3 The Indo-European Linguistic Family: Genetic and Typological Perspectives

Bernard Comrie

Introduction: Genetic and Areal Affiliations
The other chapters in this book are essentially inward-looking in terms of their Indo-European perspective, examining reasons for positing the genetic unity of the Indo-European languages and ways of accounting for their differentiation from a single ancestor language. This chapter, by contrast, is outward-looking, seeking to identify some salient features of Indo-European against a background of the range of structural diversity found across the languages of the world — in other words, to typologize Indo-European. The emphasis in this typological chapter will be on Proto-Indo-European and the early attested IE languages.

In order to put the typology of early Indo-European into historical perspective, it is important to consider with what other languages Proto-Indo-European might have genetic affiliations and with what other languages Proto-Indo-European might have been in areal contact; in some cases, the answers to the two questions might be the same, since if Proto-Indo-European does have genetic relatives, then at some point it must also have been in areal contact with them. Where early Indo-European shares typological characteristics with other languages and language families, distant genetic affiliations or areal contact are a plausible reason for these similarities.

The question of possible more distant genetic affiliations of the IE language family is one fraught with controversy, often even with emotional polemic, and no attempt will be made here to impose a particular viewpoint. One thing that is clear is that Indo-European itself is a well-defined genetic unit, that is, any two arbitrarily selected IE languages are genetically closer
to one another than they are to any non-IE language. Thus what is at issue is not so much the nature of Indo-European but whether or not Indo-European forms part of a larger genetic unit (in much the same way as the Germanic languages form a genetic unit within the larger IE family). The various suggestions that have been made for more distant genetic affiliations are far from as tightly argued as are those for the internal genetic unity of Indo-European itself, but none the less there are some tantalizing hints of broader genetic links that deserve further consideration. For instance, in Proto-Indo-European the first and second person pronouns are characterized, at least in some morphological forms, by stems containing *m- and *t- respectively. Essentially this same pattern, with m- (or sometimes another labial) for first person and t- for second person, is found across a number of other language families of northern Eurasia: Uralic, Yukaghir (which many scholars consider to form, with Uralic, a Uralic-Yukaghir genetic unit), Turkic, Mongolian, Tungusic (which three, many scholars consider to form a single, Altaic family, perhaps also including Japanese and Korean), Chukotko-Kamchatkan (in north-eastern Siberia), and Eskimo-Aleut (in northern North America). Since the probability of this same pattern being replicated independently by chance in so many instances is low, and since personal pronouns are relatively immune to borrowing, this suggests that there may be a distant genetic relation among these languages, which Greenberg (forthcoming) has subsumed, along with some others, in a putative Eurasian family. Other scholars have placed equal or more emphasis on possible genetic relations to languages to the south and east, in particular Afroasiatic (the family which includes Semitic as one of its branches, with such languages as Arabic and Hebrew), Kartvelian or South Caucasian (spoken to the south of the main range of the Caucasus and including Georgian as its major representative), and Dravidian (whose member languages are spoken primarily in southern India); they prefer Nostratic as the name of the larger genetic unit to which they would assign Indo-European (e.g. Illič-Svityč 1971–84).

In dealing with areal contacts between Indo-European and other language families, and the possibility of similarities that may have arisen through areal influence, it is important to restrict our attention to languages with which Proto-Indo-European was in contact, since later contacts between individual IE languages and non-IE languages belong properly to the history of those individual languages. The precise range of areal contacts entered into by Proto-Indo-European will of course depend on the solution to such controversial questions as the location of the PIE homeland and the geographic location of the other languages at the time in question. Suffice it to say that plausible candidates for substantial areal contact with Proto-Indo-European include the proto-language or some individual languages of the following families: Uralic, Afroasiatic (more specifically: Semitic), Kartvelian, possibly also North Caucasian (i.e. the indigenous languages of the northern Caucasus,
including North-west Caucasian and Nakh-Daghestanian or North-east Caucasian).

**Phonological Typology**

From a typological perspective, the vowel system of late Proto-Indo-European, that is, the stage immediately preceding the break-up of the parent language, would hardly elicit surprise, containing as it does the most common five-vowel system in the world (*i-e-a-o-u*), superimposed on which is a two-way opposition of vowel length, plus the ‘schwa’-vowel standing outside the length opposition. By contrast, the consonant system, more specifically the obstruent system, is highly skewed, with a preponderance of stops distinguished by a rich set of places of articulation (perhaps as many as five: labial, dental, palatal, velar, labiovelar) and of phonation types (perhaps as many as four: voiceless unaspirated, voiceless aspirated, voiced unaspirated, ‘voiced aspirated’), but only a single fricative phoneme (*s*). While obstruent systems with so many stops and so few fricatives are not unknown, being found for instance quite typically in the indigenous languages of Australia, they are certainly unusual across languages of the world as a whole. And indeed, a substantial part of the further reconstruction back to early Proto-Indo-European has been concerned with providing a typologically more plausible account of the consonant system – but with the result that in turn a new vowel system has been proposed that stands at the boundaries of typological plausibility.

Perhaps the most important step in this shift of typological perspective came with the Laryngeal Theory, which gradually stripped away the oppositions of quantity and quality in the vowel system: *i* and *u* can be treated as allophones of *y* and *w* respectively; *o* is to a large extent a morphophonemic alternant of *e* under ablaut; a neater overall account can be given of ablaut if it is assumed that the long vowels reflect an earlier sequence of short vowel plus laryngeal, and schwa can then in turn be considered a syllabic laryngeal; many remaining instances of the vowels *a* and *o* can plausibly be considered reflexes of laryngeal–vowel or vowel–laryngeal sequences if one assumes that there were three laryngeals, the late PIE differentiation of vowel qualities (*e-a-o*) reflecting an early differentiation of laryngeals (*h₁–h₂–h₃*), perhaps mirroring the palatal–velar–labiovelar opposition (see also Chapter 2, pp. 40–4). Thus early Proto-Indo-European, or at least this widely accepted version of early Proto-Indo-European, ends up with a more balanced obstruent system through the introduction of the three ‘laryngeals’ (at least some of which are probably fricatives), but by the same token ends up with a typologically highly unusual vowel system: perhaps just a single vowel, symbolized *e*. While it is probably true that no absolutely convincing case has been made for any attested and well-described language of the world that it is monovocalic, there are some languages that certainly come close,
including, interestingly, the North-west Caucasian languages (Abkhaz-Abaza, Circassian, Ubykh), which have very restricted vowel inventories and very rich consonant inventories (Hewitt 1981: 205–7). In terms of areal typology, and perhaps more distant genetic comparison, it is also worth noting that the Semitic languages, especially phonologically more conservative Semitic languages like Arabic, have relatively small vowel inventories (Classical Arabic has i–a–u, long and short) but rather rich consonant inventories. Both North-west Caucasian and Semitic languages are particularly rich in ‘laryngeal’ consonants, that is, consonants produced at the back of the upper vocal tract.

