree model, all of the different varieties of Greek, including Mycenaean, would appear as branches off a single node, the sub-group parent, which is usually called Proto-Greek or Common Greek. In his most recent paper,

Garrett draws up a set of features which could be assumed for Proto-Greek on the basis of the alphabetic Greek varieties. In doing this, Garrett performs the same exercise for the Greek sub-group that other scholars have attempted for other sub-groups, such as the Sabellian languages of central Italy, where the very earliest texts are attested at around the same date. However, for Greek there is the advantage that the features assumed for Proto-Greek can actually be compared with a language of the second millennium BC, Mycenaean Greek. We know that Mycenaean cannot be equated with Proto-Greek, since it has undergone some changes shared only with some later Greek dialects, and so it must be later. Yet all of the distinctive morphological features, and many of the distinctive phonological features, which are assumed to be distinctive for Proto-Greek can be shown not to have taken place at the time of Mycenaean. Wherever the later Greek dialects have made innovations in morphology from PIE, Mycenaean Greek appears not to have participated in that innovation. In other words, the distinctive aspects of the later Greek dialects (which they all share) arose across a number of varieties which already were distinguished one from another. It is not possible, using the shared morphological innovation criterion, to construct a unified invariant entity such as 'Proto-Greek' which is distinguishable from PIE. Certainly, Mycenaean Greek shares many specific lexical features with the later varieties, and a few phonological changes which are distinctively Greek must predate Mycenaean. But if we had more evidence for other IE languages other than Anatolian contemporary with Mycenaean, we might not be able to separate out what was 'Greek' about Mycenaean from its neighbours. The Greek sub-group was only truly formed in the period after Mycenaean, when convergence between the different dialects of Greek took place, in part related to social changes coupled with a strong sense of Greek ethnic identity. It is worth quoting Garrett's conclusion about sub-grouping of the IE languages in full, with the caveat that he is discussing IE languages apart from the Anatolian sub-group, which he accepts has branched off earlier from the rest of the languages:

> If this framework is appropriate for IE branches generally, we cannot regard IE 'sub-groups' as sub-groups in a classical sense. Rather, the loss or 'pruning' of intermediate dialects, together with convergence *in situ* among the dialects that were to become Greek, Italic, Celtic and so on, have in tandem created the appearance of a tree with discrete branches. But the true historical filiation of the IE family is unknown and perhaps unknowable. (Garrett 2006)

1.5 The time and place of PIE

The existence of the IE language family presupposes a parent to the family, as William Jones realised in the quotation given in section 1.2. The similarities between IE languages cannot be explained through convergence or borrowing from one language to another. The parent language, PIE, can be reconstructed

through the comparative method, as we shall see in the rest of this book. But the existence of a language implies speakers. Where did they come from, and when did they exist? These two questions are entirely natural, and the answer to them has been sought by linguists, archaeologists and others for the last two hundred years. Since this book is primarily concerned with linguistic rather than archaeological material, we shall not attempt to give an answer to these questions here (but see further section 7.5). It is, however, worth examining some of the linguistic assumptions which feature in the debate.

Firstly, it is useful to distinguish between the hypothetical, reconstructed 'language' which is the result of the operation of the comparative method on the IE languages, what we shall call 'reconstructed PIE', and the unattested spoken language from which we presume all the IE languages derived, which we shall refer to as 'the spoken IE parent language'. Reconstructed PIE may have some features in common with the spoken IE parent language, but it is not the same as it, and is not a real language. Reconstructed PIE is a construct which does not have an existence at a particular time and place (other than in books such as this one), and is unlike a real language in that it contains data which may belong to different stages of its linguistic history. The most helpful metaphor to explain this is the 'constellation' analogy. Constellations of stars in the night sky, such as The Plough or Orion, make sense to the observer as points on a sphere of a fixed radius around the earth. We see the constellations as two-dimensional, dot-to-dot pictures, on a curved plane. But in fact, the stars are not all equidistant from the earth: some lie much further away than others. Constellations are an illusion and have no existence in reality. In the same way, the asterisk-heavy 'star-spangled grammar' of reconstructed PIE may unite reconstructions which go back to different stages of the language. Some reconstructed forms may be much older than others, and the reconstruction of a datable lexical item for PIE does not mean that the spoken IE parent language must be as old (or as young) as the lexical form.

