

The Indo-European Beginnings

The Indo-European Roots of Greek

Greek belongs to a family of related languages which are called “Indo-European” because at the time of the discovery of this family the known languages were distributed in Europe and the Indian subcontinent (Indo-European languages were subsequently discovered in Asia Minor and central Asia). The existence of such a family was suggested by William Jones, a British scholar and lawyer who was appointed to the Supreme Court at Calcutta in 1783. Jones was an expert linguist who had taught himself Arabic and Persian at Oxford in addition to Greek and Latin; he was also a radical politician, who supported the American revolution and bitterly attacked the slave trade. When he arrived in India as a judge he learned Sanskrit, the ancient classical language of India and the sacred language of Hinduism, in order to understand the principles of the native Hindu legal tradition (he wrote several books on Hindu and Moslem law in India). In 1786 he delivered a paper in Calcutta to the Asiatic Society of Calcutta, which included the following famous words:

The Sanscrit language, whatever be its antiquity, is of a wonderful structure; more perfect than the Greek, more copious than the Latin, and

more exquisitely refined than either, yet bearing to both of them a stronger affinity, both in the roots of verbs and the forms of grammar, than could possibly have been produced by accident; so strong indeed, that no philologist could examine them all three, without believing them to have sprung from some common source, which, perhaps, no longer exists.

Throughout the nineteenth century work continued on the newly discovered family of languages, mostly in Germany, and this gave rise to the new science of linguistics in the West. In India there was a long and illustrious tradition of linguistics, going back to the late sixth century BC, when the famous grammarian Pāṇini composed his exhaustive grammar of the Sanskrit language (and the tradition of systematic thought about language in India was doubtless older than Pāṇini). There was no analogous “classic” in Greek or in Roman literature. Although in both the Greek and the Roman world there was interest in language, this was mostly related to its importance to philosophy and rhetoric in the early period; there was more technical work on language in the Hellenistic and Roman periods, but this was focused more on textual criticism and the explication of archaic and classical forms of the language for educational purposes. Europeans were still rather unsophisticated linguists in the eighteenth century. However, the kick-start given by the comparison of Greek, Latin, and Sanskrit, followed by study of Germanic and then Slavic, Celtic, and others, led to the development of what we now call historical linguistics: the study of the development of languages over time, and the reconstruction of an unattested “parent language” by systematically comparing the later languages which have survived in written form. This was the start of modern Western linguistics: at the end of the nineteenth century Ferdinand de Saussure, who had been trained in historical Indo-European linguistics, moved from considering the development of languages over time (historical linguistics) to the analysis of structural relations of languages at a given point in time (synchronic linguistics).

The Family Tree

Indo-European historical linguistics was, of course, a child of its time, and many of the linguistic models and metaphors which have become ingrained in our way of thinking about language reflect the

intellectual environment of the nineteenth century. Part of this environment was a fascination with biological taxonomy and the evolution of species: Charles Darwin's *On the Origin of Species* (1859) provoked an intellectual revolution, and it is no coincidence that much of the terminology of historical linguistics is reminiscent of biology. Languages are described as related, and form a family; one aims to reconstruct a parent language, from which the daughter languages evolve; relations between languages are set out in branching tree diagrams, like a family tree. This type of relationship between languages is called genetic. Both the model and terminology have the potential to be extremely misleading, since languages are not in fact organisms: an essential difference from the Darwinian model is that languages (or rather, their speakers) *do* pass on acquired changes. In addition to this, language is a sociocultural force which plays a central role in the self-definition of the speaker: these two facts have consequences for the way we think about language change and the model of the family tree.

