1 Proto-Indo-European or Indo-Hittite

PROTO-INDO-EUROPEAN is the traditional name given to the ancestor language of the Indo-European family that is spread from Iceland to Chinese Turkestan and from Scandinavia to the Near East. A PROTO-LANGUAGE (Gk. πρῶτος ‘first’) refers to the earliest form of a language family presupposed by all of its descendants. There will forever be major gaps in our ability to reconstruct proto-languages, but as general linguistic knowledge becomes more sophisticated, so do the tools of reconstruction.

The so-called Anatolian subfamily, consisting of Hittite, a 2nd millennium language from central Turkey, and its immediate relatives from Turkey and the Near East, is by far the most archaic branch of Indo-European. Since Anatolian was the first subfamily to break off, the ancestor family is now commonly referred to as INDO-HITTITE. Another archaic branch is Tocharian, from Chinese Turkestan. This is widely recognized as the second branch to split off from the rest. Much of the evidence for this evolutionary history is recent, and the terminology is not yet fixed. Instead of Indo-Hittite, many scholars still prefer Proto-Indo-European (PIE) as the name of the earliest reconstructable ancestor language of this particular family. Figure 1 is a recent cladistic model (from Ringe 2006: 5) of the Indo-European languages.1

Figure 1: Major divisions of the Indo-European family

```
PIE
  /|
_/ /
  |  
Anatolian  North IE
  /|
_/ /
  |  
Tocharian  West IE
  /|
_/ /
  |  
Italo-Celtic  Central IE
      /|
     /  |
    Celtic  Italic
```

Central Indo-European consists of the other subfamilies, most important for our purposes being Greek, Indo-Iranian, Baltic, Slavic, and Germanic. Figure 2 is a geographical overview of the Indo-European family.

---

1 Cladistics (from Gk. κλάδος ‘branch’) is a phylogenetic system of classification that arranges organisms (and other entities) by their evolution. The cladogram in Figure 1 is a ‘tree-like’ reconstruction of the Indo-European languages by their evolutionary history.
Figure 2: Indo-European Language Family: geographical view
Figure 3 is a more elaborated cladogram of the Indo-European languages, from Ringe, Warnow, and Taylor (2002).

The model in Figure 3 embodies the observation that Germanic shares the most features with Balto-Slavic and Indo-Iranian. This is expressed by the fact that these three subfamilies branch from the same macronode. At the same time, however, there was considerable contact between Germanic and Celtic:

This split distribution of character states leads naturally to the hypothesis that Germanic was originally a near sister of Balto-Slavic and Indo-Iranian … that at a very early date it lost contact with its more easterly sisters and came into close contact with the languages to the west; and that that contact episode led to extensive vocabulary borrowing at a period before the occurrence in any of the languages of any distinctive sound changes that would have rendered the borrowings detectable. (Ringe, Warnow, and Taylor 2002: 111)

Such crossovers create difficulties for the family tree model, as discussed recently by Labov (2007).
2 Proto-Indo-European Phonological System

The main catalogue of information about the PIE phonological system is Mayrhofer (1986). Recent synopses in English are now available in Meier-Brügger (2003) and Ringe (2006). The inventory of contrasting segments is presented in the following chart:

PIE phonological segments

<table>
<thead>
<tr>
<th>Obstruents</th>
<th>labial</th>
<th>coronal</th>
<th>palatal</th>
<th>velar</th>
<th>labiovelar</th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td>t</td>
<td>k</td>
<td>k</td>
<td>kₜ</td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>d</td>
<td>ġ</td>
<td>g</td>
<td>gₜ</td>
<td></td>
</tr>
<tr>
<td>bʰ</td>
<td>dʰ</td>
<td>ġʰ</td>
<td>gʰ</td>
<td>gʰt</td>
<td></td>
</tr>
<tr>
<td>s</td>
<td>h₁</td>
<td>h₂</td>
<td>h₃</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sonorants</th>
<th>non-syllabic</th>
<th>syllabic</th>
</tr>
</thead>
<tbody>
<tr>
<td>m</td>
<td>m₂</td>
<td>m</td>
</tr>
<tr>
<td>n</td>
<td>ŋ</td>
<td>ŋ</td>
</tr>
<tr>
<td>r</td>
<td>r</td>
<td>r</td>
</tr>
<tr>
<td>l</td>
<td>l</td>
<td>l</td>
</tr>
<tr>
<td>y</td>
<td>i</td>
<td>y</td>
</tr>
<tr>
<td>w</td>
<td>u</td>
<td>w</td>
</tr>
<tr>
<td>e</td>
<td>o</td>
<td>a</td>
</tr>
<tr>
<td>ē</td>
<td>ŏ</td>
<td>ā</td>
</tr>
</tbody>
</table>

The voiced aspirates were kept only in Indic, where they remain to this day. They are typically described as breathy voiced but they are in fact voiced and aspirated, as demonstrated instrumentally by Prakash Dixit (1975).

The palatal series merged with the velars in all the geographically western IE languages, including Ancient Greek.

