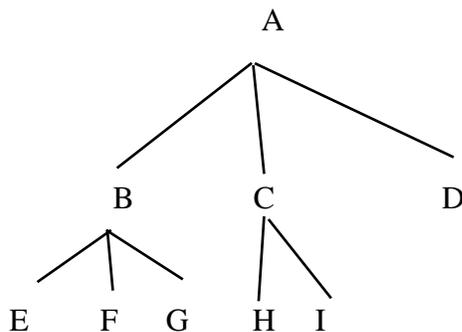# THE INDO-EUROPEAN FAMILY — THE LINGUISTIC EVIDENCE

by Brian D. Joseph, The Ohio State University

0. Introduction

A stunning result of linguistic research in the 19th century was the recognition that some languages show correspondences of form that cannot be due to chance convergences, to borrowing among the languages involved, or to universal characteristics of human language, and that such correspondences therefore can only be the result of the languages in question having sprung from a common source language in the past. Such languages are said to be "related" (more specifically, "genetically related", though "genetic" here does not have any connection to the term referring to a biological genetic relationship) and to belong to a "language family". It can therefore be convenient to model such linguistic genetic relationships via a "family tree", showing the genealogy of the languages claimed to be related.

For example, in the model below, all the languages B through I in the tree are related as members of the same family; if they were not related, they would not all descend from the same original language A. In such a schema, A is the "proto-language", the starting point for the family, and B, C, and D are "offspring" (often referred to as "daughter languages"); B, C, and D are thus "siblings" (often referred to as "sister languages"), and each represents a separate "branch" of the family tree. B and C, in turn, are starting points for other offspring languages, E, F, and G, and H and I, respectively. Thus B stands in the same relationship to E, F, and G as A does to B, C, and D. B and C, therefore, are proto-languages too, but are they are "intermediate proto-languages". It is thus possible to determine not only overall family relationship, but also degrees of relatedness — the languages in a single branch, e.g. E, F, and G, are more closely related to one another than each of them is to a language in a different branch, e.g. H, I, or D.

Many such language families can be recognized, and the languages within each exhibit striking formal correspondences — in their phonology, morphology, overall structure, and vocabulary — that link them together.

One of the earliest language families to be recognized, and thus the most thoroughly investigated of all to date, is the one that Greek belongs to, the one known as the Indo-European language family. The source language, generally called "Proto-Indo-European", was spoken some 6,500 years ago (see the article by J. P. Mallory) and has given rise to several hundred languages, in ten major branches. In the sections that follow, these ten branches are enumerated, with brief mention of some minor languages are given, the interrelationships among these languages are discussed, and the evidence linking the Indo-European languages together is presented.

1. The Major Branches of Indo-European

The ten major branches of the Indo-European family are listed here roughly in the order of their location east-to-west at the point of their first attestation, with an indication of the oldest representative languages in each and their earliest attestation (generally literary in nature).

1.1. Tocharian. Two closely related languages, generally referred to simply as TOCHARIAN A and TOCHARIAN B, make up this easternmost branch of Indo-European. Though extinct by the 10th century AD, these languages were discovered in documents dating from the 6th to 8th centuries AD that were found in the Central Asian region of Chinese Turkestan in the late 19th and early 20th centuries.

1.2. Indo-Iranian. This branch, consisting now of hundreds of modern languages found mainly in South Asia, is represented by two large sub-groups, IRANIAN and INDIC (also known as INDO-ARYAN), both with important ancient testimony. The earliest-attested Iranian languages are OLD PERSIAN, known from rock-cut inscriptions of the Achaemenid kings in the 6th to 4th centuries BC, and AVESTAN, the sacred language of Zoroastrianism, known from orally transmitted texts from at least as early as the 7th century BC. On the Indic side, the major representative is SANSKRIT, which in its most archaic form, known as Vedic, is the language of the orally transmitted sacred Hindu texts the Vedas, the oldest being the Rig Veda, conventionally dated to about 1200 BC. A vast literature in the highly archaic Vedic Sanskrit and in the somewhat later Classical Sanskrit, dating from the 6th century, including detailed native grammatical treatises, make this language especially important for Indo-European studies.