The late PIE obstruent system provides another apparent typological anomaly in the range of phonation types. Since the evidence for the voiceless aspirated series is rather marginal, perhaps reflecting an innovation of Indo-Iranian rather than a feature of the proto-language, the phonation types can plausibly be reduced to three: voiceless unaspirated, voiced unaspirated, and voiced aspirated. But it was then argued that, typologically, the presence of a voiced aspirated series necessarily presupposes the existence of a voiceless aspirated series. Either one goes back to the fourway opposition of phonation types, or one reanalyses the phonetic nature of one or other of three other phonation types. Actually, the term ‘voiced aspirate’ is something of a misnomer, since in strict phonetic terms voicing and aspiration are incompatible; rather, what is meant is ‘breathy (or murmured) voice’, and the question therefore boils down to whether a breathy voiced series is possible in the absence of voiceless aspirates. If one assumes that it is not, then phonetic reanalysis is required, as in the Glottalic Theory (Gamkrelidze and Ivanov 1984; ch. 1), which holds that the so-called voiced unaspirated stops were in fact glottalized stops, with the voiceless unaspirated stops being phonologically classifiable rather as voiced unglottalized, while the so-called voiced aspirated stops can be classified phonologically simply as voiced stops (see Chapter 2, p. 38). The typological debate has, however, been reopened by the discovery of languages in West Africa that have essentially the threeway opposition of phonation types proposed for Proto-Indo-European (Stewart 1989: 231–9). It should be noted that the typological argument just outlined is not the only evidence in favor of the Glottalic Theory, and if the Glottalic Theory is accepted then Proto-Indo-European would share a striking typological feature with North Caucasian, Kartvelian and possibly Semitic (the so-called ‘emphatics’ show up in some Semitic languages, e.g. Amharic, as glottalized consonants; their value in Proto-Semitic is unclear). (For a more detailed consideration of these phonological problems, reference may be made to Comrie 1993.)

We have already alluded to the most striking morphophonemic alternation of Proto-Indo-European, namely ablaut, whereby the basic vowel e appears, under appropriate circumstances, as o or is even deleted altogether. Somewhat similar patterns of changes in vowel quality and loss of vowels have been
investigated from the viewpoint of areal typological similarities to Indo-European in the Kartvelian languages (Gamqrelize and Maçavariani 1965).

**Morphological Typology**

In speaking of the morphological typology of Proto-Indo-European, there are three main topics that deserve attention: (a) the general morphological type, in terms of the classification into isolating, agglutinating and fusional languages; (b) the distinction of words into word classes or parts of speech; (c) the most salient categories that are expressed morphologically. Kuryłowicz (1964) is a masterly discussion of this area of Indo-European linguistics.

**Morphological Type**

In terms of the classification of morphological types, early Indo-European is fusional. It is distinct from an isolating language, in which each word consists of one and only one morpheme, in that it has many words consisting of more than one morpheme, for example, Lat. *viri* 'man (gen. sg.)', where the one word includes expression of the lexical item ('man'), of singular number, and of genitive case, that is, three morphemes. It is distinct from an agglutinating language in that it is frequently impossible to segment a word into morphs (formatives, sequences of phonemes) corresponding to its constituent morphemes. In Lat. *viri*, for example, there is no segmentable morph expressing singular number (cf. acc. sg. *virum*), nor is there a segmentable morph expressing genitive case (cf. gen. pl. *virōrum*). Indeed, the fusional nature of early Indo-European morphology is exacerbated by the fact that the major parts of speech fall into inflectional classes, such that the same combination of morphemes often finds different expression in different inflectional classes: thus, genitive singular can also be realized in Latin as *-ae* (e.g. *mensae* 'table (gen. sg.)'), *-is* (e.g. *regis* 'king (gen. sg.)'), and so on. In the classification of morphological types, a fourth type is sometimes introduced, namely polysynthetic, referring to a language where typically a large number of morphemes are combined into a single word. Although early Indo-European is fusional, it is not polysynthetic. In nouns, for instance, only two categories are expressed inflectionally (case, number), while in verbs the range of categories expressed is of the order of person–number (of the subject), tense–aspect, mood, voice, finiteness, in sharp contrast to typical polysynthetic languages which readily include a dozen morphemes in a single word, such a word often functioning as a complete sentence. Indeed, it is remarkable that despite the complexity of the morphology of the early IE languages, easily testified to by anyone who has had to master the morphology of an early IE language, this complexity is due not to any large number of morphological categories expressed, but rather to the fusional expression of the few categories that are expressed.

While fusional morphology characterizes the earliest attested IE languages
and is indeed still characteristic of many modern languages (e.g. Icelandic and the Baltic and Slavic languages), and forms part of the reconstruction of late Proto-Indo-European, the application of internal reconstruction to the product of comparative reconstruction has been able to strip away some, but by no means all, of the fusional nature of early Indo-European and point to an earlier stage of Proto-Indo-European which was more agglutinative; for rather strikingly different reconstructions of this type, reference may be made to Szemerényi (1989) and Beekes (1990). The fusional nature of Late Proto-Indo-European stems in part from intervening phonological changes that obscured similarities and to the influence of analogy, which further obscured the original pattern. A simple example is provided by the accusative singular in Ancient Greek, whose two major exponents, -n and -a, though quite different synchronically, both derive from PIE *-m, non-syllabic and syllabic respectively. Historically, Ancient Gk -n is the reflex of PIE non-syllabic *m and is thus expected after a syllabic segment, as in lógon 'word (acc. sg.)', while -a is the reflex of syllabic *m and is thus expected after a non-syllabic segment, as in phylaka 'guard (acc. sg.)'. But dental stems with unaccented final syllable take the -n allomorph, with deletion of the stem-final dental, as in érin, accusative singular of éris (stem erid-) 'strife' (cf. elpída, accusative singular of elpís 'hope', with an accented final syllable). Even this distribution is disturbed by exceptions like kleís (stem kleid-) 'key', whose accusative singular is usually kleín, rarely kleída. Such subsequent phonological and analogical changes even lead to difficulty in separating stem from inflection, so that in declining a Latin noun like hortus 'garden' synchronically, it is usual to segment along the lines nom. sg. hort-us, voc. sg. hort-e, gen. sg. hort-i, dat.–abl. sg. hort-o, thus losing sight of the etymological stem horto-(of which horte is an ablaut variant).

Of the languages with which early Indo-European was in areal contact, fusional morphology is particularly characteristic of the Semitic and Kartvelian languages.