The laryngeals (⁎hl, etc.) were preserved only in Anatolian (see below). They also had a syllabic counterpart, generally represented as */a/, e.g. Skt. pitār- ‘father’ = Gk. patēr, Lat. pater, Eng. father < *pə-tēr-, phonologically */ph₂-tēr-. This is typically described rather as insertion of epenthetic *ə and deletion of the laryngeal (LHE 79).

The non-vocalic syllabic sonorants (also called resonants) subdivide into three categories. The glides */y, w/ alternate with vowels */i, u/. The syllabic nasals remained intact nowhere. In Greek and Sanskrit, they became a, in Latin en/in, em/im, and in Germanic un, um. So, for instance the PIE negating particle *p yields Gk. a(ᵣ)-, as in atheist or an-archist, Lat. in-, as in in-secure, and Gmc. un-, as in un-likely.
The syllabic liquids remained intact only in Indo-Iranian. In Germanic, they developed like the syllabic nasals, i.e. -*uR* (R = any resonant), e.g. PIE *wlkʷ*-o-s > Skt. *vṛkas*, PGmc. *wulf-az ~ *wolf-az > wolf.

3 Anatolian

Because of its archaic character, Anatolian is generally regarded as the first subgroup to separate from Proto-Indo-European. Many scholars now refer to PIE as Indo-Hittite (discussion in Meid 1979; Lehman 2001; Zeilfelder 2001; Melchert 1992, 2001).

One of the archaic features of Anatolian is preservation of at least some of the Indo-European so-called laryngeals, usually transcribed *h* in Hittite. The number and phonetic nature of these sounds is disputed (most scholars assume three), and they are variously transcribed. Watkins (2000), for instance, writes α₁, α₂, α₃; others h₁, h₂, h₃. There is also a convention that writes h₁ etc. when consonantal, α₁ etc. when syllabic (e.g. *po-tēr ‘father’ > Skt. *pītā / pitār, G *patēr, L *pater*), and H (or X) when the precise laryngeal is indeterminate or irrelevant. Some core reflexes follow.

\[
\begin{align*}
*h₁e & > /e/ & *ch₁ & > /e/ \\
*h₂e & > /a/ & *ch₂ & > /a/ \\
*h₁e & > /o/ & *ch₁ & > /o/ \\
*Ho & > /o/ & *oH & > /o/ \\
\end{align*}
\]

At least some long vowels result from contraction of a vowel plus laryngeal, and */o/ was not colored by any laryngeal. Since */h₁/ is not written in any of the Anatolian scripts and had no vowel-coloring effects, it was likely a glottal stop */h/. The vowel-coloring laryngeals may have been pharyngeals, e.g. */h/ for a-coloring, and lip-rounded */v/ for o-coloring (Beekes 1989, but see Job 1994; Rasmussen 1994b).


*hektes-* ‘is’: Hitt. ēš-zi, Gk. es-tí, Lat. est, PGmc. *isti > Goth. ist, E is

*sēh₁-mn ‘seed’ (Lat. sēmen ‘seed; SEMEN’) / collective *sēh₁-mō > PGmc. *sēmō > OS, OHG sāmo ‘seed’ (LHE 74)

*h₂ent-i ‘in front’ > *h₂ánti: Hitt. hānza, Gk. antí ‘against; ANTÍ-‘, Lat. ante ‘in front; ANTE-‘, PGmc. *andi ‘in addition; and’ > (O)E and

*peh₂ ‘protect; feed’ (Skt. pā-ti ‘protects’): *peh₂-trom / *peh₂-dhlon > *pah₂-trom / *pah₂-dhlon > Gmc. *fōdra- FODDER / Lat. pābulum ‘food; fodder; nourishment’ PAB(U)LUM; enlarged *peh₂-s- > *pah₂-s-: Hitt. paḥš- ‘protect’, Lat. pāstor ‘shepherd’ PASTOR
*-ēh₂ (factitive suffix) > *-ah₂, e.g. *new-ēh₂ ‘make new’: Hitt. *nēw-ahh₄ ‘renew; recopy’, Lat. (re)nov-ā-re ‘to renew’ RENOVATE
*peh₂wṛ ‘fire’ > *pah₂wṛ, Hitt. pāḥḥur, zero-grade *ph₂ur- > *puh₂r- (by metathesis) > Gk. pūr (Pyre); Gmc. *für-i- > OE ċyrt FIRE
*h₂ost- ‘bone’ > Lat. os / oss- (oss-ify); cf. *h₂ost-ef-o- > Gk. ostéon oṣtē(o)-; collective *h₂ést-ōi > Hitt. haštāi ‘bone(s)’
*h₂ep-en-e/ont- > *h₂op-en-e/ont- ‘rich’: Hitt. ḫappenant-, Lat. opulent-us
OPULENT (same root [*op-] as Lat. opera ‘works’ etc.)
*pehᵶ(i)- > *pohᵶ(i)- > *pö(i)- ‘drink’: Hitt. pāš- ‘take a swallow’ < *pehᵶ-r- (Hittite lost *h₁ in many environments); Lat. pō-tō ‘a drink’ POTION, POISON; zero-grade *phᵶ-r- > *p̥rᵶ-r- > Gk. *po-ti- > po-si- in sumpósion ‘a drinking together; drinking party’ SYMPOSIUM

Since our concern is largely with the historic languages, laryngeals are included only in the earliest reconstructable forms.