1.3. Armenian. Though showing considerable dialect diversity, this branch is represented essentially by a single language, ARMENIAN, spoken now mainly in the Armenian Republic

and in eastern Turkey, and attested from the 5th century AD through Bible translations into what is referred to as Classical Armenian.

1.4. Anatolian. Several of the ancient languages of Anatolia, in what is now Turkey, came to light in the early 20th century in cuneiform archives discovered at Boğazköy, east of Ankara, and were soon recognized as Indo-European languages. Though the oldest of these is PALAIC, an extinct language even in the 18th century BC, the best represented by far is HITTITE, attested from the 17th century BC up through the 13th century BC. Also important are LUVIAN, contemporaneous with Hittite but spoken into the first millennium BC, and LYCIAN and LYDIAN, both attested from as early as the 5th century BC. All of the Indo-European Anatolian languages were extinct by late Hellenistic times.

1.5. Greek. Like Armenian, GREEK is essentially a single language throughout its long history, yet constitutes a separate and distinct branch of Indo-European, though it too has considerable dialect diversity at all points in its history. Greek is attested first in Linear B texts from (perhaps) as early as 1400 BC, with the later Homeric texts showing considerable archaism as well.

1.6. Albanian. Attested quite late, only from the 15th century AD, ALBANIAN, in its two major dialects Geg and Tosk, is a separate branch of the Indo-European family. Its prehistory is most unclear, though some connection with an ancient language of the Balkans, possibly Illyrian or Thracian, is often assumed.

1.7. Balto-Slavic. This branch consists of two well-represented subgroups, the BALTIC languages and the SLAVIC languages. Grouping them together into a single branch is somewhat controversial but is generally accepted and is justified by some significant innovations they share, particularly in the accentual realm. The oldest attested representative of Baltic is the now-extinct OLD PRUSSIAN, attested from the 14th century AD, but the most substantial documentation for Baltic comes from LITHUANIAN and LATVIAN, both attested from the 16th century and still spoken today. The Slavic languages are attested from the 9th century AD, with the earliest text being a Bible translation prepared, at the invitation of the ruler of Moravia, by Cyril and Methodius of Thessalonica writing in a southern Slavic dialect now referred to as OLD CHURCH SLAVONIC.

1.8. Germanic. Three subgroups with important older representation make up the GERMANIC branch of Indo-European. The earliest attested Germanic is found in the so-called "Runic" inscriptions from as early as the 2nd century AD, though the evidence is sparse compared to the rich literary material of later centuries. Earliest among this richer documentation is GOTHIC, the sole (and now-extinct) representative of East Germanic, attested first through a 4th century AD Bible translation. West Germanic is represented by

OLD ENGLISH, attested from the 7th century AD, and by OLD HIGH GERMAN, attested from the 8th century. North Germanic is attested earliest in OLD NORSE, from the 12th century.

1.9. Italic. Along with Anatolian, Indo-Iranian, and Greek, the other branch of Indo-European with substantial attestation from before the common era is ITALIC, covering many of the languages spoken in ancient Italy. The primary representative of this branch is LATIN, attested first in the 6th century BC through short inscriptions with much more extensive documentation coming in the 3rd century BC and later (and note Latin's modern offshoots, the so-called "Romance" languages). Other Italic languages include FALISCAN, OSCAN, and UMBRIAN, all somewhat spottily attested and rendered extinct in ancient times by the spread of (Roman) Latin.