**Word Classes (Parts of Speech)**

The most readily distinguishable word class in Proto-Indo-European is the verb, which has a number of categories not shown by any other word class, e.g. person-number, tense–aspect, and mood. Within the remainder, the core class is composed of substantives (nouns, in the narrow sense of this term), which are characterized by the categories of case and number. These same categories characterize a number of other word classes that are distinguishable, to different degrees, from substantives. Thus, pronouns show the same categories as substantives, but follow a different inflection in some respects, for example, with nominative singular neuter in *-d rather than *-m, e.g. Lat. illud 'that (nom. sg. n.)' versus bonum 'good (nom. sg.)'; the personal pronouns, originally restricted to the first and second persons (demonstratives being used in place of third person pronouns) have an even more idiosyncratic
declension, with some suppletion between the nominative and oblique cases, a phenomenon to which we return in *Morphological Categories*, pp. 81ff. The distinguishability of a separate class of adjectives is even more questionable for Proto-Indo-European, and in the early IE languages adjectives differ little from substantives, the most important difference being that adjectives show a fully productive inflectional category of gender (e.g. Lat. m. bonus, f. bona, n. bonum), whereas with nouns gender is at best a derivational category, for example, victor ‘conqueror (male)’, victrix ‘conqueror (female)’. In individual early IE languages, however, further distinctions arise between substantives and adjectives. In Ancient Greek, for instance, they may have different accentual patterns, as when the genitive plural of first-declension nouns is always accentuated -ōn (e.g. khōrōn, genitive plural of khōrā ‘land’) while that of first-declension adjectives is not so accentuated if the nominative singular does not have final accent (e.g. aksiōn, genitive plural of aksiā ‘worthy (f.’)). The treatment of adjectives, that is, of items expressing prototypically qualities rather than entities (substantives) or activities (verbs), as similar or identical to substantives seems to be particularly characteristic of languages of Europe, northern Asia and northern Africa, that is, those languages with which Indo-European has been in areal contact and with which it may have distant genetic affiliations. According to ongoing typological investigations by R.M.W. Dixon, outside this area this treatment of quality-words is widespread only in Australia, parts of Mexico and California, and the Philippines; over most of the world quality-words are either treated like verbs or form a small, closed class of adjectives (with other quality-words treated either as verbs or as substantives). Participles, incidentally, are not strictly a separate word class: they are rather adjectives derived from verbs and thus showing some verbal categories in addition to the categories expressed in an adjective.

Many adverbs are etymologically, and sometimes even synchronically, case forms of nominals, for example, Lat. partim (< acc. sg.) ‘partly’, Skt divā (< instr. sg.) ‘by day’, or nominals with case-like suffixes, e.g. Ancient Gk pōthen ‘whence?’ However, there is also a set of primarily short adverbs for which at least no such substantival origin is clear, for example, *h₁en ‘up’ (cf. Ancient Gk anā), *upēr(i) ‘above’ (cf. Skt upāri), pro ‘forwards’ (cf. Skt prā). Many of the last-named items also show up as prepositions (or postpositions, which we may together call adpositions) and also as verbal prefixes in some early and most contemporary IE languages, but one of the striking typological characteristics of Proto-Indo-European seems to have been the lack of adpositions. The items in question were earlier adverbs, and their use as adpositions or prefixes is secondary. In Proto-Indo-European the grammatical and semantic roles of noun phrases were expressed primarily by means of case, with those adverbs that were later to become adpositions at best further specifying the semantics of the case in question; the reinterpretation in favour of an adposition governing a particular case came
later. Clear traces of the earlier system are visible in Homeric Greek, in comparison with the later use of the items in question as strict prepositions in classical Attic prose. Thus, the adverbial use of peri ‘about’ is clear in: gelasse dē pāsa peri khthōn (Iliad 19.362) ‘and all the earth laughed around’. While kātā ‘down’ appears as a verbal prefix in kāt-ēdomai ‘eat up’, it is still possible for it to be separated from its verb, as in hoi kātā boīs Hyperionos Ėellyoio ēsthion (Odyssey 1.8–9) ‘they ate up the cattle of Hyperion the Sun-god’. And even in adpositional usage, while preposition is the norm, postposition is also found, as in theōn ek thēspata ēidē (Iliad 5.64) ‘he knew the decrees from the gods’ (cf. Chapter 9, p. 258). This wide range of usages is still found with some items in English, for example, up, which is a preposition in up the Nile River, a prefix in the verb to uplift (used primarily in metaphorical senses), and an adverb in to lift up (used primarily in the literal sense; see Chapter 14, p. 448). And while across most of Indo-European such items, in becoming adpositions, became prepositions, there are also languages, such as Hittite, where postpositions are the norm, and individual items that are postpositional even in languages that normally have prepositions, for example, Lat. mēcum ‘with me’. Weakly developed adpositions, or adpositions that are for the most part clearly of secondary origin, alongside a well-developed case system are also characteristic of Uralic and Kartvelian languages, in this respect contrasting sharply with Semitic, with its limited case system (e.g. Classical Arabic has only nominative, accusative and genitive) and prepositions of no clear non-prepositional origin (e.g. Classical Arabic fi ‘in’, min ‘from’).

Although the numerals form a clearly defined semantic class, they are not to be treated as a single word class in morphological terms, combining items that are adjectival (or, in some instances in some languages, pronominal), invariable and substantival. Whether Proto-Indo-European had clause-combining conjunctions is questionable, since clause-combining probably originally involved the use of non-finite verbal forms (see Clause Combining, pp. 94–5); the most plausible candidate is the word for ‘if’, Ancient Gk ei, Goth. ei suggesting originally a locative of the anaphoric pronoun *h₁.e ‘that, the aforementioned’. The enclitic coordinating conjunctions *-kʷe ‘and’ and *-ue ‘or’ have widespread reflexes. Finally, Proto-Indo-European had a number of invariable particles and, of course, interjections.

**Morphological Categories**

As noted in Morphological Type, pp. 78f., despite the complexity of early IE morphology the actual number of categories expressed morphologically is rather small. While the reconstruction of the PIE forms for most of these categories is relatively straightforward, the reconstruction of their precise meaning, or more generally function, is typically fraught with greater difficulty, given the extent to which etymologically identical forms have diverged in use in different languages.
The only morphological categories expressed in nominals (nouns, adjectives, pronouns) are gender, number and case, and of these gender is inflectional only in adjectives and some pronouns, perhaps reflecting an earlier stage where gender was always derivational rather than inflectional and where adjectives were indistinct from nouns. The absence of a category of possession (i.e. affixes for ‘my’, etc.) is in striking contrast to Uralic and Semitic, although in Uralic, at least, the development of this category is probably quite recent (as seen, for instance, in the number of languages attaching possessive suffixes after case suffixes).

Speculations on the areal typological affinities of the gender system are hampered by uncertainty over whether the PIE system was basically masculine–feminine–neuter (as in Sanskrit, Ancient Greek, Latin, etc.) or animate–inanimate (as in Hittite: see Chapter 7, p. 177). While the formally distinct feminines in *-h₂ (e.g. *-ā < *-e-h₂, *-ī < *-i-h₂) are clearly a relatively late grammaticalization of a derivational pattern, it is not clear how late this process is. Of the areally related languages, gender is completely absent from Uralic and Kartvelian, while Semitic has a masculine–feminine opposition, including apparently idiosyncratic gender assignments reminiscent in type of the early (and many modern) IE languages; North-west Caucasian has a masculine–feminine–neuter system, but with the gender assignment of nouns predictable directly from their semantic humanness and their biological sex. The number system, singular–dual–plural, of Proto-Indo-European is widespread among the languages of the world, being found for instance in some Uralic and Semitic (though not Kartvelian) languages; simplification of such a system to a singular–plural opposition, as has happened over most of Indo-European, is again widespread cross-linguistically, again with parallels in Uralic and Semitic.