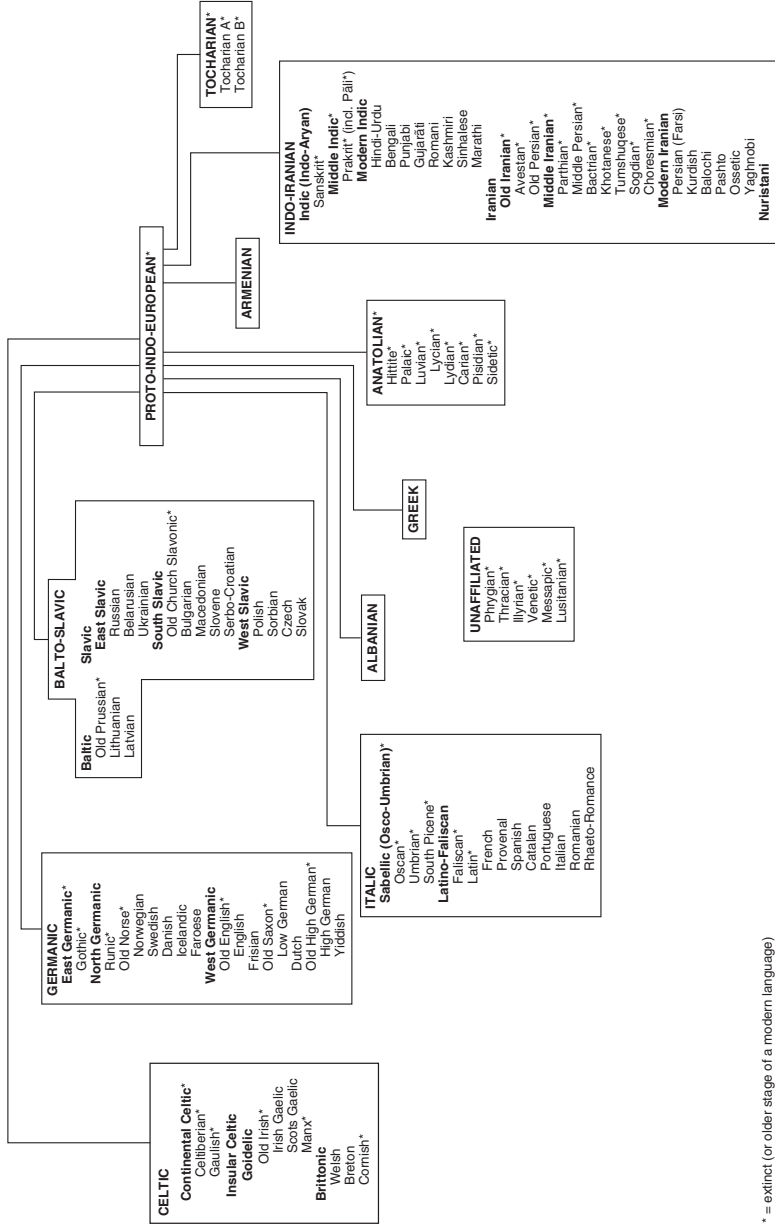
It is true that most speakers learn a variety of the native language from parents (or older speakers in general); in this sense a language may be said to be "inherited." But the metaphor does not bear pressing: for in fact a speaker learns not just one native idiom, but a variety of idioms from a variety of different speakers. In addition to grandparents, parents, and siblings, most children are exposed to different varieties of the language from the community at large. A competent native speaker is capable of recognizing a wide range of varieties (and their social connotations), and has mastery of quite a few varieties which are employed in different social situations. This reflects that fact that the notion of a language is to some extent a social construct: a language typically consists of a variety of different idioms and dialects, and in many cases is not clearly distinguishable from neighboring languages. And even when neighboring languages are in fact distinct, they may still form part of the speaker's linguistic competence (monolingual cultures are exotic in the world, not the norm). Of course, in many cultures there is a prestigious standard language which many speakers think of as *the* language (and other varieties may be seen as inferior by comparison to this standard), but this perception is a cultural and political phenomenon, rather than a reflection of linguistic reality.