1.10. Celtic. The westernmost branch of Indo-European at the time of its first attestation is the group of CELTIC languages. Although Celtic languages were spoken over much of the western European continent in ancient times, with traces attested in GAULISH and CELTIBERIAN inscriptions from as early as the 3rd century BC, the main representatives of this branch are found in the British Isles. The most important Celtic language for Indo-European studies is OLD IRISH, attested in short inscriptions from the 4th and 5th centuries AD and in extensive literary documents from the 8th century; WELSH, too, is important, attested also from the 8th century.

2. Some Minor Indo-European Languages

Besides the languages mentioned above in §1, there are several rather poorly attested languages which are clearly part of the Indo-European family, to judge from what the scanty remains reveal about their vocabulary and grammar, but whose place in the family is not at all clear. Among these are the two languages mentioned above in connection with Albanian, THRACIAN and ILLYRIAN, for which the evidence consists mainly of place names, personal names, and some occasional words mentioned in ancient testimony.   Additional "Restsprachen", as these are sometimes called due to their residual nature, include PHRYGIAN, known mainly through tomb inscriptions in Asia Minor from the 6th and 5th centuries BC and the 2nd to 4th centuries AD; MESSAPIC, a non-Italic Indo-European language of ancient Italy, attested in inscriptions from the 6th to the 1st centuries BC; MACEDONIAN, the first language of Philip's ancient empire in what is now northern Greece, known mainly through glosses in late Hellenistic Greek lexicographical works and personal names; and several others even more spottily attested. For the most part, these languages add little to our overall understanding of the Indo-European family, though new finds may change that situation.

3. The Evidence Uniting the Indo-European Languages

As mentioned above, what provides the basis for positing an Indo-European family and for relating the various languages listed in §§1 and 2 is a set of striking correspondences of form among all these languages. These correspondences come at all levels of grammar, involving the sounds, the morphology, the lexicon, and the syntax. An especially significant aspect of these correspondences is that those involving sounds are regular, internally consistent (in the sense that they generally do not conflict with one another), and exemplified each by a large number of matching words and morphemes across the various languages, and that the matching words and morphemes show parallels in meaning and/or grammatical function.

For example, the words in the various languages for 'father', shown in (1), reveal several correspondences of sounds, as listed in (2), all of which recur in other sets of compared words, as indicated in bold-face in (3), some in different positions, showing they are to be defined more broadly (forms from languages not represented in (1) and (2) are given in parentheses in (3)); the forms in (3) also provide other (recurring) correspondences, some of which are indicated in (4) (NOTE: Ø = the absence of a sound in the relevant position; also, not all the languages are represented in these lists, though with other words they would be, and material irrelevant for the comparison is in parentheses; some simplification, e.g. regarding the Germanic reflexes in correspondences involving **t/th**, has been necessary here, though with no distortion of the overall results):

(1) Tocharian B *pācer* = Sanskrit *pitar-* = Armenian *hayr* = Greek πατέρ- = English *father* = Latin *pater* = Old Irish *athair*

(2)   a. word-initial position: p = p = h = π = f = p = Ø
    b. word-medial position: c = t = y = τ = th = t = th
    c. stem-final position: r = r = r = ρ = r = r = r
    d. vowels: ā = i = a = α = a = a = a
             e = a = Ø = ε = a = e = ai

(3)   a. 'mother': Tocharian B m**ā**c**er** = Sanskrit m**ā**tar- = Armenian m**ayr** = Greek μητερ- = English mo**ther** = Latin ma**ter** = Old Irish ma**thair** (cf. also
        Albanian mo**tr**ë ('sister') = Lithuanian mó**t**(ina) = Old Church Slavonic ma**t**i)
    b. 'feather, wing': Sanskrit **p**a**t**(t)**r**a- = Greek π**τ**ερο- = Old English **f**e**ther**
    c. 'heat, hot': Sanskrit **t**a**p**(as) = Latin **t**e**p**(-idus) = Old Irish **t**eØ(-ë) (cf. also Old Church Slavonic **t**o**p**(l-))
    d. 'seven': Tocharian B suk**t**- = Sanskrit s**apt**(-a) = Greek ε**πτ** (ά) = Latin s**ept**(em) (cf. also Hittite si**pt**(am-) = Albanian sh**t**(atë) = Lithuanian se**pt**ynì)

e. 'build': Sanskrit **ta**k(-s-) = Greek ΤΕΚ(-ΤΟΝ-) 'builder' = Latin **te**x(-ere)