The case system of early Indo-European represents a moderate-sized inventory (seven plus the vocative, on the standard reconstruction), certainly richer than that of Semitic, but not nearly so rich as that of Hungarian or some Daghestanian languages. As one might expect typologically from an inventory no larger than this, other means were used to encode finer distinctions, recourse being had in particular to those adverbs that later became adpositions (Word Classes, pp. 79f.). Indeed, even looking at the early IE languages one can discern a continuing development away from reliance on case towards reduction in the number of cases or at least in their functional range and towards greater reliance on adpositions; witness, for instance, the demise of the accusative as a means of indicating motion towards (e.g. Lat. Römam ‘to Rome’), its replacement in this function by prepositional phrases (e.g. Vulgar Lat. *ad Roma(m)), and the restriction of its functions to the syntactic one of indicating direct objects. Extrapolating backwards, one might hypothesize that at an earlier stage the cases had even more semantic and less syntactic content. This would fit in with some other claims made on the basis of internal reconstruction with respect to early Indo-European.
For instance, the distribution of the PIE nominative singular suffix *-s is rather strange, being absent in particular from neuter nouns. There are three main kinds of case-marking systems for subjects and objects found across the languages of the world. In the ‘nominative–accusative system’, one case (nominative) is used for both intransitive subjects and transitive subjects (or agents), while another case (accusative) is used for the direct object (patient) of transitive clauses. In the ‘ergative–absolutive system’, one case (absolutive) is used for both intransitive subjects and the direct objects (patients) of transitive clauses, while another case (ergative) is used for the subject (agent) of a transitive clause. In the ‘active–inactive system’, one case (active) is used for both subjects (agents) of transitive verbs and for semantically more agent-like (whence: active) subjects of intransitive verbs, while another case is used for both direct objects (patients) of transitive verbs and for semantically more patient-like (whence: inactive) subjects of intransitive verbs. (See, for instance, Harris 1990 and references cited there.) The early IE languages are traditionally described in terms of the nominative–accusative system, although this loses sight of the fact that neuter nouns in fact never show a distinction between nominative and accusative and indeed in o-stems have a nominative–accusative inflection *-m which is identical to the accusative of masculine nouns of this declension. In Proto-Indo-European, the most clear-cut nominative–accusative system is found with the personal pronouns, which in the first person typically go as far as to have suppletion between the stem used in the nominative and that used in the accusative and other oblique cases, for example, Lat. nom. ego ‘I’, acc. mē, Skt nom. vayām ‘we’, acc. asmān.

If those paradigms where the so-called nominative singular is in *-s are not analysed, at least for early Proto-Indo-European, as nominative–accusative, the question arises whether they formed part of an ergative–absolutive or of an active–inactive case-marking system. Here, cross-linguistic studies of languages with split case-marking systems, that is, which combine two or more of the three case-marking systems in various ways, can throw light on the problem (Silverstein 1976). Crucial to this area of study is the so-called ‘Animacy Hierarchy’, which arranges noun phrases in terms of their inherent animacy, as follows (though individual languages may make more or fewer distinctions): first/second-person pronouns > human noun phrases > other animate noun phrases > inanimate noun phrases. In languages which have a split in their case-marking system conditioned by the Animacy Hierarchy and which combine nominative–accusative case marking with one or more other case-marking systems, it is always the case that the nominative–accusative system is used for noun phrases highest in animacy. This fits in perfectly with our observation that PIE personal pronouns, restricted to first and second persons, show the most clear-cut nominative–accusative case-marking system. In languages where the ergative–absolutive system is combined with other case-marking systems, it is always the case that the ergative–absolutive system is found with noun phrases lowest in animacy. This is not consistent
with the observation that PIE *-s is most characteristic of masculine nouns and least characteristic of neuter nouns, the latter being typically lowest in animacy in Proto-Indo-European. The only way in which the IE evidence can be reconciled with the cross-linguistic distribution of ergative–absolutive case marking would be to claim that *-s was originally an ergative case marker, that neuter nouns rarely or never appeared as agents of transitive verbs, and that neuter nouns thus rarely or never showed up in the form in *-s. Subsequently, the use of *-s spread from transitive to intransitive verbs, thus becoming a gender-restricted nominative. If, however, we assume that the *-s was originally a marker of the active case, then the need to adopt this kind of argumentation evaporates: the inflection was simply used to mark those noun phrases that are more ‘active’, whether subjects of transitive or intransitive verbs, thus including masculine nouns in preference to neuter nouns; this has the effect of assigning greater semantic content to this suffix. The problem of the original function of *-s still cannot, however, be regarded as definitively solved: in languages with active–inactive case marking, for instance, the nature of the verb (dynamic or stative) is usually the crucial factor in determining the case of the intransitive subject, whereas on the active–inactive analysis of the PIE case-marking system the verb plays no role, only the inherent animacy of the subject noun phrase. In sum, while the PIE case system has anomalies in terms of the nominative–accusative case marking, the assumption of an ergative–absolutive or active–inactive case-marking system serves only to eliminate these anomalies at the expense of introducing others (see further Comrie 1994).

The major verbal categories are person–number, tense–aspect, mood, voice and finiteness, though with the exception of person–number the semantic characterization of categories is even more elusive with verbs than it is with nouns. The person–number inflections quite straightforwardly encode the person–number of the subject, irrespective of whether or not the subject is agentive, that is, following a strict nominative–accusative system; there is no inflection on the verb for any other argument, in contrast to the object agreement found in Semitic and some Uralic languages and the polypersonal agreement found in Kartvelian languages (subject, direct object, indirect object) and especially in North-west Caucasian languages (where verbs may in addition agree with adpositional objects).

On the basis primarily of Ancient Greek and Indo-Aryan, the conventional reconstruction of the PIE tense–aspect system comprises the following categories: present, imperfect, aorist, perfect. This system rests on three oppositions: that between non-past and past tense, seen clearly in the opposition between present and imperfect; that between imperfective and perfective aspect, seen in the opposition between present and imperfect on the one hand, and aorist on the other, and that between perfect and non-perfect, corresponding to the distinction between the perfect and the rest (for these terms, see Comrie 1976). Of these, the imperfect is clearly a more recent
development, essentially a past tense based on the present stem, so a more accurate picture might be the following: present, aorist, perfect. Since, in this system, there is no opposition of tense, it can be reformulated as imperfective, aorist (perfective), perfect. It is noteworthy that in this reconstruction, aspect is more important than tense, although by the time of the early IE languages tense, in particular the past–non-past opposition, has become crucial. Many branches of Indo-European have simpler systems, though often with traces of the earlier richness. Latin and Germanic, for instance, have no synchronically distinct tense–aspect as a reflex of the PIE perfect, although etymologically perfect formations lie behind the Latin preterite *pepercī *I have spared’ and the present nōvī ‘I know’, behind the Gothic preterite gaigrōt ‘I have wept’ and present wait ‘I know’. Hittite is unusual, for so ancient an IE language, in having not even traces of the earlier richness, since its verbal system operates in terms of a past–non-past opposition with no distinctions of aspect; whether this represents innovatory loss or the retention of an archaic system predating that of the IE languages remains unclear.