There are clear consequences for the genetic metaphor of language relationship and language change when we replace the idea of

a uniform language inherited from parents with that of a continuum of language varieties taken over from across the language community. First, it can be seen that the native speaker's competence has multiple sources, and is subject to continuing development, so one cannot contrast the validity or purity of a genetic relationship with "contamination" or "influence" from other sources. The second point is closely related to this: a language change occurs when a majority of speakers adopt for use in a majority of situations a variant which was previously used by a minority of speakers, or in a restricted social context, or both. The reasons that prompt speakers to adopt such changes are complex: sociolinguistic research indicates that these decisions – like decisions pertaining to clothing and personal appearance – are the result of the speaker's desire to shift his or her identity with regard to a particular section of the community. This type of behavior is easy to observe in adolescents, but research indicates that it persists in a subtler form in people of all ages. Speakers may be unconscious of many of the linguistic shifts they are making.

Since the growth of sociolinguistics enabled linguists to understand how languages change, it has become common to emphasize the importance of "areal" factors in describing linguistic change and language relationships, at the expense of the traditional "genetic" family tree. This shift in emphasis offers important insights into the historical development of Greek, even though we have seen that the distinction itself is slightly dubious. "Genetic" can be applied, metaphorically, to features of a language which were observable in an earlier stage of that language, while "areal" covers features which have entered the language from elsewhere.

The language groups which are now derived from the Indo-European parent language are: Albanian, Baltic, Anatolian, Armenian, Celtic, Germanic, Greek, Indic, Iranian, Italic, Slavic, Tocharian. Very poorly attested languages or groups include Illyrian, Phrygian, Thracian; it seems certain that many other languages have disappeared without trace. The relationships between these groups are not identical: for example, Indic and Iranian are so close that they are generally grouped together as "Indo-Iranian," and Celtic, Germanic, and Italic show overlapping similarities which are best explained by their contiguity in the northwestern area of the Indo-European world. It is generally agreed that the Anatolian group must have split off from the parent language earlier than the others, since



* = extinct (or older stage of a modern language)

Figure 1.1 Family tree of the Indo-European languages. Source: Benjamin W. Fortson IV, *Indo-European Language and Culture: An Introduction*, 2nd edn. (Oxford: Wiley-Blackwell, 2010), Figure 1.1.

it has peculiarities which sets it apart from the rest of the “family” (such as lack of a separate feminine gender). There are a number of different types of reason for thinking that these languages are related.

Systematic correspondences in the phonology

This means that for Indo-European an inventory of phonemes (sounds) is reconstructed by comparing the daughter languages: we use words which appear to have a similar form and meaning across the I-E languages to build an inventory of phonemes for the parent language, and to postulate a number of sound-change rules for the daughter languages.

Example: the word for “foot” (accusative case):

Greek	Latin	Sanskrit	Gothic	Hittite
πόδα [poda]	<i>pedem</i>	<i>pādam</i>	<i>footum</i>	<i>pada</i>

In this example Greek [p] corresponds to a [p] in Latin, Sanskrit, and Hittite, and to an [f] in Germanic (Gothic). On this basis a phoneme [p] is reconstructed for Indo-European (written **p*), and a sound change **p > f* is posited for Germanic. This is known as the comparative method, and is fundamental to historical linguistics. The comparative method does not like sound changes to have exceptions; if we state that an I-E **d^b* (aspirated *d*) becomes Greek *t^b* (aspirated *t*, written *θ*) in one word, then *the same change has to operate in all words*.

Abandoning this principle of regularity means that any random, haphazard, or frankly lunatic etymology can be constructed for any language, and this was regularly done from antiquity until the eighteenth century. Compare, for example, the etymologies of the Roman scholar Varro (5.20):

Apri ab eo quod in locis asperis, nisi a Graecis quod hi *kaproi*. *Caprea* a similitudine quadam caprae. *Cervi*, quod magna cornua gerunt, *gervi*, *G* in *C* mutavit ut in multis. ... *Volpes*, ut Aelius dicebat, quod volat pedibus.