4 Tocharian

Tocharian (Toch.) was discovered in western China (Xinjiaäng) in Buddhist manuscripts (mostly translations). Ancestors of the Tocharians, with type O blood, light hair, and other Europoid genotypes, seem to be well preserved in mummies from the Tarim Basin [c. –1800+] (Mallory and Mair 2000; Day 2001: 353–5).

Tocharian texts [c.500–700] are in two very divergent dialects: East Tocharian (TochA) from the Karashar-Turfan region, and West Tocharian (TochB) from Turfan, Qarašahr, Šorčuq, and Kuča. TochB has, besides Buddhist texts, medical and magic texts, caravan passes, and other commercial documents. For some differences, compare the kinship terms: A pācār (GEN pācīr) : B pācār (GEN pāṭrī) ‘father’; A mācār (pl mācīr) : B mācār (OBL māṭār) ‘mother’; A pracēr (OBLpl pracēs) Krause-Thomas 1964: 2.14) : B procer (OBL protār) ‘brother’; A cēkār : B tkēcār ‘daughter’.

Though easternmost of the Indo-European languages, Tocharian shares features with Italic and Celtic (for instance, a mediopassive in -r) and patterns phonologically, in part, with western Indo-European in having a velar stop (e.g. TochA känt, B kante ‘hundred’ = Lat. centum [kēntum], G he-katōn, etc.) where the (other) eastern dialects have a palatal continuant, e.g. Skt. šatām, Lith. šimūtas ‘hundred’ < PIE *kṃtōm ‘id.’.
5 Germanic

Germanic separates itself from the rest of Indo-European by means of its consonant shift known as Grimm’s Law (cf. Grimm 1822), traditionally formulated as follows:

Grimm’s Law

\[
\begin{array}{cccccc}
\ p & t & k & k^w & f & h \\
\ b & d & g & g^w & b/b & ð/d \\
\ b^h & d^h & g^h & g^w & \ & \ \\
\end{array}
\]

\[\rightarrow \ \begin{array}{cccccc}
\ p & t & k & k^w & f & \chi/h \\
\ b & d & g & g^w & b/b & ð/d \\
\ b^h & d^h & g^h & g^w & \ & \ \\
\end{array}\]

EXAMPLES

*pēd/-pēd- ‘foot’ (Lat. ped-, Gk. pod-) : OE fōt FOOT (< *pōd-)
*pisk-o- (Lat. piscis ‘fish’) > Gmc. *fiskaz > Goth. fisk-s, Olce fisk-r, E fish
*gho-sth₂-i- (‘standing apart’ LSDE 48, w. lit.) > *ghosti- ‘stranger; guest; host’ (Lat. hostis ‘enemy’ [< ‘foreigner’]) > Gmc. *gastiz > ON gestr GUEST
*dékmd ‘ten’ (Gk. déka, Lat. decem; for the final dental, cf. krmp-óm ‘hundred’ < *(d)krmp-ó-)⁴ > Gmc. *teþun(t) > Goth. taihun [tehun], OE Angl. tēn(e) TEN
*kannabi- (Gk. kánabas ‘hemp; hemp seed’) > Gmc. *χαναπις ~ χαν(α)paz (HGE 159; LHE 297) > OE hænep, E hemp (both the plant and its fiber)⁵
*kʷod (Lat. quod ‘which, what’) > Gmc. *h*at, OE hwæt (> what)

---

² The reflexes of IE *gʷh are complicated. Internally, one reflex is /w/, as in *snoigʰh-o-s > Goth. snaiwōs, Eng. snow, but see also the discussion on *dhogʰ hos ‘day’ below. For word-initial position, a reasonable example is *gʰhōr-mo- (Skt. ghamā- ‘warmth’, Gk. thermós ‘hot’, Lat. for-ceps ‘firetongs’) > Goth. warmjan ‘(to) warm’, Eng. warm. This is now rejected by most scholars. For instance, AHDR 35 puts burn under *gʰh-er-, since this tradition claims that /h/ is the only reflex of word-initial *gʰh; cf. bane from Gmc. *banjǭ ‘wound’ < *gʰhón-veh̄- [geh-ven ‘strike’]; cf. Gk. phōnos ‘murder’. See Ringe (2006: 105–12). Polomé (1999b: 284) declares this view erroneous (cf. Huld 1998: 122), which is too strong; cf. *gʰbreh- ‘smell’ > Gmc. *brē- in OE breh ‘smell, vapor’ BREATH. Nevertheless, I find it difficult to discard the equation of Gk. thermós ‘hot’ (etc.) and Eng. warm. GED W36 derives warm from *wer- ‘burn, cook’, which otherwise attests no *mo- derivative. Perhaps at the stage *γʰor-mo- there was a spread of lip-rounding from the other segments in the word to strengthen the labialized component.