Two developments in Proto-Indo-European served to develop the past–non-past opposition in the system. One is the opposition between primary and secondary inflections. Despite the terminology, the secondary inflections are original, the primary inflections being formed from them with a final -i. The function of the final -i was to indicate present time reference, as in the Ancient Greek present tithēmi ‘I put (pres.)’ in contrast to the imperfect etithēn ‘I was putting, used to put’, with -n from *-m. The development of the primary inflections gives a form with clearly present tense, in principle sufficient to distinguish this from other time references. The second is the development of the augment, e-, perhaps originally a temporal adverb, to indicate past time reference, as in the Ancient Greek imperfect form just cited and in the aorist élyse ‘he loosed’. The augment is attested only in Indo-Iranian, Greek, Armenian and Phrygian, and may have been a local innovation. Thus, at least for those branches of Indo-European with the augment, there is a distinct present tense with -i and a distinct past tense with e-. If some branches never had the augment, then they are characterized throughout by a distinct present tense with -i.

In this general schema of tense in late Proto-Indo-European, the only incompatibility for the secondary inflections is with present time reference, thus leaving them in principle free to express either past or future time reference, or indeed absence of time reference. Traces of this situation can be seen in the use of secondary inflections with the optative (since wished-for states of affairs can be seen as standing outside real time) and in the Vedic injunctive, morphologically an augmentless past, which often has future time reference, for example, indrasya nū vīryāṇī prá vocam (RigVeda I.32.1) ‘I will now proclaim the heroic deeds of Indra’, kō no mahyā āditaye pūnar dāt (RigVeda I.24.1) ‘who will give us back to great Aditi?’ (see Chapter 4, p. 116). Subsequently, however, the secondary inflections came to be more
specifically tied to past time reference, at least in the indicative, so that in most early IE languages that retain the imperfective–perfective aspectual opposition we find the past–non-past opposition reflected in that between present on the one hand and imperfect and aorist on the other, and the imperfective–perfective opposition reflected in that between present and imperfect on the one hand and aorist on the other (Comrie 1990).

It will be noted that there is no future tense reconstructible for Proto-Indo-European. All of the forms that can be referred to as future tenses in IE languages are later formations, paralleling ways of forming future tenses that are found in many languages of the world (Ultan 1976), for instance modal formations (e.g. Lat. erō ‘I shall be’ from the subjunctive, Eng. I will go with an auxiliary originally meaning ‘want’ and I shall go with an auxiliary originally meaning ‘must’), inchoative formations (e.g. Ger. ich werde gehen ‘I will go’ with an auxiliary originally meaning ‘become’, OCS miněti načěšto ‘they will think’ with an auxiliary originally meaning ‘begin’), or simply the present tense with future time reference (e.g. Ger. ich gehe morgen ‘I (will) go tomorrow’). As in some other languages that lack a distinct future tense (e.g. Finnish, Hungarian, Georgian), the use of a present-tense form with telic or perfective meaning often gives future-tense reference, most clearly in the West and East Slavic languages, where the perfective of the present has basically future time reference, e.g. Russ. ja brošu (pfv.) ‘I will throw’, cf. ja brosaju (ipfv.) ‘I throw’. It will be noted that this presupposes the development of the PIE present tense, with its primary endings, into a more general non-past tense.

This leaves the perfect. Although in some early IE languages the perfect seems to have been well integrated into the verbal paradigm, for instance in Sanskrit and to a large extent in classical Attic Greek, there are clear signs that its relation to the rest of the paradigm was once much looser, the perfect being in origin a stative derivative of the verb stem, as seen for instance in the diathesis changes that characterize many perfects in relation to the rest of the paradigm, for example, in Ancient Greek peithō ‘persuade’ has as one of its perfects pépoitha, with the meaning not ‘I have persuaded’ but rather ‘I trust’, that is, as a result of having been persuaded. Where we find such diathesis changes, the phenomenon always involves expression of a state of affairs where the subject is affected. The inflections of the perfect in Proto-Indo-European are also close to those of the middle voice (see below). From this one can conclude that the IE perfect was originally a voice-neutral derived form and that its original value was to express a state, this state being attributed naturally to that entity most affected. Since the perfect was derived typically from dynamic verbs, its value was more specifically expression of a state resulting from a previous action, as ‘to trust’ is the state resulting from a previous action of persuasion. The changes in diathesis, while perhaps appearing idiosyncratic in the synchronic description of early IE languages, find a natural explanation on the basis of this assumption about
the original value of the perfect and general typological observations on the nature of resultative constructions (for which see Nedjalkov 1988). In many languages, the perfect subsequently shifted value semantically, de-emphasizing the resultant state and emphasizing more the earlier action, with the result that such forms often became general perfective pasts, supplanting or merging with the aorist, as in Sanskrit and Latin, or even general past tenses, as in Germanic. This is a diachronic shift that has taken place cyclically in a number of IE languages: for instance, Latin developed a new periphrastic formation with resultative meaning, for example, *habeō litterās scriptās* ‘I have the letters written’, which has eventually given the perfective past *ho scritto* ‘I wrote’ of spoken Italian.

We may now compare the reconstructed PIE tense–aspect system with that of other languages plausibly in areal contact or possibly genetically related to Indo-European. The closest parallel to the core present–imperfect–aorist systems of many early IE languages is perhaps found in the Kartvelian languages, with its core present–imperfect–aorist system. The basic distinction between past and non-past is found in nearly all languages of Europe, including Basque and the Finno-Ugric languages. The system is radically different from that found in the Semitic languages. However, it should be recalled that even within Indo-European we can plausibly reconstruct back from the system of such early IE languages as Ancient Greek to a rather different earlier system with much heavier emphasis on aspect than on tense, so the similarities between Indo-European and many of its neighbours may well be more representative of areal contact than of any deeper genetic relations that may exist. (For further discussion of the typology of tense–aspect systems in languages of Europe, see Comrie 1990.)

The distinct moods in Proto-Indo-European were the indicative, subjunctive, optative and imperative, perhaps also injunctive, although as a morphologically clearly distinct form the injunctive is limited to Sanskrit. Given the rather undeveloped state of cross-linguistic studies on the typology of mood systems, it is difficult to say much of typological relevance with respect to moods. The role of the subjunctive in the development of future tense forms has already been alluded to. One interesting observation with respect to the imperative is that both Sanskrit and Latin, and therefore probably Proto-Indo-European, distinguish between an imperative with more proximate time reference (e.g. Skt *ihū*; Lat. *ī* ‘go’) and an imperative with more distant time reference (Skt *itātī*; Latin *ūnō*).