The word for wild boar [*aper*] comes from the fact that they have a rough [*asper*] habitat; unless it is from Greek, because the Greek word

is *kapros*. The roe deer [*caprea*] is named from a certain resemblance to the she-goat [*capra*]. Stags [*cervus*] are so called because they bear [*gerunt*] large horns, the *G* of *gervus* has changed into a *C*, as often happens. ... The fox [*volpēs*] is so called because, as Aelius said, it flies [*volat*] with its feet [*pēs*].

The comparative method does not allow such random deletion and substitution of sounds: when sounds change, they do so in accordance with clearly defined rules. The Latin word *cervus* “stag” is in fact derived from the I-E **ker-* “horn”, which gives the Greek κέρας [*keras*] “horn” (as well as the English word *horn*).

However, a force that can undo regular sound change is analogy, which plays an important role in all aspects of human language. In English, for example, the verb *to dive* had an original “weak” past tense *dived*, but in some dialects this has changed to *dove* on the analogy of “strong” verbs such as *drive* → *drove*. In Greek an *s* inherited from Indo-European first became an *h* and finally disappeared between vowels, as in the nom. plur. of γένος [*genos*] “family”: **genes-a* > **geneha* > *genea*. But in some cases the force of analogy led to the retention of an intervocalic *s*. For example, the future tense in Greek was created by adding an *-s-* to the verbal stem:

<i>klep-</i> “steal” →	future stem <i>kleps-</i>	<i>klepsō</i> “I shall steal”
<i>lu-</i> “release” →	future stem <i>lus-</i>	<i>lusō</i> “I shall release”

Normally we would expect the intervocalic *-s-* in *lusō* to disappear; but in this case the *-s-* was maintained or restored on the analogy of consonant-stem verbs like *klepsō*. It would have been inconvenient for the future marker to disappear: this would have given *luō*, identical in form to the present tense.

Fundamental similarities in the morphology

Indo-European clearly had a complex inflecting morphology, since all the daughter languages have preserved elements of this. An inflected language is one in which grammatical significance is carried

by changes in the form of the word, usually in the ending and often in the stem as well. A typical Indo-European word is built as follows:

root + suffix + grammatical ending

The root of a word carries the basic meaning; adding a suffix to it creates a stem to which the endings can be added. Thus, to take the verb “to release” that we considered above: *lusō* “I shall release” breaks down as:

<i>lu-</i> (root) +	→	<i>lus-</i> (future stem) +	→	<i>lusō</i> “I shall
suffix <i>-s-</i>		1st person ending <i>-ō</i>		release”

We could also add the “agent” suffix *-tēr* (related to Latin *-tor* as in *pastor*, Engl. *-er* as in *maker*) to the root *lu-* to make an agent noun: *lutēr* “one who releases, deliverer.”

In languages such as modern English much of the meaning is carried instead by a fixed word order rather than by endings, and by “auxiliary” words such as *will*, *had*, etc. The older Indo-European languages preserve the complex morphology that we can see in Greek and Latin: of course, they have all changed and innovated in various ways, but on the whole the basic morphological building blocks (the morphemes) are the same, or very similar. For example:

(i) I-E verb “to be” (root $*h_1s-$): $*h_1s-mi$ “I am,” $*h_1s-ti$ “s/he is”

Greek	Sanskrit	Latin	Hittite	Gothic
<i>emi, esti</i>	<i>asmi, asti</i>	<i>sum, est</i>	<i>esmi, eszi</i>	<i>im (< *immi), ist</i>

(ii) I-E noun “sheep”: nominative $*h_3ewis$ → accusative $*h_3ewim$

Greek	Sanskrit	Latin	Luwian (Anatolian)
<i>o(w)is, o(w)in</i>	<i>avih, avim</i>	<i>ovis, ovem</i>	<i>hawis, hawin</i>

In example (i) Latin *sum* is the result of a complicated process of sound change and analogical pressure; apart from that, differences between the forms are the result of regular sound changes. In example (ii) all the differences between the forms are the result of regular sound changes: in Greek the phoneme [w] is found in many dialects, but not in classical (Attic) Greek.