³ This root illustrates the guest-host reciprocal relationship in IE society (Benveniste 1969: i. 87–96). From Lat. hostis itself is borrowed E hosť, hostile, etc. An IE derivative *ghosti-pot-i- ‘guest-master’ (with *pot- ‘master’, as in Gk. despótēs [DESPOT] < *dems-pot- ‘master of the house’) yields Russian gospodin, an address form used as a title for non-Russians, and Lat. hospe(hospit-) ‘guest; host’ (whence hospitality, etc.; cf. Panagl 1992b: 313). Lat. hospit- enters English as host by way of OF hospit, along with derivatives OF hostel (HOSTEL) > F hôtel (HOTEL).

⁴ The second d in *dekmd for expected t is a due to word-final position and the preceding sonorant (Ringe 2006: 20, w. lit).

⁵ Gk. kánnabis, Lith. kanapēs, etc., were independent borrowings from a non-IE language. The plant was known to the Greeks from Scythia in c–5. Herodotus [c.484–25] describes its use as new. Streitberg (1943 [1896]: 136) and Voyles (1992: 78) speculate that the Germanic consonant shift applied around the time hemp was borrowed, i.e. during the –400s at the earliest, to accommodate *kanab- (cf. Pers. kanab).
*nizdó- (Lat. nidus ‘nest’ < *ni-sed- ‘sit-on’) > Gmc. *nistaz > OE nest

*kn-o-m (Gaul. caron ‘horn’ DLG 91) > Gmc. *hurną > runic horna, (O)E horn

*ğr̥h₂-n̥-m ‘crushed, ground’ (Lat. grānum ‘grain’) > PGmc. *kurną >

Goth. kaurn [korn], OE corn CORN

*g(e)rbh- ‘scratch’ (Gk. gráph-ō ‘I scratch, write’) > E carve

*bhrātēr (Skt. bhrātā, Attic Gk. prhrētēr [‘clansman’], Lat. frāter)

‘brother’ > Goth. brōpar, OE brōpar (> brother)⁶

*h, roudh-o- ‘red’ (cf. Gk. eruth-ró-s ‘red’, Lat. rub-ēsc-ent- ‘turning red’)

> Gmc. *rāudā-z > Olce rauð-r, OE rēad (> red)

*dhogʷ-h-o-s (Lith. dągas ‘summer heat’ [*dhegʷ-h- ‘burn’])

> PGmc. *daŋʷ/az > Goth. dags, OS dag, OE dæʒ DAY⁷

*sengʷ-h-on-om ‘chanting; to chant’ (cf. *songʷ-h-ē, > Gk. omphē ‘divine voice’;

*sóngʷ-h-os > PGmc. *sangʷaz > Goth. saggws [saŋgws], (O)E song) >

PGmc. *singwaną ‘to sing’ > Goth. sīgwan, ON syngva, OE sīgan SING

(note that PGmc. *gʷ from *gʰ h remained only after a nasal: LHE 106f.)

*gʷʰ-ti- ‘act of killing’ > *gʷʰunti- > PGmc. *gunþi ‘battle’ > ON guðr

(cf. remodeled OE gûþ ‘id.’)

The final example is supposedly important for establishing a relative chronology of the changes (Ringe 2006: 92, 106):

1) syllabic resonant > *-uR-

2) delabialization before /u/

3) Grimm’s Law (GL)

However, since *gʷ was the initial GL reflex of *gʰ h, other orderings remain possible, e.g. *gʰ hanti- > *gʰ niþi- (GL) > *gʰunþi- (*-uR-) > *gunþi-. The only change that cannot precede *-uR- is that of *gʷ to *b, or the result would be **bunþi-.

The second and third entries (*pisko-, *ghosti-) are instructive for the operation of Grimm’s Law (GL). They show that GL did not apply to a consonant preceding by /s/; cf. PIE *sth₂-tī- > *sto₂-tī-s ‘act of standing’ (cf. Lat. statiō ‘standing; STATION’) >

⁶ Goth. e and o are always long and therefore by convention not so marked; henceforth, a Gothic form such as brōpar will be written simply brōpar.

⁷ Also suggested is *dḥogʷ-o-s [*dhegh-]; cf. *dghyees- > Lat. heri ‘yesterday’ (Puhvel 1987; Bomhard and Kerns 1994, entry #82), but the relationship to Lith. dągas is appealing and accepted by many (Fraenkel 1962: 85f.; GED 86; Voyles 1992: 42, 53ff.; Kluge-Seebold 1995; Polomé 1997: 122; Orel 2003: 66). The older derivation from *aḡh²- plus initial *d- from the neuter demonstrative (cf. Dunkel 1987: 14), though continued in AHDR, is unlikely (Friedman 1999: 52) and not widely accepted. Blending of *dągʷaz with *agar (cf. Skt. áhar ‘day’ etc. < *aḡhr/n-) is plausible (GED 86f.).
PGmc. *staðiz ‘place’ > Goth. stāps, OE stede STEAD. This is reminiscent of the English distribution, according to which an initial voiceless stop is aspirated (pot [pʰat]) unless /s/ precedes (spot [spat]). In theoretical terms, the [+ spread glottis] feature does not spread over the entire cluster, as shown by the non-change of /t/ in *ōk tôw > Goth. ahtaU ‘eight’, etc. This is taken by Calabrese and Halle (1998) to be due to a temporal limitation on glottal opening, so that only the first segment of the cluster becomes aspirated, then spirantized. Notice that voiced consonants were not so affected: *nizdôs > *nîstaz > nest.