Indo-Europeanists have attempted to construct models for PIE which bring the language into a closer approximation with the spoken parent language. The most influential model of this type is the 'Space-Time' model originally drawn up by Meid (1975). According to Meid's model, the spoken IE parent language existed over time and space. As time progresses, the number of speakers increases and the language spreads over a larger area. Hence one can draw a triangular representation of the language, as in figure 1.5, with the dispersal of the language over space plotted as the horizontal axis, and the time-scale plotted on a vertical axis.

Meid originally plotted reconstructed PIE onto this model, dividing up the language into three stages (Early, Middle and Late), and attempted to assign features and relative chronologies to different stages, and indeed to plot real dates on the time-line. There are two problems with this. Firstly, the technique of comparative reconstruction has the aim of reducing variation, by giving a single ancestor to phonemes, morphemes or words which differ in daughter languages. The method favours the reconstruction of everything to a single point, and it is not always clear on what grounds the linguist can separate out different features into

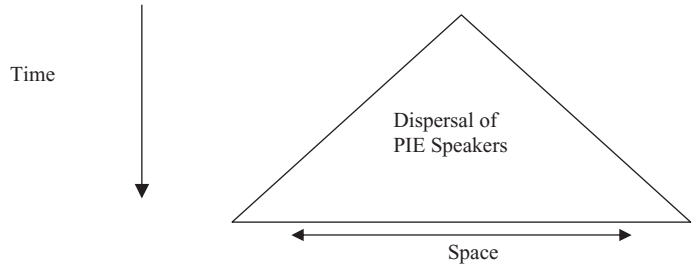

Figure 1.5 *Meid's Space-Time model*

different stages. Secondly, dating the model in real time is obviously problematic, since it is not clear how one can date a feature such as a reconstructed case marker or verbal paradigm, although it may be possible to assign some absolute dates to items of material culture, such as wheels or the terminology for spinning wool.

Meid's model also relies on various underlying assumptions about the expansion of the speakers of IE. They are thought to have spread over ever greater space and to have come into existence in a vacuum. Indeed, Meid explicitly states that at the earliest stage of IE there would have been no dialectal variation, because the speakers were most probably a very small, tight-knit band. The Space-Time model is extrapolating backwards from the spread of IE speakers in historical times and in recent prehistory, and assuming that since they expanded, they can be projected back to a single point. There is also an attempt to conflate reconstructed PIE, which is reduced to an invariant linguistic system as a consequence of the comparative method, and the spoken IE parent language, which is made to go back to a dialect-free stage in order to comply with the reconstructed language.

The Space-Time model is now looking distinctly unstable, and the idea that the spoken IE parent language was once invariant and spoken in a single place and time is far from certain. The assumption that in linguistic prehistory individual languages spread by fanning out over a wide area has been challenged by work on Australian and other areas, where there have been long periods of linguistic convergence, and it can no longer be assumed that there was a prehistoric rapid expansion of speakers of IE languages. The Space-Time model is also under attack from the renewed interest in cladistics discussed in the preceding section. As most scholars now believe there was an early split between the Anatolian languages and the other IE languages, we can no longer so easily bundle all the languages together, but we must rather reconstruct two separate stages, a pre-Anatolian PIE and a post-Anatolian PIE. All this makes discussion of the location in time and space of the speakers of the IE parent language rather more complex than has been reckoned in some previous accounts.