'weave' (cf. also Hittite **t**ak(s-) = Old Church Slavonic **te**s(ati))

f. 'night: Sanskrit nak**t**(-am) **=** Greek nuk**t**- = Latin noc**t** = Old Irish (i-n)noch**t**

(cf. also Hittite nek^w**t**- = Albanian na**t**(ë))

g. 'who/what/which (of two)': Tocharian B k$_u$se = Sanskrit ka(-**tar**as) = (Ionic)

Greek ko(-**ter**os) = Albanian kush = Old Church Slavonic ko(-**ter**yj-) **=**

Old English hwæ(-**ther)** = Oscan pú(-**ter**ei) = Old Irish cía (and note also

Hittite k^{w}is = Latin k^{w}is [spelled *quis*])

(4) a. m = m = m = m = m (etc.)

b. n = n = n = n (etc.)

c. s = s = s (etc.)

d. Sanskrit/Hittite/Greek/Albanian/Latin k$^{(w)}$ = Tocharian k$_u$ = Germanic h$^{(w)}$

Evidence like this can be multiplied, giving sets of interlocking sound correspondences across all the branches that would be hard to explain by reference to chance, borrowing, or universality, leaving genetic relationship as the only explanation. The matchings of form and meaning are especially significant, all the more so when they involve recurring irregular patterns; for instance, the substitution evident in the English positive/comparative pair *good/better* is matched by German *gut/besser*; thus confirming a Germanic subgrouping, and the vowel-ful singular ('he/she is') vs. vowel-less plural forms ('they are') of the present tense of 'be' in (5) are similarly conclusive of a genetic relationship among the various languages, indicating a linguistic family:

(5) Sanskrit *as-ti* / s-anti = Hittite es-t^{s}i / (a)s-ant^{s}i = (Doric) Greek ἐσ-τί /ἐντί =

Old Church Slavonic *jes-tъ* / s-ǫtъ = Latin *es-t* / s-unt = Gothic *is-t* / s-ind **=** Old

Irish *i-s* / Old Welsh *h-int*

This evidence grows stronger when other forms are added in, such as the singular forms Tocharian B *ste*, Lithuanian *ês-ti*, and Albanian *(ë)shtë*, for these then constitute matchings like those in (3), except that they involve not just lexical items but grammatical markers as well, e.g. third person singular present *-ti* = *-t^{s}i* = -τι, etc., third person plural present *-anti* = *-ant^{s}i* = -εντι, etc.

Besides confirming the Indo-European family unity, these correspondences and matchings allow for the reconstruction of the sounds and forms of the parent language Proto-Indo-European, the reasoning being that the testimony of the offspring languages gives some insight into what the starting point must have been like. Thus, Proto-Indo-European lexical items and grammatical markers corresponding to the forms above are generally reconstructed as in (6), where the asterisk (*) indicates a reconstructed form:

(6) *pə̯ter- 'father', *mā̯ter- 'mother', *pet-(e)ro- 'feather', *tep(os) 'heat, hot', *sept(e)m- 'seven', *tek-s- 'build', *nekʷt- 'night', *kʷo/i- 'who, what', *-tero-'(of-two), COMPARATIVE', *-s 'NOMINATIVE.SINGULAR', *es- 'be', *-ti '3SG', *-onti '3PL'

and much more detailed and extensive reconstruction of Proto-Indo-European forms is possible once a greater amount of data is considered from other correspondences and matchings.