Given the change in Hellenistic Greek of /pʰ, tʰ, kʰ/ to /f, ã, χ/, and the similar change in Avestan (Skt. pathás = Avest. papô ‘way’ GENsg), it appears that the change from an aspirate to a continuant is a natural process. This is the motivation for one traditional formulation of the stages of GL:8

Stages of Grimm’s Law (one possible account)

1) Voiceless stop becomes aspirated unless an obstructant precedes
   *pisk-os > *pʰisk-as
2) Voiced stop becomes voiceless
   *dekʰmd > *tekʰmt (> *tekʰun)
3) Aspirates become continuants
   (and voiced continuants become stops in certain positions)
   *pʰisk-as > *fisk-as
   *tekʰun > *teχun (> Goth. taihuN [tehun])
   *bʰendʰ-an- > *b/bend-an- (> Goth., OE bind-an ‘to bind’)

Shortly after Grimm’s Law came Verner’s Law (VL). Karl Verner (1875) resolved some apparent discrepancies in the Germanic reflexes of Grimm’s Law. When the Indo-European accent followed, a continuant became voiced, e.g. *po₂tér (Skt. pitā, G patēr) ‘father’ > Gmc. *fapkēr > *fādēr > *fāðer > Goth. fādar, Olc fāðir, OE fēder (> E father; see Lutz 1991: 90ff.). Other examples:

*h2wap- ‘evil’ (Hitt. ḫuwappaš) / suffixed *h₂up-člo-s ‘evil, bad’ > *ufčelas (GL) >
*uβelaz (VL) > *uβelaz (accent shift) > *uβilaz (unstressed e-raising) > Goth.
ubils, OE yfēl EVIL (HGE 433; LHE 102)

---

8 There have been many formulations and reformulations of Grimm’s Law, some of which are contingent on a revision of the Indo-European consonant system. For instance, in place of the voiced series, a set of glottalics is often posited, in which case the voiced aspirates need not have been distinctly aspirated in PIE. These issues will not concern us here. For discussion, see, for instance, Mayrhofer (1986: 94–101); the contributions to Vennemann (1989); Salmons (1993); Beekes (1995: 130–4).
*tek-nó- (*tek- stretch out the hand; touch; receive’, hence ‘follower, retainer’ (Zimmer 2000) > Gmc. *peχnáž (GL) > *peynáþ (VL) > *péynaz (accent shift) > OE þeó(n) ‘follower, attendant, retainer, THANE’ (HGE 418)

*př-tū- (cf. Lat. portus [‘passage’ >] ‘harbor, port’ < *pértu- / *pṛteú- ‘crossing’) > *furþús (-uR-, GL) > *furðúz (VL) > *furðuz (accent shift) > OE ford

FORD (HGE 119; LHE 274), frequent with animal names, e.g. Ox(en)ford, Swinford, Shefford, Catford, Hartford, etc. (Gelling 1984: 67–72)

*teut-ēh, ‘tribe; people’ (OIr. túath, Oscan tauto ‘citizenry, state’) > *πευβό (GL) > *πευδό (VL) > *πευδό (accent shift) > Goth. πiuda, OE þeod ‘people, nation’ (HGE 423; LHE 103)

Note that in Old English, Gmc. */ð/ became /ð/ prehistorically. English /ð/ is in part by intervocalic voicing of /þ/ in Old English, and partly from Old Norse.

Verner’s Law is explained by Calabrese and Halle (1998) as follows. Given the hypothesis that voicing and pitch depend on vocal fold tension, the claim is that [+stiff vocal folds] yields [–voiced] obstruents and high(er) pitch in sonorants. Stressed vowels typically bear high tone and unstressed bear low tone (low tone is unmarked for vowels). Therefore a stressed vowel is assigned [+stiff vocal folds] and an unstressed vowel [–stiff vocal folds]. Verner’s Law, then, is an assimilation (spread) of the laryngeal feature [–stiff vocal folds] from the vowel to the nearest following continuant.

---

9 This is the root of Dutch, Germ. Deutsch ‘German’, etc., first attested in Gothic piudisko ‘like the heathens’ (translating G ethnikós ‘heathenly; foreigny’), a derivative in *-isk- (> E -ish) of piuda ‘people; nation’ (< Gmc. *πευδό < *teutā ‘people’; cf. Caesar’s Teutoni ‘Teutons’, etc., probably of Celtic origin). Compare the application by Christians to themselves of the Jewish term Gentiles for ‘heathens’ (L. gentilēs ‘people of the gentēs ‘tribes’) (Spitzer 1948: 171–8). From Gothic, piudisk- spread to the Romans of Gaul, who applied it to the language of the Franks, who adopted it. In Medieval Latin [768+], it is used of various Germanic languages. In the Strassburg Oaths [842], Lodhuinicus swears an oath rōmānā linguā ‘in the Roman(ce) language’ and Karolus swears teudiscā linguā. The preface to the Old Saxon Heliand uses Theudisca for Old Saxon. Otfrid [c.800-c.870] uses L. theoticē of the language of his Liber evangeliorum, but in the Germanic dialect portion he calls it frenkisg ‘Frankish’. ML theodisca was adopted into Old High German, where it became nativized as diutiscā/diutisk-, first attested as a gloss [c9–11], then in Notker, but remained rare until Middle High German. The name of the German and Dutch people, then, is a secondary substantivization of the adjective that was originally applied to their language by others. For details, see Must (1992).
6 Indo-European culture

Roots that exhibit archaic derivational patterns and/or have a broad range of occurrence around the Indo-European languages have a good chance of being lexical items of Proto-Indo-European. Useful lists of those roots appear in Buck (1988 [1949]), Delamarre (1984), and especially Mallory and Adams (1997).