We saw in section 1.4 that the New Zealand team which reconstructed an IE phylogeny from a comparison of word-lists of living IE languages incorporated dates into their family tree. The attempt to use the changes in the lexicon to calculate the age of language families and sub-groups is not new. Lexical change was first used as a clock to measure the age of languages by American scholars

working on basic vocabulary lists in the 1950s, in an enterprise that became known as 'glottochronology'. However, the first wave of glottochronological research has become largely discredited, since there was a simple reliance on a constant rate of linguistic change. It was easy for other scholars to demonstrate that in documented language history lexical replacement takes place at wildly differing rates across different languages and in different periods of a single language's history (as shown by Bergsland and Vogt 1962). The New Zealand team do not make the same mistakes as the glottochronologists do. They use models which were originally designed to build phylogenies based on DNA and other genetic information, which do not assume a constant rate of change. Instead, their model accepts that the rate of change varies, but it constrains the variation within limits that coincide with attested linguistic sub-groups. For example, it is known that the Romance languages all derive from Latin, and we know that Latin was spoken 2,000 years ago. The rates of lexical change in the Romance family can therefore be calculated in absolute terms. These different possible rates of change are then projected back into prehistory, and the age of the parent can be ascertained within a range of dates depending on the highest and lowest rates of change attested in the daughter languages. More recently (Atkinson *et al.* 2005), they have used data based not just on lexical characters, but on morphological and phonological information as well (reproduced here in figure 1.6). In this figure, two competing archaeological models for the spread of speakers of the IE languages are indicated on the family tree. One theory, first put forward by Colin Renfrew in 1987, relates the spread of languages to the spread of farming from the Ancient Near East. By the New Zealand team's hypothesis the likely dates for the spread of farming would relate to the split of the Anatolian sub-group from the rest of the IE family. The alternative theory, labelled the Kurgan theory, follows the work of Gimbutas and others (and is most conveniently summarised in Mallory 1989). This links the speakers of IE to nomads on the steppes of southern Russia who gain a technological advantage over other societies through the use of wheeled transport. As the figure shows, the dispersal of the IE languages, apart from Anatolian, could be made to fit with this model.

However, the findings of the New Zealand team must be used with caution. Although the use of advanced statistical techniques and complex mathematical models enables them to come up with real numbers, there is currently controversy over whether the dating given to the family should be accepted. Although the mistakes of the glottochronologists have been avoided, the correct use of different statistical models is still hotly debated, with members of the Pennsylvania team uncertain that the models used by the New Zealand team are in fact appropriate for what they want to do, and others questioning their use of data. Furthermore, the use of language changes in historical time to project back into language change in prehistorical time is itself questionable. Language change, and vocabulary change in particular, may be affected by the size of the population speaking the language and the level of shared literacy. Large literate populations may retain words better than small illiterate populations. A linguistic innovation does not need so long to take hold in a small population, especially if there is no retarding influence from