The overall evidence of these parallels among the Indo-European languages also allows for closer relationships to be discerned among the different branches (just as Balto-Slavic in effect links two sub-sub-groups together into a single branch). This is especially true when two or more branches share the same feature which is innovative with respect to the proto-language starting point. These dialectal groupings within Indo-European are, however, quite controversial and not well-worked out. Still, Italic and Celtic show special affinities, as do Greek and Armenian, and Balto-Slavic and Indo-Iranian, to name a few suggestive Indo-European internal groupings.

4. Conclusion — Beyond Indo-European?

The unity of the Indo-European family is beyond question and its recognition is one of the crowning successes of 19th and 20th century comparative linguistics. Questions still remain, however, as to the exact membership of the family, for with some poorly attested ancient languages, most notably Etruscan, spoken in ancient Italy, there simply is not enough positive evidence to count it among the Indo-European languages. More tantalizing yet is the question of deeper connections that Indo-European enters into, for just as the various branches of Indo-European are related to one another, having sprung from Proto-Indo-European, so too is it reasonable to suppose that Indo-European is related to other language families, at a greater time depth, and that it is merely a branch of a larger "super-family". And, there are suggestive parallels and correspondences that have led to many hypotheses. A favorite candidate for such a linkage with Indo-European is the Uralic family (including Finnish, Hungarian, Samoyed, and others), and a connection with Semitic (Hebrew, Arabic, etc.) has also been argued for. More broadly, many consider Indo-European to be part of a large family, often called Nostratic, that includes Uralic, Hamito-Semitic, Kartvelian, among others. Such connections are controversial, for they are difficult to prove conclusively, but it is clear that Proto-Indo-European cannot simply have emerged *ex nihilo* 6,500 years ago; the origin of language is just too much farther back in the distant past for Proto-Indo-European to be viewed as being at the dawn of human language (see the article by A.-Ph. Christides and S. Tsohatzidis).

5. Selected General References

Baldi, Philip. 1983. *An introduction to the Indo-European languages*. Carbondale &
    Edwardsville: Southern Illinois University Press.

Beekes, Robert S. P. 1995. *Comparative Indo-European linguistics: An introduction*.
    Amsterdam & Philadelphia: John Benjamins Publishing Co.

Buck, Charles D. 1949. *A dictionary of selected synonyms in the principal Indo-European
    languages.* Chicago: University of Chicago Press.

Burrow, Thomas. 1973. *The Sanskrit language.* London: Faber.

Collinge, N. E. 1985. *The laws of Indo-European*. Amsterdam & Philadelphia: John
    Benjamins Publishing Co.

Gamkrelidze, Thomas V. & Vjacheslav V. Ivanov. 1994. *Indo-European and the Indo-
    Europeans*. Berlin: Mouton de Gruyter.

Lehmann, Winfred. 1993. *Theoretical bases of Indo-European linguistics*. London:
    Routledge.

Meillet, Antoine. 1937. *Introduction à l'étude comparative des langues indo-européennes*.
    Paris: Hachette.

Palmer, Leonard. 1954. *The Latin Language*. London: Faber & Faber.

Palmer, Leonard. 1980. *The Greek Language*. Atlantic Highlands, NJ: Humanities Press,
    Inc.

Pedersen, Holger. 1959. *The discovery of language*. Bloomington: Indiana University
    Press.

Ramat, Anna Giacalone & Paolo Ramat, eds. 1993. *Le lingue indoeuropee.* Bologna: il
    Mulino.

Robinson, Orrin W. 1992. *Old English and its closest relatives: A survey of the earliest
    Germanic languages.* Stanford: Stanford University Press.

Sihler, Andrew. 1995. *New comparative grammar of Greek and Latin*. New York &
    Oxford: Oxford University Press.

Watkins, Calvert W. 1985. *The American Heritage dictionary of Indo-European roots*.
    Boston: Houghton Mifflin Co.

Department of Linguistics
Columbus, Ohio USA 43210-1298
[bjoseph@ling.ohio-state.edu]