A morphological oddity that is evident in all I-E languages is the alternation of the vowel *e* with the vowel *o*. This is not a sound change but a morphological marker of Indo-European: thus the Greek verb *pherō* “I carry” has an *e* in the stem *pher-*, but the related noun *phoros* “tribute” has an *o* (stem *phor-*): the same process in the same root can be seen in English *bear* versus *burden*. A third possibility is that the vowel disappears completely: compare the I-E root **genh₁-* “procreate, family” in Greek *genos* “race, family”, *gonos* “offspring”, and *gnēsios* (adj.) “belonging to the family, legitimate” (the root appears here as *gn-*, as in Latin *gnātus* “son”). This process, known as ablaut, is fossilized (no longer productive) in the daughter languages, and has suffered analogical interference, with the result that the *e/o* variation appears almost random (hence Latin *ped-* “foot” but Greek *pod-*).

A large number of lexical roots in common

It is clear that related languages are likely to have a large amount of vocabulary in common (although sound changes may have changed the form of the words to some extent): we have already come across some examples above. However, languages very often change the meanings of words, and drop words for no apparent reason (dropped words may be replaced by borrowings, or by other words which have been pressed into service, or which have had their meanings extended). Words which are more likely to resist replacement include the so-called “core” vocabulary: numerals, body parts, family members, and certain others. It is rare, however, for a lexical root to survive in all the major attested branches.

I-E	Greek	Sanskrit	Latin	Anatolian	Germanic
<i>*māter-</i>	<i>māter-</i>	<i>mātar-</i>	<i>māter</i>	—	<i>mother</i>
<i>*ph₂ter-</i>	<i>pater-</i>	<i>pitar-</i>	<i>pater</i>	—	<i>father</i>
<i>*d^hugh₂ter-</i>	<i>t^hugater-</i>	<i>dubitar-</i>	—	<i>tuwatri</i> ⁱ	<i>daughter</i>
<i>*nas-</i>	—	<i>nas-</i>	<i>nāsus</i>	—	<i>nose</i>
<i>*wed-r/n-</i>	<i>budr-</i>	<i>udn-</i>	<i>unda</i>	<i>wadar</i> ⁱⁱ	<i>water</i>
<i>*dwo</i>	<i>duo</i>	<i>dvā</i>	<i>duo</i>	<i>dā</i> ⁱⁱ	<i>two</i>
<i>*g^wous</i>	<i>bous</i>	<i>gauḥ</i>	<i>bōs</i>	<i>uwa</i> ⁱ	<i>cow</i>

Notes: ⁱ Luwian (hieroglyphic)

ⁱⁱ Hittite

Common vocabulary by itself is not a reliable indicator that two languages are related, since languages borrow words from each other freely: common vocabulary needs to exhibit regular phonological correspondences, and similar morphological patterns (this is important, as morphology seems to be one of the areas of language which is most resistant to transfer across language boundaries).

Similarities in syntax and certain widespread poetic features

Since the I-E languages inherited very similar morphological systems (complex inflected morphology), their syntactic patterning is on the whole similar, at least in the early attested languages. Word order is free rather than bound: there is a tendency for the main verb to come at the end of a clause or sentence, and enclitic words generally follow the first accented word in the sentence. All I-E languages have relative clauses introduced by a relativizing pronoun (as in Engl. “The man *who* came to dinner”): I-E **yos* gave the Greek relative “who” (Gk. *hos*, Skt. *yah*), while other languages use the interrogative and indefinite stem **kʷi-*/**kʷo-* (> Lat. *quī*, Hitt. *kwis*). This stem survives in Greek *tis* (and in Latin *quis*) with interrogative and indefinite functions “who?”/“a certain.”