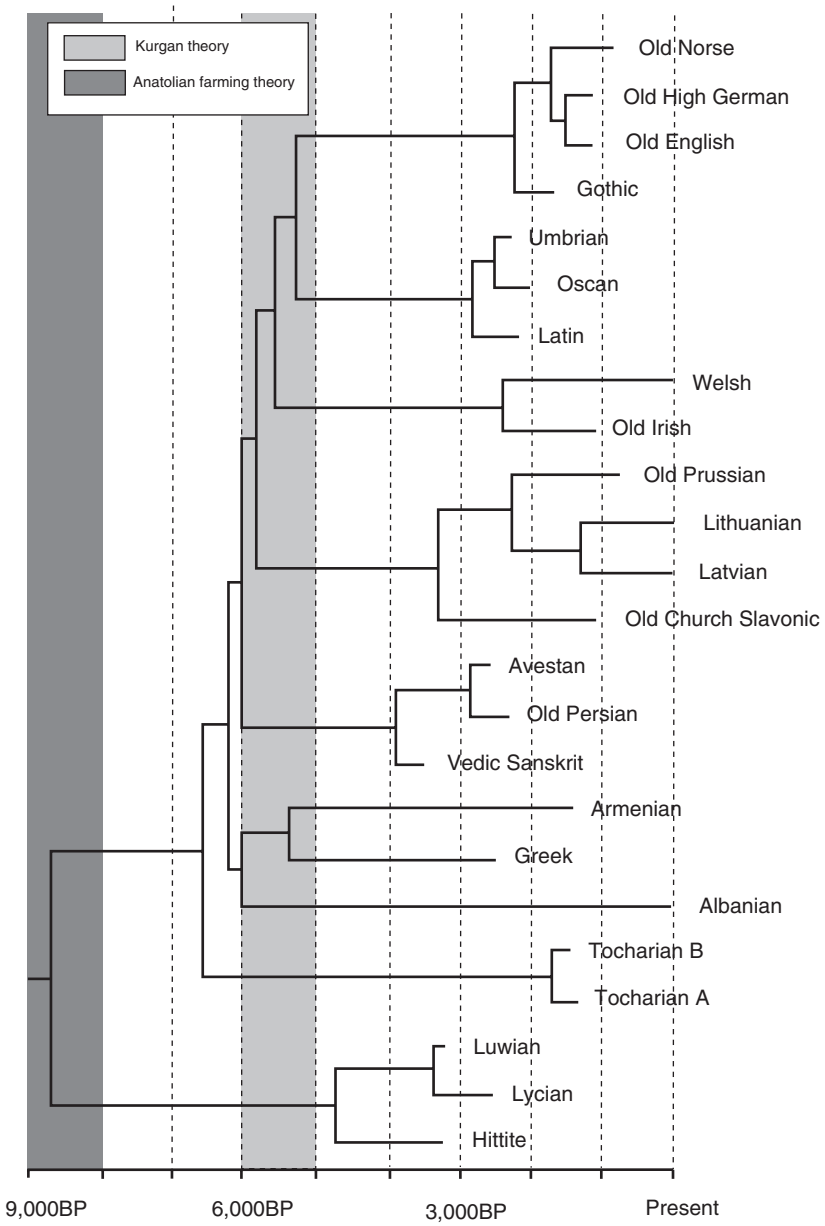

Figure 1.6 *The New Zealand family tree with dates*
Reprinted by permission of Blackwell Publishing Ltd: *Transactions of the Philological Society* 103 (2005)

the written word and education. This may have the effect of making linguistic change slower in modern languages than in languages spoken in prehistoric times, and therefore skewing the rates of linguistic change that are projected back into prehistory. For the moment, then, the jury is still out on whether phylogenetic dating can help solve the problem of how old the IE language family is.

1.6 Nostratic and other supergroups

In the history of IE studies there have been many attempts to link IE and other language families together into larger genetic groupings, which can be called language *phyla* (singular *phylum*). At the very beginning of IE studies, in the early nineteenth century, the German Indo-Europeanist Franz Bopp had attempted to connect IE with 'Malayo-Polynesian', and over the following two hundred years claims were made for links with Uralic, Afro-Asiatic, Kartvelian, Eskimo-Aleut, Ainu, Etruscan and practically every other language in the Old World, and some in the New. In recent years, the phylum known as 'Nostratic', first proposed by Holger Pedersen, has received much discussion after the reconstruction of Proto-Nostratic attempted (originally independently) by two linguists working in Soviet Russia, Illič-Svityč and Dolgopolsky, and the revised reconstruction attempted by the American Bomhard. Nostratic is generally thought to comprise the language families of IE, Uralic, Kartvelian, Afro-Asiatic, Dravidian and 'Altaic' (or, according to some, the individual groups of Turkic, Tungus and Mongolian), and some scholars include other languages from among Korean, Japanese, Ainu, Sumerian, Etruscan, Chukchi-Kamchatkan, Nilo-Saharan, Eskimo-Aleut and Gilyak (also known as Nivkh). There have also been suggestions that Nostratic is itself related to other hypothetical large language phyla such as Greenberg's Amerind or Sino-Caucasian, and there have been a few brave attempts to reconstruct some lexemes in 'Proto-World', the hypothetical ancestor of all human languages. It is beyond the scope of this book to consider all these claims in detail, but we shall briefly examine the methodological basis upon which the reconstruction of Proto-Nostratic (and, by extension, such families as Greenberg's 'Eurasiatic') is based. It is not possible to disprove the premise that IE is distantly related to language X or belongs to language phylum Y, and no one would deny that IE did not arise *in vacuo*. However, by briefly looking at the case for Nostratic, it will be seen why the reconstruction of groupings at a higher level than PIE is not very plausible given our current state of knowledge. Furthermore, even if the Nostratic hypothesis were correct, it is unlikely to be fruitful for IE studies.