Typical of agricultural societies, tools and farm implements were known (40+ entries), e.g. *h₂erh₁-tr-o-m Lat. arātrum ‘plow’; *yug-o-m YOKE, Lat. jugum subjugate. The Indo-Europeans tended (*k”el- ‘turn’, Lat. col-e-re / cultus CULTIVATE) various barnyard animals: *gh(a)id-o- GOAT; *uk”sōn OX; *gʰo-OW- COW, Lat. bōs / bov- BOVINE; *h₂ew-i-s EWE, Lat. ovis ‘sheep’ OVINE; *peku FEE ‘cattle, movable wealth’, Lat. pecū ‘flock’ PECUNIARY; *seuh-/*sū- (sow / swine) ‘adult pig’; *pork-o- ‘young pig’ FARROW / Lat. porcus PORK (Benveniste 1969: i. 27–36).

The very large number (100+) of reconstructable animals and the types best attested invite the conclusion that the Indo-Europeans dwelt in a forested region of Eurasia with a temperate climate. Congenial to this hypothesis are the wolf *w₁lkʷ-o-s, bear (*h₂tK-o-s Gk. árktos ARCTIC, Lat. ursus URSUS), elk *elk-, beaver *bhe-bhr-u-s, otter *udr-o-s (Gk. húdros ‘water-snake’ cf. HYDRA), eel *ang”h-i-s (Lat. anguis ‘snake’ ANGUINE, anguilla ‘eel’ ANGUILLIFORM), salmon [more likely, salmon trout] (*laks- or *lokas-o- LOX), goose *ghans-, duck (*h₂enh₁t- Lat. anas / anat- ANATIDÆ), bee *bhei-, etc. The absence of grouse excludes the far north, and absence of reconstructable words for crocodile, elephant, cobra, python, camel, monkey, lion, tiger, cat, ass would seem to suggest a temperate Eurasian homeland. If it could be proven that such words were not simply lost after the emigration of the Indo-European tribes (cf. Cowgill 1986: 66ff.), a southeast Asian, Levantine, or African homeland could be excluded, as also on the evidence of the absence of certain types of vegetation, e.g. palm, vine, olive, banana, banyan, bamboo, sandalwood, rice, lotus.

Some fifteen well documented tree types also suggest (in no way prove) a temperate forested area, the far north being excluded by absence of a common word for fir or spruce. Attested trees include beech *bhāg-o-, birch *bhērhḵ-, and oak (*perk”u- Lат. quercus QUERCINE). The problem is that most of the tree-names are shared with Finno-Ugric (and Altaic), e.g. *apsā- ASPEN (cf. Siberian Altai apsak), *ey-wo- YEW (cf. Vogul jiw ‘tree, wood’), and it is not clear whether the correspondences are due to (distant) genetic relationship, contact, or both (Campbell 1990). The apple *ab(e)l- was not PIE but a northern Eurasian word.

The common Indo-European words for snow *sneig’h-/*snoig’h-o-, ice *eis, and frost *prus-to- (cf. *prus(w)-ina Lat. pruīna ‘hoarfrost’ PRUINOSE) support a temperate climate, as do the seasons: winter *ghei-m- / *ghyem- (Lat. hiēms HIEMAL), spring
Keprhom (Dlfr in 2000: ogrum -r), together with the absence of a word for ocean or island, these considerations localize the Indo-Europeans in Eurasia.

Indo-Europeans knew how to build (*dem-) houses (*dom-o-s Lat. domus DOMICILE), possibly mainly of wood, on the evidence of the Germanic derivative, Gothic tim(b)jan ‘to build’, E timber, but specialization in a region of abundant wood is possible. They cooked (*pek*- Lat. coqu-e-re COOK) their foods. Their beverages included the honey drink MEAD *medhu (cf. Gk. mēthu ‘wine; mead’ AMETHYST), and possibly wine (*woin-o- Lat. vinum WINE/VINE). The limited distribution of the latter (Lat. vinum, Gk. (w)oīnos, Arm. gini, Hitt. wiwyana-) and the presence of the same word in Semitic (Arabic wain), Caucasian (Georgian ɣwino), and other language families (Gamkrelidze and Ivanov 1995: 557–62), favor a non-Indo-European beverage (pace Beekes 1995: 35). Mead is better attested: Bomhard and Kerns (1994: 665f.) set up a Proto-Nostratic\(^\text{10}\) root (#543) *mədəw-/*mədw- ‘honey, mead’, but no Proto-Nostratic root for wine, which was probably a Near Eastern/Mediterranean drink.