In most general linguistic studies of voice, the paradigm case of voice has usually been taken to be the opposition between active and passive, that is, the relation as illustrated in English *The dog bit the cat* versus *The cat was bitten by the dog*. Following the terminology of Klaiman (1991), we may refer to this as ‘derived voice’, since the passive can be viewed as derived from the active (with preservation of at least the basic meaning). In Proto-Indo-European, however, there is no evidence for an active–passive
distinction, and the forms (synthetic or, more usually, analytic) that are specifically passive in individual IE languages are all clearly later developments. The only voice distinction characteristic of Proto-Indo-European is rather that between active and middle, a voice distinction of a type that Klaiman refers to as ‘basic voice’, since the active and the middle do not share the same meaning. The characteristic of the middle voice is that the subject is affected by the action of the verb, a distinction well illustrated by the traditionally cited opposition between Ancient Gk λῦω (act.) toûs aikhmalótous ‘I free the prisoners’ and λύμαι (mid.) toûs aikhmalótous ‘I ransom the prisoners’, where in the second example the subject gains benefit from the action. As shown by Klaiman (1991: ch. 2) such basic voice systems are by no means infrequent across the languages of the world. However, I am not aware of basic voice systems in the languages likely to have been in direct areal contact or possible distant genetic relation with Proto-Indo-European. The closest would perhaps be the systems of verbal derivatives (Hebrew binyanim) in the Semitic, and more generally Afroasiatic, languages. In addition, the Dravidian languages have a basic voice system very similar to that of Proto-Indo-European.

Non-finite verbal forms are characteristic of all the early IE languages and in most instances the parallels in formation are so clear that PIE forms can be reconstructed with minimal difficulty. As will be seen in Clause Combining, pp. 94–5, non-finite verbal forms played a crucial role in clause combining in Proto-Indo-European. Verbal adjectives (participles) include the present participle active (e.g. Skt bhárant-, Goth. bairands ‘carrying’), the perfect participle active (e.g. Ancient Gk eî̂dos ‘knowing’, OCS nesù ‘having carried’), the middle participle (e.g. Ancient Gk hepômenos ‘following’), and the *-to-/*-no- verbal adjective (e.g. Skt syūtā, Lith. siūtas ‘sewn’). Verbal nouns were of various morphological formations, originally part of the derivational rather than the inflectional morphology, though various case forms of such verbal nouns became generalized lexically and specialized syntactically to give infinitives, a category found in most IE languages but not going back to a single PIE ancestor form. There is less clear evidence for forms reconstructible to Proto-Indo-European that are specifically verbal adverbs (gerunds or converbs, in Slavicist and Turkologist terminology, with such meanings as ‘while doing . . .’), such forms as the Russian gerund (verbal adverb) čitaja ‘(while) reading’, etymologically a frozen nominative case form of a participle, being later language-specific formations. Most of the languages with which Proto-Indo-European may have been in areal contact or genetic affiliation have systems of non-finite forms, reaching particular richness in the Uralic and especially the Altaic languages, which latter in particular have verbal adverbs (usually called gerunds or converbs).
Syntactic Typology

Constituent Order

Much of the discussion of the syntactic typology of Indo-European in recent years has centered on the question of word-order typology, with quite different assessments being given on such questions as the order of subject (S), verb (V) and object (O) in Proto-Indo-European. Thus, Lehmann (1974) argues that Proto-Indo-European was basically an SOV language, with other constituent orders that one would expect, cross-linguistically, to co-occur most frequently with this word order, such as the adjective (A) and genitive (G) before the noun (N), that is, AN, GN, and adpositions (Ad) following the dependent noun phrase (NP), that is, NPAd. In terms of the dependency approach to syntax, in which each constituent typically has a head and one or more dependents (arguments), Proto-Indo-European would thus be a head-final language, with the head (verb of a verb phrase, noun of a noun phrase, adposition of an adpositional phrase) at the end of that phrase. Lehmann’s database consisted primarily of citations of examples consistent with these orders from the early IE languages, although he was, of course, aware that deviations from some or all of these constituent orders are frequent in the early IE languages: these deviations were to be treated in terms of departures in the individual early IE languages from the word order of Proto-Indo-European, or as marked constituent orders allowed in Proto-Indo-European. Friedrich (1975) argues, on the basis of a statistical analysis of the early IE languages, that for several of the constituent-order parameters there is at least as much evidence for assuming some other constituent order, with a trichotomization of the early IE languages into SOV (e.g. Indo-Aryan, Anatolian, Tocharian), SVO (e.g. Greek, on Friedrich’s assessment), and even VSO (Celtic, also Proto-Slavic on Friedrich’s assessment). Watkins (1976) suggests that it is methodologically unsound to rely on statistical analyses of early IE languages (perhaps especially given that some of the earliest attestations of individual branches of Indo-European are separated by thousands of years from the break-up of late Proto-Indo-European); rather, emphasis should be placed on expressions that can otherwise be judged as archaic in the attested texts. In cases where such archaic formulations can be clearly identified, they point rather clearly to Proto-Indo-European as basically an SOV language, and both this conclusion and Watkins’ methodology have been generally accepted in the field. The evidence for whether Proto-Indo-European was AN or NA and whether it was GN or NG is indecisive. As noted above, the proto-language probably lacked adpositions, so the question of postpositional versus prepositional belongs properly to the history of the individual branches.

It should be noted that many individual branches show changes in basic constituent order over their recorded histories. While in some cases this involves a movement from more SOV to more SVO, as with the Romance and
Germanic languages, it is equally likely to involve the opposite movement, as in Indo-Aryan: modern Indo-Aryan languages are much closer to consistent SOV order, and in particular to the requirement of verb-finality, than is Sanskrit. It is thus not possible to take directions of development found in the historically attested languages and project these same directions back indefinitely into the past.

While Proto-Indo-European was probably basically SOV, it also, on the basis of the attested early IE languages, allowed considerable freedom of constituent order, for instance with constituents being proposed for purposes of pragmatic highlighting. This relative freedom of constituent order was facilitated by the rich morphology, in particular the marking of case on noun phrases, permitting recovery of grammatical relations irrespective of constituent order, to a lesser extent by verb agreement with the subject, since this serves as an additional means of identifying the subject. As indicated in Agreement, pp. 91f., agreement among the individual words of a noun phrase, for instance, meant that it was even possible to move words out of constituents, so that an adjective might be separated from its head noun; in this sense, Proto-Indo-European can be characterized as having freedom of word order, and not just of constituent order (Chapter 14, p. 449).