It has been claimed that the methodology for reconstructing Nostratic is exactly the same as the methodology for reconstructing Germanic or IE or any language family: the comparative method. Items in different languages are compared, correspondence sets between sounds are established and the parent language is thus reconstructed. The frustration evident in many of the statements of Nostraticists is clear: they are using the same methods as IE linguists, yet their results are not accepted by most IE linguists for reasons which are seldom clearly articulated. This confusion arises from a misunderstanding about the comparative method. The (partly) successful operation of the comparative method over a non-restricted field of (open-class) vocabulary does not furnish proof that two languages are genetically related, rather the comparative method is used to reconstruct the parent

language of any two languages which are already hypothesised to be related. The hypothesis that the IE languages are related is, as we have already seen in section 1.2 above, based upon a self-evident affinity between languages. This affinity is manifested in the obvious similarity between inflectional morphemes and vocabulary in restricted semantic sets (kinship terms and numerals are two standard examples; note that these vocabulary sets are not reconstructable for Nostratic). Operation of the comparative method may elucidate the relationships between languages, but it does not 'prove' the hypothesis of relationship.

An advocate of Nostratic might object that when dealing with a language family of the time depth of Nostratic, it is unrealistic to look for the same sort of evidence as we have for IE in order to construct a hypothesis of language relationship; furthermore, it might be inappropriate to seek out the sort of morphological agreement we find in IE in a language without inflectional morphology, as Nostratic is hypothesised to have been. The Indo-Europeanist, our advocate of Nostratic might add, is in the fortunate position of having well-documented histories of most of the languages in the family. Let us consider again the example of Modern Irish and Sinhala (see section 1.2 above). If we were to attempt to compare them without any access to their history, the Nostraticist would argue that their relationship could only be unearthed through vocabulary comparison. The Indo-Europeanist might counter that, without our historical knowledge of the two languages, they cannot be shown to be related beyond reasonable doubt. The debate only serves to illustrate the gulf between the two sides.

The most effective counterblast to Illič-Svityč and Dolgopolsky's Nostratic has come from the Indo-Europeanist Ringe (see Ringe 1995 and 1999), who has claimed that the large number of vocabulary comparisons amassed by Illič-Svityč in support of Nostratic is not indicative of a genetic relationship, but in fact is equally likely to be the result of chance similarity between very broadly similar phonetic forms and (usually rather vague) meanings. Ringe plots the number of roots which are attested in the Nostratic family. In most cases, a Nostratic root appears in two separate branches of the family. A smaller number of roots appear in three branches, and the number diminishes up to a very small number of roots that appear in all six branches of the family. Ringe shows that the spread of roots across the number of language families closely parallels the expected statistical distribution if any single root from one language family has a 40 per cent probability of being matched to another root in another branch. One way of understanding this is to imagine the six language families as if they were six bags of differently coloured marbles. Finding matches for roots across the Nostratic family is analogous to finding marbles of the same colour in different bags. If the chance of finding a matching marble is 40 per cent, the distribution of the number of marbles of the same colour found in two bags, three bags, etc. approximates the distribution of roots in the Nostratic languages. Ringe also examines the distribution of vocabulary correspondences over the IE family, and here the distribution is not the same. There are more roots shared across a number of languages than would be found if there were just a 40 per cent chance of

Table 1.3 *Nostratic and PIE stop comparisons.*

PN	PIE (Moscow)	PIE (Bomhard)
**d	*dh	*d^h (= 'traditional' *dh)
**t´	*t	*t´ (= 'traditional' *d)
**t	*d	*t^h (= 'traditional' *t)

finding a match between two languages, which implies that there is a better basis for reconstructing a family.