Language is also, of course, used for poetic and aesthetic purposes: in most of the major I-E languages there are traditions of epic poetry which show some interesting commonalities. Now, similarities between poetic or literary traditions do not prove a “genetic” relationship, since these things travel by processes of imitation and osmosis as well: there are also striking thematic similarities between Greek and non-Indo-European traditions of the ancient Near East (for example, the Mesopotamian *Epic of Gilgamesh*), which must be indicative of regional influence. Nevertheless, the I-E poetic traditions come from areas as far removed as Ireland and India, and often raise the possibility of tying thematic echoes to common linguistic forms. In 1853 the German scholar Adalbert Kuhn noticed that the Homeric phrase (*Iliad* 9.413) “undying fame”, κλέος ἄφθιτον [*kleos aphthiton*], was exactly cognate with the Sanskrit phrase *śrávas ... áksitam* (*Rig Veda* 1.9.7). This concept is an important part of the ideology of

the epic poetry (typically, it is the reward earned by the brave warrior), and the phrase may have roots in an ancient tradition of heroic praise poetry. Since then much work has been done on inherited features of language which go beyond pure phonology and morphology, and which give a sense of which “larger” aspects of the surviving languages may go back to an earlier period, from everyday turns of phrase to poetic themes and ideas. Similarities in poetic meter have also been studied in an effort to identify inherited metrical patterns.

Phonemic Inventory of Indo-European

The reconstructed phonemic inventory of Indo-European is set out below: some aspects of it are uncertain, but on the whole it represents a modern consensus:

Consonants

	<i>voiceless stop</i>	<i>voiced stop</i>	<i>voiced aspirate stop</i>	<i>fricative</i>
bilabial	<i>p</i>	<i>b</i>	<i>bʰ</i>	
apical	<i>t</i>	<i>d</i>	<i>dʰ</i>	<i>s</i>
velar ⁱ	<i>k</i>	<i>g</i>	<i>gʰ</i>	
labiovelar	<i>kʷ</i>	<i>gʷ</i>	<i>gʷʰ</i>	

Resonants and semivowelsⁱⁱ (consonantal ~ vocalic)

nasals	<i>m ~ m̥</i>	<i>n ~ n̥</i>
liquids	<i>l ~ l̥</i>	<i>r ~ r̥</i>
semivowels	<i>w ~ u</i>	<i>y ~ i</i>

Laryngealsⁱⁱⁱ

<i>h₁</i>	<i>h₂</i>	<i>h₃</i>
----------------------	----------------------	----------------------

Vowels and diphthongs

<i>e</i>	<i>o</i>	<i>ē</i>	<i>ō</i>	<i>ei</i>	<i>oi</i>	<i>eu</i>	<i>ou</i>
<i>a</i>		<i>ā</i>		<i>ai</i>		<i>au</i>	

Notes: ⁱ Velars. The reconstruction of I-E velars is complicated by apparent irregularities in their development in the daughter languages: for

example, **ǵ* gives *ǵ* in Greek and Latin, and usually gives *j* (as in Engl. *jam*) in Sanskrit:

ǵenu*/ǵonu* “knee” > Gk. *gonu*, Lat. *genu*; Skt. *jānu*

But in some cases a velar is continued as a velar in Sanskrit too:

**yugom* “yoke” > Gk. *zugon*, Lat. *iugum*; Skt. *yugam*

Since the comparative method (above) does not allow such irregularity, it is necessary to reconstruct two series of velars: a regular series **ǵ*, **k*, **ǵʰ* (for **yugom*), and an additional series, known as the palatal velars, **ǵʲ*, **kʲ*, **ǵʰʲ* (for **gonu*). However, since the reconstruction of two series is not necessary to explain data within Greek (or Latin) we shall ignore the distinction.

ⁱⁱ Resonants/semivowels. These phonemes can be either consonants or vowels, depending on where they appear in the word: [w] is simply [u] in consonantal function. Vocalic liquids and nasals may seem unfamiliar: but *ŋ* and *l̥*, for example, appear in the final syllable of *button* and *bottle* in normal spoken English.