A full array of Indo-European weaponry is known (13 in Delamarre: *ag*si- AXE; *skei- / *skoī-o-m SWORD, Lat. scūtum ‘shield’ (E)SCUDO, etc.), but these and the limited sailing terms (*neh₂u- / *nāu- ‘boat’: Lat. nāvis ‘ship’ NAVAL; *mh₂-s- / *mazdo-s SAIL; *h,reh₁-t- SAIL, Gk. eretmōs ‘oar’; etc.) provide no evidence to localize the homeland.

7 Homeland of the Indo-Europeans

Despite the volumes written about the homeland of the Indo-Europeans, speculations range from Scandinavia to the steppes of southern Russia to Babylonia, and there is no agreement among scholars, succinctly summarized in Baldi (1999: 39–44) and Raulwing (2000: 67–75). The previous section outlined in broad terms evidence in favor of a Eurasian origin. This section presents a little of the more technical evidence.

One alternative to the Eurasian steppe origin of the Indo-Europeans is the Anatolian homeland (Renfrew 1988a, b; 2001). This hypothesis encounters several major problems. The first involves Greek vocabulary. If the Indo-Europeans migrated from Anatolia to Greece in the middle of the –7th millennium, “it is curious that the Greeks themselves, though ensconced on the Mediterranean littoral for five hundred years

---

\(^{10}\) Nostratic is one cornerstone of the hypothesis that all language families descend ultimately from a common ancestor. Proto-Nostratic would have been spoken ca.15,000, and includes Indo-European, Kartvelian (South Caucasian), Afroasiatic, Uralic-Yukaghir, Elamo-Dravidian, and Altaic (Bomhard and Kerns 1994: 34). For discussion, see the papers in Salmons and Joseph (1998).
before the first appearance of their language in the Linear B tablets, should nevertheless have been largely dependent on a foreign vocabulary for their characteristic flora and fauna…” (Jasanoff 1988: 802). The second problem is Hittite vocabulary. That the Hittites were relatively recent in Anatolia is suggested by the fact that Sumerian and Akkadian have almost no words borrowed from Hittite.11 By contrast, on the count of Tischler (1979: 266f.), Hittite has only some 420 words with good Indo-European etymologies versus 240 core vocabulary items of assured foreign provenience.

Lexicostatistics/glottochronology, the completely discredited method of measuring time depth by means of a list of putative common words, has recently been bolstered by computational methods derived from evolutionary biology. Using this model, Gray and Atkinson (2003) conclude that the initial break-up of Indo-European occurred no later than c.–8000. While Indo-Europeanists have objected to this early date, for the initial split of Indo-Hittite, it does not seem entirely unreasonable, given the major differences between the Anatolian subgroup and the rest of Indo-European. However, Gray and Atkinson also conclude that this supports Renfrew’s hypothesis, which is a total non-sequitur. In principle, any dating schema is compatible with any of the hypotheses about the proto-homeland of the Indo-Europeans.

Vocabulary and typological features shared with the Kartvelian languages12 of the Caucasus might localize the Indo-Europeans in northern Mesopotamia, but the lexical and grammatical similarities could indicate distant genetic relationship rather than contact (cf. Hayward 1989). This is possibly supported by the seventy Nostratic roots with seemingly regular correspondences in Bomhard (1990) or the potential 601 Proto-Nostratic roots in Bomhard and Kerns (1994).

Gamkrelidze and Ivanov (1984, 1990) claim that Indo-European cattle breeding, agriculture, and wheeled transport suggest a homeland in the Near East, to which these were confined until the end of M-4. This would be consistent with loanwords (e.g. *tauro- = Sem. *tawr- ‘bull’, *k(o)m- = Sem. *qarn- ‘horn’), if scholars could agree which words were borrowed.13 Moreover, association of the Indo-Europeans with horses and chariots is problematic. The ensuing discussion is from Raulwing (2000). The Indo-Europeans probably knew the horse (*h₁êkw-o-s / *h₁êk-wo-s Lat. equus ‘horse’

---

11 There was, however, a pre-equestrian IE language, which Whittaker (2004) calls Euphratic, that had settled in Mesopotamia between –3200 and –2500 and left some fifty loanwords in Sumerian, e.g. ṣuwi ‘ewe’ (*h₁ow-‘sheep’), ṣud(n) ‘ox, bull’ (*g₁ow- ‘ox, cow’), ṣubur ‘udder’ (*uḥ₃ur ‘id.’), ṣağ ‘arable tract’ (*h₂eg₄-o- ‘field’), ṣurūn ‘eagle’ (*h₂ur-en- ‘id.’), ṣurū ‘heav’y’ (*g₂rh₂-ū- ‘id.’), etc.

12 Root alternations like Kartvelian *derk’- : *drek’- : *drk’- ‘bend’ are claimed to resemble IE *derk’- : *drek’- : *drk’- ‘see’ (Gamkrelidze and Ivanov 1984: 252–63; Bomhard and Kerns, pp. 73–90).