Ringe's arguments assume that there is a 40 per cent chance of finding a parallel root somewhere else in the Nostratic phylum to a root chosen at random. This may seem a high probability if there is no genetic relationship between the languages, but if one considers the comparisons offered in dictionaries of Nostratic, it does not look so unlikely. Root morphemes are generally short (roots are monosyllabic in PIE), and many reconstructed roots have a wide semantic range. In large language families such as IE, Uralic and Afro-Asiatic there is a huge body of data from which to draw vocabulary, and there may be a number of roots reconstructed for these families which are themselves the result of chance similarities. Furthermore, Nostraticists have tended to be lenient to some inexactitude in the phonological and semantic correspondences between roots.

The sheer weight of evidence produced in support of Nostratic has normally been the biggest argument in favour of the hypothesis. Although it may be possible to find coincidental matches for some roots, is there really a 40 per cent chance of finding a cognate to so many vocabulary items across so many language families? Strong corroborative evidence for Ringe's position comes, however, from the work of the Nostraticists themselves. There are two leading schools of Nostratic reconstruction. The reconstruction of Nostratic put forward by Illič-Svityč and Dolgopolsky and followed by the Moscow school has been revised by Bomhard in the light of a different reconstruction of PIE stops. Table 1.3 gives a comparison of the equivalences between Proto-Nostratic (PN) and PIE in the two systems (note that PIE reconstructed forms are denoted by an *, Proto-Nostratic forms by a double asterisk preceding them).

As table 1.3 shows, the PIE correspondences to PN **t´ and **t are reversed in the Moscow and Bomhard versions of Nostratic. The same reversal affects the other stop series, **p´ and **p, **k´ and **k. However, many of the same PIE roots containing the disputed sounds are given Nostratic etymologies by both the Moscow school and Bomhard; clearly, they cannot both be right. This demonstrates that, even if Nostratic is indeed a valid language phylum, it is possible to find a considerable number of false positives – i.e. matches for PIE roots which are invalid. There has not yet been any rebuttal of Ringe's claim about the distribution of Nostratic roots across language families, and if Nostratic is to gain more credence among the wider academic community, the number of convincing etymologies across languages must be increased.

The verdict given by one leading Indo-Europeanist (Watkins) that the results of Nostratic research 'even if true, are not very interesting' provoked an outraged response from a number of scholars. For the purpose of IE studies, there is a kernel of truth about Watkins' remark. The reconstruction of Nostratic at present adds nothing to our understanding of PIE, and it is difficult to see how further Nostratic research can improve this situation. Some scholars have argued that the reconstruction of Nostratic can help settle many of the existing uncertainties over the phonology of PIE, including debates concerning the number and nature of laryngeals; the reconstruction of glottalic consonants in PIE; the reconstruction of two or three velar series; and the reconstruction of voiced aspirates. However, closer examination shows that the Nostratic contribution to these debates consists only of ambiguous and doubtful data which do not add to our existing knowledge.