ⁱⁱⁱ Laryngeals. The exact phonetic value of these sounds can only be guessed at. As consonants they disappeared from all branches of Indo-European except Anatolian, where at least one of them survived as an *h*. However, they left important tracks in the vowel system of Greek:

**h*₁ leaves an *e*, and does not affect a neighboring *e* (neutral or E-laryngeal).

**h*₂ leaves an *a*, and turns a neighboring *e* into an *a* (A-laryngeal).

**h*₃ leaves an *o*, and turns a neighboring *e* into an *o* (O-laryngeal).

Greek is the only major I-E language in which the three vowel colors are maintained (Phrygian seems to differentiate them too, but the language is very poorly attested): all other languages merge them into a single vowel (Latin has *a*, Sanskrit has *i*). In the parent language they may have been varieties of laryngeal [ʔ] (glottal stop) and pharyngeal [ʕ] (Arabic ‘*ayin*). This category of sounds is hard to define in normal phonetic terms: although generally classed as consonants, the way they affect the air-flow is peculiar compared to

regular consonants (which stop or impede it in the oral cavity), and they often behave like semivowels. In Greek their behavior can be summarized as follows:

**h* between consonants became a vowel: **ph₂ter-* > Gk. *pater-* “father”.

**h* after a vowel disappeared, but lengthened the vowel (and “colored” an *e*): **si-steh₂-mi* > Gk. *histāmi* “I set up, stand”, **di-deh₃-mi* > Gk. *didōmi* “I give”.

**h* before a vowel disappeared (and “colored” an *e*): **h₂enti* > Gk. *anti* “facing, in exchange for” (cf. Hittite *hants* “in front”), **h₃ewis* > Gk. *owis* “sheep” (cf. Luwian *hawis*).

Indo-European Language and People

If there was an Indo-European language there must, presumably, have been a group of people who spoke it. Since the late nineteenth century a huge amount of effort has been invested in trying to find out who these people were, where they lived, and how they lived. There are two principal sources of conjecture. Firstly, efforts have been made to identify Indo-European speakers with archaeological material. Secondly, reconstructed words have been used as evidence: this has been termed “linguistic palaeontology.” In its simplest form the method supposes that if a common word can be reconstructed for an object, the speakers of the reconstructed language must have known that object. For example, since we can reconstruct words for wheel, plough, yoke, horse, and various type of stock animal (pigs, sheep, cattle), it seems likely that Indo-European speakers were familiar with these objects and animals: by extension, it has been concluded that, before dispersal, I-E speakers practiced agriculture. Various other conclusions of a similar nature have, with varying degrees of caution, been arrived at. A problem is that the meaning of a reconstructed word is often not secure: while “mother,” “father,” “sheep” are clear, many terms for plants and animals – which could give a clue both to the location of the homeland and to the speakers’ way of life – have clearly changed meaning in the daughter languages: for example, the Greek word for oak, *φηγός* [*phēgos*], is the exact cognate of Engl. *beech* (and Lat. *fāgus* “beech”).

Much more dangerous is speculation about social structure, religion, or culture: a collection of asterisked and in varying degrees hypothetical words is simply insufficient evidence for anything but the most basic of observations. Furthermore, our reconstructed language is anachronistic in the sense that the comparative method is not good at sorting out different chronological or even dialectal layers in a language: we have a mish-mash of lexical items which we call a language, but which may have been in use at different periods and in different areas of the Indo-European area.

Language has often been thought of as an expression of the soul or psyche of a people: the Roman poet Ennius famously said that he had three souls because he spoke Latin, Oscan, and Greek. Perhaps for this reason there is always a strong temptation to connect a language not just with a group of speakers, but with a “race,” a notoriously undefined term. In the nineteenth century there was much speculation about an Aryan race which spoke the newly discovered parent language. The word Aryan was in itself a perfectly harmless term, being the word that the Indo-Iranian peoples used of themselves (Skt. *ārya-* and Old Persian *ariya-*, cognate with the word *Iran*): it was conjectured (wrongly) that it was the common Indo-European self-designation. The term was then adopted by European and North American racial theorists who believed in a hierarchy of races (their own at the top, by odd coincidence), and passed into the paraphernalia of Nazi Germany, along with the equally innocent swastika sign (Skt. *svastika* “good luck charm”).