13 For Bomhard and Kerns (1994: 327f.), ‘bull’ was a Proto-Nostratic word (#148: *₄₃jarw-), and ‘horn’ is derived from IE *kér- ‘head’; Gamkrelidze and Ivanov (p. 876) take the Semitic root from IE.
EQUINE), but the word could be a borrowing (pp. 106–9). Domesticated horses were probably brought to Anatolia by Indo-Europeans (pp. 33ff.). The speakers of PIE “were acquainted with the general principle of a wheeled vehicle drawn by paired bovids under the yoke” (p. 18). Nevertheless, these were carts or wagons; the spoked-wheel chariot, not to be confused with two-wheelers of M–3, first appears at the beginning of M–2 in central Anatolia, and later in the Levant, Mesopotamia, and Greece (pp. 14, 98ff.). The training of chariot horses, drivers, and warriors first became feasible in the city states of the ancient Near East, and this is supported archaeologically (p. 99). While individual languages adapted Indo-European roots to the chariot and its parts (*aks-axle, Lat. axis ‘id.’ [AXIS]; etc.), many of the terms are more easily explained as borrowings (pp. 109ff.). Even the wheel gives evidence of several reinventions, with different words (Huld 2000), one from the root *k*el-‘revolve’: *k*e-/*k*lo-: Gk. küklos ‘circle’ CYCLE; OE hweowol / hwæol wheel (~ hweogol) (LHE 108). These chronological facts are difficult for the hypotheses of Renfrew and Drews.

Proto-Anatolian and Proto-Indo-European seem to have split east of the lower Dnieper, “between the arrival of ox traction and the arrival of the wheel, probably in the first half of the fourth millennium” (Darden 2001: 220). Woolly sheep feature prominently in Darden’s account as well as that of Barber (2001) who, likewise, places Proto-Indo-European in the steppe region of Caucasus east of the Dnieper, where certain woollen goods originated c.–3000. The same plaid twills are identified also with the ancestors of the Celts in Hallstatt, Austria [1200–400], and the likely ancestors of the Tocharians near Hami in Chinese Turkestan [1200–700] (Barber 2001: 13). Ancestors of the Tocharians, with type O blood, light hair, and other Europoid genotypes, are preserved in mummies from the Tarim Basin [c.–1800+] (Mallory and Mair 2000; Day 2001: 352–5).

---

14 Against the borrowing of ‘horse’, see Jasanoﬀ (1988: 802). Still, a calque remains possible. It is unknown where and when the horse was ﬁrst domesticated (Levine 1999). In the Black Sea steppe region as late as the Early Bronze Age there is still no evidence that horses were harnessed or ridden (Rassamakin 1999). According to Drews (2004), horseback riding did not appear until the Iron Age. Nevertheless, horses are known since M–5 (cf. Huld 2007).

15 This form is problematic: *k*e-klaus > *h*e-halas (GL) > *h*yalaz (VL) > *h*e-welalz > OE hweowol. Ringe (2006: 108) attributes the g of hweogol to analogy with other w – g alternations. This assumes that it is accidental that an epenthetic /u/ developed in precisely the environment where */g* became [g] (§5 above). Surely it is at least as plausible that a regional early development of *h*e-walalz to *h*e-walalz was preserved only in an Old English alternant.

16 If the recent dates obtained by more sophisticated computational methods are at all accurate, the dates in this section will have to be adjusted. For instance, nothing precludes the possibility that Indo-Hittite split earlier and that the wheel spread as a major cultural item, necessitating the creation (or adaptation) of a word for it. The initial split could have occurred early and subsequent splits and migrations in the more generally agreed time-frame.
Rapid dispersal over Eurasia, “perhaps by elite dominance or folk migrations, would account for archaic traits cropping up even on the peripheries of the IE world” (Day 2001: 303). This is also consistent with the cranioskeletal evidence that the steppe groups of Eurasia essentially coincided with the regions of later Indo-European speakers (Day, pp. 306ff.). Light hair and skin pigmentation also correlates and presupposes endogamy at least in the upper classes (Day, pp. 306–10). On reaching certain areas, such as India and Anatolia, Indo-European spread by contact-induced language shift (Day, pp. 314ff.). This is further evidence against an Anatolian origin of the Indo-Europeans. Day concludes (pp. 317ff.) that the origin had to be in the Western Eurasian steppe.

PIE may have split c.–8000. In any event, the best evidence is that Indo-Europeans were dwelling in the Dnieper-Ural region (zone 1 in Mallory 2002) c.–5000. Anthony (1995) proposes an expansion of the Indo-Europeans from the Ukrainian steppes along the Dnieper c.–3500. The Hittite subgroup seems to have settled Anatolia in M–3, around the same time Indo-Europeans settled the Pontic steppe (the “European” branch). Greece and India were most likely settled in M–2 (Drews 1988: 41-55, 149ff.). Other migrations to account for dialect interconnections, as posited, e.g. by Gamkrelidze and Ivanov (1984, 1990) and others, remain speculative, given the difficulties of matching specific peoples with archaeological remains (Thomas 1991; Raulwing 2000).

References


DLG = Delamarre (2001).


GED = Lehmann (1986).


HGE = Orel (2003).

STREITBERG, WILHELM (1943 [1896]). *Ugermanische Grammatik*. Heidelberg: Carl Winter.