It may be instructive to conclude this section with a consideration of one of the most widespread and semantically plausible Nostratic roots, and its pay-off for IE studies. The Nostratic reconstruction is ***kälU* meaning 'female relation of the opposite moiety', and this is held to be the direct ancestor of PIE **gl̥ō*- 'husband's sister' (Latin *glōs*, Greek *gál(o)ōs*, Late Church Slavonic *zŭlŭva*, Armenian *tal*, all meaning 'husband's sister', and the recently added (and unknown to the proponents of Nostratic) Sanskrit cognate *giri*- 'sister-in-law'). For the Indo-Europeanist, the formal reconstruction of this word is problematic for two reasons: the original inflectional pattern is difficult to reconstruct, and the word appears to contain two vowels side by side with no intervening consonant – vocalic **l̥* and **ō*. The Nostratic reconstruction helps with neither of these problems: the vocalism is deemed to be an entirely IE development (vowels are often particularly problematic in Nostratic etymologies), and there is as yet no comparative morphology for this word. Furthermore, the Sanskrit cognate shows that the word must be reconstructed with initial **g*- (as opposed to **g´*-, see section 2.4 for the significance), which cannot derive directly from ***kälU* according to current theory. The Nostratic comparison is no more helpful when it comes to semantics: the exact meaning of the PIE form (which is opposed to terms for 'husband's brother', 'husband's mother', 'husband's father' and 'husband's brother's wife', see section 7.4) is replaced by a catch-all term 'female-in-law'. Considering the present state of research in Nostratic, the Indo-Europeanist can afford to limit the time and space devoted to its study.

Further reading

General

There are numerous surveys of the IE language family, the reconstruction of PIE and of the individual languages in the family. Among recent studies Fortson (2004) gives a reader-friendly overview of much of the current work on reconstruction and also includes chapters on the individual branches of

the family, with sketches of their diachronic development from PIE and indications of reliable editions of texts, grammars and lexica. Meier-Brügger (2003) is also excellent for bibliographical surveys of the field and indications for further reading on current debates. Further useful surveys of the family and the major IE languages and language groups may be found in Bader (1997), Ramat and Ramat (1997) and Woodard (2004). Lusitanian and Tartessian are two of the languages of Spain for which information has been made more widely available recently thanks to the work of Untermann (see Untermann 1997 for a survey and texts in Lusitanian and Tartessian).

Work on reconstructing the IE family tree and cladistics has recently received an upsurge of interest, with scholars from genetics and computer science joining with linguists to work out the best phylogeny for the IE family. Two volumes of papers, McMahon (2005) and Forster and Renfrew (2006), offer the most recent reconstructed trees and discussion of the most appropriate methods of computer-based quantative comparison. Many earlier scholars had proposed that Anatolian and Tocharian had broken off earlier from the PIE family than the other language branches on the basis of phonological, morphological and lexical features (see, for example, Klingenschmitt 1994 (for Anatolian) and Ringe 1988–90 (for Tocharian)).

There has been much debate about the location of the speakers of PIE in time and space since the publication of Renfrew (1987). Much of the discussion is centred around archaeological, rather than linguistic, data; Mallory (1989) and Mallory and Adams (2006) present alternative theories to Renfrew. The edited volumes by Blench and Spriggs (1997–9) and McMahon *et al.* (2000) include articles by historical linguists and Indo-Europeanists addressing the methodological problems of connecting archaeological and linguistic data, and assigning a time-depth to a reconstructed proto-language.

A very good introductory account of long-range comparison in general is given by Trask (1996, Chapter 13). Much of the Russian work on Nostratic is untranslated, including the still incomplete, posthumously published work of Illič-Svityč (1971–84), although some of the key Russian articles are translated in Shevoroshkin and Markey (1986). In English, Manaster Ramer (1993) gives a sympathetic overview of Illič-Svityč's work and the history of Nostratic studies after Illič-Svityč, and presentations of reconstructed Nostratic are given by Kaiser and Shevoroshkin (1988) and Dolgopolsky (1998 and 1999). A very brief sketch, together with putative connections between Nostratic and other language groups and the reasons why the Nostratic hypothesis is 'plausible and fruitful', is found in Shevoroshkin and Manaster Ramer (1991). Bomhard and Kerns (1994) reconstruct their own brand of Nostratic, with which Greenberg's 'Eurasiatic' (Greenberg 2000) is in close accord (Eurasiatic is Nostratic without Afro-Asiatic or Dravidian, or Kartvelian, but with Eskimo-Aleut, Gilyak, Chukotian and Japanese-Korean-Ainu). The argument against the misuse of the comparative method is well put by Nichols (1996). Ringe's (1992) work dealing with the mathematics of comparing Nostratic correspondences with pure chance is superseded by Ringe (1995 and 1999). Critical comments on the Nostratic