The area which was inhabited by speakers of Indo-European is not known, though there have been many suggestions. There is as yet no consensus over the various efforts that have been made to identify Indo-European speakers with archaeological material. Scholarship since the 1950s has in general put the Indo-European homeland near the rough geographical center of the Indo-European speaking world, between the Black Sea and the Caspian Sea (the Pontic-Caspian Steppe): this region has been argued for by archaeologists who identify the “Kurgan” culture of the steppe with Indo-European speakers (*kurgan* is the Russian word for a burial mound, borrowed from Turkic). All arguments over the geography are bound up with arguments over the date of the parent language, and the method of its dispersal. The traditional view has been that the last period of common Indo-European dates to somewhere in the early or

mid-fourth millennium BC. A different view, first propounded in the 1980s, saw the geographical starting point in Anatolia: this view has not been widely accepted, partly because proponents of an Anatolian origin push the date of the parent language back as far as the eighth millennium. However, the debate led to a useful discussion of the ways in which languages spread, which has implications for the whole of the Indo-European area: the older view of migration and conquest by Indo-European speakers (using superior warfare techniques such as horses) is now seen as simplistic: an interesting feature of the Anatolian theory was that it connected the spread of the language with the spread of farming and associated technology, rather than with large movements of people.¹

Whatever the geographical origin of the Indo-European languages, a date in the fourth millennium still seems more attractive, partly because archaeologists are clear that the products associated with farming and wheeled vehicles (wheels, axles, yokes, wool, etc.) are not found earlier than the fourth millennium: since we can reconstruct Indo-European words for these items, if we were to push the dispersal of the language back to an earlier period we would have to assume that these words – which are found widely across the Indo-European languages – were innovated independently in each language group. In the case of the four farming terms mentioned above, for example, both English and Greek preserve the Indo-European words:

* <i>k^wek^wlos</i>	κύκλος [<i>kuklos</i>]	<i>wheel</i> (< OE <i>hweowol</i>)
* <i>aks-</i>	ἄξων [<i>aksōn</i>]	<i>axle</i>
* <i>yugom</i>	ζυγόν [<i>zugon</i>]	<i>yoke</i>
* <i>w₁lh₂n-</i>	λίθος [<i>lēnos</i>]	<i>wool</i> (< OE <i>wull</i> < Proto-Germanic * <i>wulno-</i>)

From Indo-European to Greek

Sometime between the last period of Indo-European (perhaps around the mid-fourth millennium BC) and our earliest surviving Mycenaean texts of around 1400 BC, speakers of one or more dialects of Indo-European arrived in the south Balkan peninsula. Since this region was later known (more or less) as Greece, the new arrivals

are sometimes known as “proto-Greeks” and their language as “proto-Greek.” These terms need to be used with some caution: the people who later called themselves Greek were a mixture of the newcomers and the people they found already living in the region, and their language was similarly the result of development of (probably) various closely related Indo-European dialects in interaction on Greek soil.

It is hard to date the arrival of these people, because there is no indisputable evidence from archaeology of a dramatic break in the culture of mainland Greece. This evidence might take the form of both widespread destruction of earlier settlements, and signs of the arrival of a new material culture: for example, new styles of pottery, new architectural forms, or a different style of burial. Furthermore, archaeologists are divided over whether the arrival of a new group of people (let alone a new language) is always reflected by changes in the archaeological remains, and vice versa (this has been dubbed the “pots=people” debate). Such evidence as there is has often been taken to point to a date at the end of the early Bronze Age, around 2100–1900 BC (the period known to archaeologists as Early Helladic III). This is perfectly plausible from a linguistic perspective.

Note

- 1 The Kurgan hypothesis was proposed by Marija Gimbutas (1931–1994), the Anatolian farming hypothesis by Colin Renfrew (1937–). Conveniently summarized with bibliography in Mallory (1989: chapter six). Linguistic arguments for the later date in Garrett (2006).