One systematic set of exceptions to the general constituent-order principles outlined above is provided by enclitics, that is unstressed items that are necessarily pronounced as a single stress unit with the preceding stressed word (see especially Hittite, Chapter 7, pp. 187f.). Of particular interest are sentence enclitics, whose position is determined relative to the sentence as a whole, since in many early IE languages such enclitics appear as the second word in the sentence. This set of enclitics includes in particular unstressed forms of object pronouns, such as Ancient Gk *mé* ‘me (acc.)’, *moi* ‘to me (dat.)’, and a number of sentence particles, such as Ancient Gk *té* ‘and’. (In citing Ancient Greek enclitics in isolation, it is usual to mark them as if accented. Despite the use of the accent in all environments, such Ancient Greek particles as *gar* ‘for, because’ and connective *de* have all the word-order properties expected of enclitics.) This sentence-second position for enclitics has come to be called, at least in IE studies, Wackernagel’s Law after an early formulation of the regularity by Jacob Wackernagel (cf. Chapter 2, p. 70). In many instances the notion of sentence-second position is interpreted absolutely literally, so that an enclitic will, if appropriate, separate the constituents of a single phrase, as in Ancient Gk *hé gar eióthýá moí mantiké ...* (Plato, Apologia 40A) ‘for the divination that is customary to me ...’. Recent work on the behaviour of clitics cross-linguistically has established that second-position enclitics, including instances where this is interpreted literally as positioning after the first stressed word of the sentence, is widespread across the languages of the world (see, for instance, Klavans 1982).

Although recent discussions of PIE constituent order have concentrated primarily on the order of constituents, it is also worth commenting briefly on
the order of morphemes within the word. The inflectional morphology of Proto-Indo-European, and still of most modern IE languages, is almost exclusively suffixing, that of nominals exclusively so. The only inflectional verbal prefixes are the augment and reduplication; the only infix in Proto-Indo-European is the present stem infix -n- of the verb, for example, vincō ‘I conquer’, compare perfect vici and past participle passive victus. This restriction of inflections to suffixes is also found in the Uralic (and Altaic) languages, and is in striking contrast to the frequent use of prefixes in reflexional morphology of the verb in Kartvelian and Semitic languages and to an even greater extent in the inflectional morphology of the North Caucasian languages. The derivational morphology of Proto-Indo-European is likewise primarily suffixing: suffixing derivational morphology is found virtually without exception in Turkic languages and also for the most part in Uralic languages. Prefixing in the early attested IE languages consists primarily of preverbs (i.e. adverbs that have been fused to the verb as prefixes, see Word Classes, pp. 79f.), in which case one is dealing etymologically more with compounding than with prefixing, and it is significant that in those Uralic languages that have widespread prefixing, such as Hungarian, the prefixing is likewise primarily of this kind, for example, el-megy ‘go away’, where el- is in origin an adverb meaning ‘away’.

**Agreement**

One of the characteristics of the syntactic structure of the early IE languages is the extent to which the constituents of a multi-word phrase must agree. Within a noun phrase, adjectives (including attributive pronouns, such as demonstratives used attributively) must agree in gender, number and case with their head noun. (There is no agreement of a genitive with its head noun, a possibility found in a limited way in Kartvelian languages.) Agreement of adjectives with their head nouns is found in the Kartvelian and Semitic languages, though not in Turkic or, for the most part, Uralic languages, and indeed the Balto-Finnic branch of Uralic, which does have adjective agreement in noun phrases, is the branch of Uralic most subject to IE influence.

Within the clause, the finite verb must agree in person–number with its subject, while predicative adjectives must agree in gender, number and case with their subject. The major exception to this generalization is that finite verbs normally stand in the third-person singular (arguably the default form) when the subject is neuter plural in a number of early IE languages; however, this may simply reflect the fact that the neuter plural was originally a collective formation (analogous to the singular of ā-stems, i.e. *h2-stems), rather than a true plural. (Agreement of predicative adjectives in case is most clearly visible when the subject is in some case other than the nominative, for example, with the accusative subject of an infinitive in Latin, as in lēgem brevem esse oportet ‘it is fitting for a law to be brief’ (Seneca,
As already noted in *Morphological Categories*, pp. 81f., agreement of the predicate with the subject is found in most other languages of the area, including Uralic, Turkic, Kartvelian, North Caucasian and Semitic, although some of these, in particular Kartvelian, North Caucasian and Semitic, also allow indexing in the verb of arguments other than the subject, a possibility not found in the early IE languages and not reconstructible for Proto-Indo-European.

In terms of the typological distinction between head marking and dependent marking introduced by Nichols (1986), Proto-Indo-European is overwhelmingly dependent marking: within the clause, the relation between predicate (head) and arguments (dependents) is shown primarily by case marking the arguments; within the noun phrase the relation between the head noun and its attributes (dependents) is shown primarily by having the dependents marked to show agreement with the head noun. Subject–predicate agreement is the only instance of head marking, since here the predicate (head) is marked to show agreement with one of its arguments (dependents). As noted by Nichols (1986: 75–6), cross-linguistically head marking at the clausal level is not unusual in languages that are otherwise basically dependent marking, and patterns similar to that found in Proto-Indo-European are also found in most Uralic, in Semitic and in Kartvelian languages.

In a sense, it is perhaps misleading to refer to this phenomenon as agreement in Proto-Indo-European and many IE languages, since it is typically possible to omit the subject whose person–number is indexed in the verb and the head noun whose gender, number and case are indexed in the attribute, if the omitted items are recoverable from the broader context, as when Latin first-person singular *amō* means ‘I love’, third-person singular *amat* means ‘(s)he loves’, or masculine singular *bonus* means ‘good man’. Furthermore, it is possible for the constituents of a noun phrase to be separated from one another, for instance for the attributive adjective to be separated from its head noun, as in Lat. *Catōnem vidi in bibliothēcā sedentem multīs circumfúsūm Stōicōrum librīs* (Cicero, *De finibus bonorum et malorum*, III.2, 7 (536)) ‘I saw Cato sitting in a library, surrounded by many books of the Stoics’, where *Catōnem*, *sedentem* and *circumfúsūm* are all separated from one another and *multīs* is separated from *librīs*. On this basis, one might argue, following Hale’s (1982) characterization of non-configurational languages, that in Proto-Indo-European and at least residually in some early IE languages, items in agreement do not really form a single constituent, but are rather distinct constituents in apposition, that is, *puer amat* is ‘he, the boy, loves’ rather than ‘the boy loves’, and *bonus puer* is ‘the good one, the boy’ rather than ‘the good boy’. Over the recorded history of the IE languages, however, these possibilities for breaking up phrases in this way have decreased sharply, and the configurationality of most attested IE languages is hardly to be doubted.
Clause-structure Typology

At their earliest attestations, the various branches of the IE family show rather consistently nominative–accusative clause structure, with the same case (nominative) being used for both intransitive and transitive subjects, and a different case (accusative) being used for direct objects; verb agreement is consistently with the subject, whether transitive or intransitive. Attested deviations from nominative–accusative structure are of two kinds. First, some (originally) derivational forms have ergative–absolutive syntax in PIE languages, as in many other languages of the world, for example, the resultative participle in *-to-/*-no-, as in Skt gatá- ‘having come’ vs syūtá- (having been) sewn’ (see Nedjalkov 1988 for the typological background). Second, forms based on such elements are quite likely to have ergative–absolutive syntax, such as the periphrastic perfect in Sanskrit.