theory are also found in many of the papers in Joseph and Salmons (1998), and Ringe (2002) is a devastating review of Greenberg's work on Eurasiatic. Watkins' verdict on the uninteresting nature of Nostratic, followed by comments from various linguists, is found in Rowenchuk (1992). For Sanskrit *giri-*, and literature on the word for 'husband's wife' in IE, see Mayrhofer (1986–2001: I 487f.).

Discussion points

1. Why study Indo-European rather than another large language family?
2. Why are morphological correspondences taken to be crucial as an indication of language relatedness?
3. In a famous article (Trubetzkoy 1939), the linguist Trubetzkoy stated that it was not possible to isolate a series of lexical and morphological elements which were present in all IE languages and absent from all non-IE languages. He proposed instead that languages could be classed as IE on the basis of the presence of the following six structural features: i) absence of vowel harmony; ii) absence of any restriction on which consonants could stand at the beginning of a word; iii) possibility to derive new words through prefixation (e.g. English *unkind*, derived from *kind*); iv) use of vocalic alternations within the lexical stem in morphology (e.g. English *ride*, *rode*, *ridden*); v) use of alternations of consonants within morphology (e.g. English plural morpheme has form /s/ in *cats* but /z/ in *dogs*); vi) the subject of a transitive verb having the same form as the subject of an intransitive verb. How well do Trubetzkoy's structural criteria apply to any IE languages you know? Are structural features a better way of grouping the IE languages together than lexical and morphological elements?
4. The Etruscan language is increasingly better understood. Some features of Etruscan grammar and morphology are given below:

Pronouns:	*mi* 'I' *mini* 'me' (accusative)
	ita 'this' *itan* / *itun* / *itn* this (accusative)
Nouns:	Case and number marking
	nominative / accusative – no affix
	genitive *-s* or *-l*
	locative *-i*
	plural (human) *-r*
	plural (non-human) *-χva*

Affixation is basically agglutinative, so, for example, *clan* 'son' has a plural *clen-ar*, and the genitive plural is *clinii-ar-as* (the vowel changes in the stem are the result of particular sound changes which affect this stem and are not general).

Verbs:	Tense and mood marking
	-ke past active
	-χe past passive

Verbs seem not to change for singular and plural, or for different persons.

Numbers

θu 'one'	*θunz* 'once'	*θunur* 'single'	*θusna* 'first'
zal 'two'	*eslz* 'twice'	*zalar* 'double'	
ci 'three'	*ciz* 'three times'	*ciar* 'treble'	
zaθrum 'twenty'			*zaθrumsna* 'twentieth'
cialχ 'thirty'			

Vocabulary

clan 'son', se*χ* 'daughter', *apa* 'father', *api* 'mother', *ruva* 'brother', *puia* 'wife', *nefts* 'nephew', *am-* 'be', *tur-* 'give', *ar-* 'make, put', *lup-* 'die', *θi* 'water', *vinum* 'wine', *-c* 'and', *sval* 'alive'.

Sample texts

ein θui ara enan
not here put anything-ACC
'don't put anything here'

itun turuce venel atelinas tinas cliniiaras

this gave Venel Atelina Zeus-GEN sons-PL.GEN
'Venel Atelina (a name) gave this to the sons of Zeus'

Arguing from the basis of any IE languages which you know, what arguments can be constructed a) for the inclusion of Etruscan in the language family; and b) against the inclusion of Etruscan in the IE language family? What do we need to know to make the argument conclusive in either direction? (You may wish to return to this question after reading later chapters.)