There are, however, certain features of Proto-Indo-European that appear somewhat anomalous against the background of the nominative–accusative system, and these have led a number of investigators to propose recently that Proto-Indo-European at some stage had either ergative–absolutive or active–inactive clause structure (Gamkrelidze and Ivanov 1984: ch. 5). Perhaps the most striking piece of evidence is the existence of the nominative singular in *-s, occurring in many declensional classes but noticeably absent from neuter nouns (see Morphological Categories, pp. 83f.). However, even if it is accepted that the case-marking system of Proto-Indo-European was ergative–absolutive or active–inactive, it is important to note one phenomenon that has been noted repeatedly in the cross-linguistic study of case-marking systems in relation to clause structure: the case-marking system of a language does not necessarily mirror its clause-structure type (e.g. Anderson 1976). It is thus misleading to regard such terms as nominative–accusative, ergative–absolutive or active–inactive as providing uniform characterizations of the whole of a language’s structure.

In the early IE languages, there is no direct evidence for inherited non-nominative–accusative clause structure. With respect to ergative–absolutive structure, as noted in Morphological Categories, pp. 83f. there are problems even in assuming an ergative–absolutive case-marking system for Proto-Indo-European. In languages with active–inactive case marking, nouns appear in the active when they denote the active participant in an event (e.g. the man in the man hit the dog, or the man ran away), in the inactive when they denote the inactive participant in an event (e.g. the man in the dog bit the man, the man died, or the man is good). Crucially, a given noun phrase as single argument of an intransitive predicate can appear in either the active or the inactive. This is not, however, found in the early IE languages; even though there are a few impersonal verbs that take a single non-nominative argument (e.g. Lat. mē pudet ‘I am ashamed’), adjectives, which are the prime instances of inactive predicates, take nominative subjects. The direct evidence shows only that Proto-Indo-European had an opposition between more animate
nouns (with nominative singular in *-s) and less animate nouns, but distinctions of this kind are widespread in languages with nominative–accusative clause structure (e.g. the Bantu languages) and languages with ergative–absolutive clause structure (e.g. the Australian language Dyirbal), having no necessary correlation with active–inactive clause structure. The evidence of the early IE languages points to nominative–accusative clause structure, and the reconstruction of active–inactive or ergative–absolutive clause structure or even case marking for Proto-Indo-European involves a high degree of speculation. In areal terms, nominative–accusative case marking and clause structure is shared with the Uralic and Semitic languages, and more distantly with the Altaic languages, while Caucasian languages typically have a high degree of ergative–absolutive or active–inactive structure in at least case marking or verb agreement.

Clause Combining

Although subordinate clauses, introduced by subordinating conjunctions, are the prime means of combining clauses into longer sentences in most of the modern IE languages, very few of these conjunctions can be reconstructed back to Proto-Indo-European (cf. Word Classes, p. 81). One recent study (Beekes 1990: 266–7) lists only a handful of derivatives of the relative pronoun (*jo- or *kwo- according to dialect) plus the conditional *hxei ‘if’ (apparently the locative of an anaphoric pronoun), and it should be noted that no single reconstructed form is given for the relative pronoun (but see below on correlative relative clauses). It is thus highly plausible that Proto-Indo-European made relatively little use of finite subordinate clauses, preferring instead various non-finite constructions, which are richly reflected in the early IE languages.

Participial constructions serve readily as substitutes for many relative clauses, for instance substituting ‘the man having fed the dog’ for ‘the man that fed the dog’ and ‘the man having been bitten by the dog’ for ‘the man that the dog bit’. The forms of the participles are clearly reconstructible back to the parent language (see Morphological Categories, p. 88). They can also serve in adverbial function, most obviously in cases where the subject of the adverbial construction is the same as that of the main clause, for example, substituting ‘having fed the dog, the man went home’ for ‘when the man (had) fed the dog, he went home’; the appearance of special verbal adverb forms, such as the absolute participle in Sanskrit or the gerund in many Slavic languages, is a later development. However, a participle in the appropriate case can also be assigned to a noun in a case other than the nominative, thus substituting ‘the dog bit the having-fed-it man’ for ‘the dog bit the man when he (had) fed it’. Finally, the existence of so-called absolute constructions permits the expression of such adverbial notions even where there are no coreferential noun phrases in the two clauses, as when ‘the man having fed the dog (or: the dog having been fed by the man), the cat ran away’ is
substituted for ‘when the man (had) fed the dog, the cat ran away’. Although different early IE languages differ in the case used in the absolute construction (locative in Sanskrit, genitive in Greek, ablative in Latin, dative in Gothic, Baltic and Slavic), the basic syntactic pattern is shared by all these languages and seems therefore reconstructible back to the parent language.

While no infinitive as such is reconstructible for Proto-Indo-European, the infinitives of individual languages represent specializations of case forms of action nominals that are reconstructible back to the proto-language (see Morphological Categories, p. 88). Occasionally more than one case form survives, as in the contrast in Old Church Slavic between the infinitive (e.g. dělati ‘to do’, deriving from a dative) and the supine (e.g. dělat, deriving from an accusative). Thus Proto-Indo-European had verbal nouns rather than an infinitive, the infinitives of the individual branches representing dialect-specific integrations of verbal nouns into the inflectional system (Disterheft 1980).

The widespread use of non-finite devices for combining clauses links Indo-European particularly closely typologically to the Uralic and Altaic languages. Nearly all attested Altaic languages make almost exclusive use of non-finite means for combining clauses (the few exceptions being under clear IE influence, for example, the use of the finite subordinating conjunction ki borrowed from Persian in Turkish and, especially, Azerbaijani). In Uralic, finite subordinating clauses occur primarily in those languages that have been under the longest influence from Indo-European, for instance Balto-Finnic (including Finnish), where they are none the less paralleled by a rich set of non-finite constructions that are used especially in more literary and archaic styles. There is, however, one difference between Indo-European on the one hand and Uralic and Altaic on the other. In comparison with the rich syntactic and semantic range of non-finite forms found in the Uralic and, especially, the Altaic languages, the early IE languages have a much more restricted set of oppositions among non-finite verb forms, and there is no reason to suppose the Proto-Indo-European had a system that was much richer than this (given that the early attested languages have non-finite verb systems very close to one another).

One kind of finite subordinate-like construction that Proto-Indo-European, on the evidence of the early IE languages, may well have had is the correlative construction, in particular as a way of expressing relative clauses, as in constructions of the type ‘who/which man feeds the dog, he/the man is good’ for ‘he/the man who feeds the dog is good’. The richer set of subordinating conjunctions and subordinate clauses found in even the early IE languages plausibly reflects a further development of this kind of construction, whence the close formal links between many subordinating constructions and the relative pronoun (see further Haudry 1973).
References


Gamqrelige, Tamaz (= Gamkrelidze, T. V.) and Mačavariiani, Givi (= Mačavariiani, G. I.) (1965) Sonantta sistema da ablauti kartvelur enebsi, Tbilisi: ‘Mecniereba’; (includes Russian résumé; German translation, primarily of the résumé: T. V. Gamkrelidze and G. I. Mačavariiani, Sonantensystem und Ablaut in den Kartwelsprachen, Tübingen: Gunter Narr, 1982.)


