

Piotr Garbacz  
University of Oslo, Oslo  
piotr.garbacz@iln.uio.no

## THE INDO-EUROPEAN FAMILY TREE AND ITS SPLIT – A REPORT FROM THE WORKSHOP *THE INDO-EUROPEAN FAMILY TREE*<sup>1</sup>

**Słowa kluczowe:** drzewo języków indoeuropejskich, rozpad praindoeuropejskiej wspólnoty językowej, italo-celtycki, greko-ormiański, języki germańskie  
**Keywords:** Indo-European family tree, split off of PIE, Italo-Celtic, Graeco-Armenian, Germanic

### Streszczenie

W niniejszym przyczynku autor prezentuje kilka schematów drzewa języków indoeuropejskich wziętych z aktualnych podręczników indoeuropeistyki i kontrastuje je z rezultatami warsztatów *The Indo-European family tree* (Uniwersytet Kopenhaski, 15–17 lutego 2017). Szczególną uwagę poświęca argumentom za przyjęciem istnienia podgrup italo-celtyckiej oraz greko-ormiańskiej, jak również pozycji gałęzi języków germańskich. Wspomina także użycie nowych metod archeologicznych, kladystyki obliczeniowej oraz badań DNA i ich ewentualnego znaczenia dla językoznawstwa historycznego.

There is little doubt today about the fact that Proto-Indo-European was spoken in the steppes of what today is Ukraine (the Pontic-Caspian hypothesis), a hypothesis that was recently re-corroborated in the works of Chang et al. (2015) and Haak et al. (2015). However, there is still some disagreement as to which branches of the Indo-European family tree split and in which order. The goal of this contribution is to show the newest trends in the research on the split off of the Indo-European languages and on the family tree of the language group, especially in light of a work-

---

<sup>1</sup> I would like to thank here Eirik Tengedal for his great assistance, both linguistic and technical during the writing process. I remain solely responsible for all errors of fact and interpretation.

shop held at the University of Copenhagen: *The Indo-European family tree* 15–17 February 2017 within the Roots of Europe project.

The outline of the paper is as follows. First, I present the standard opinions on the structure of the Indo-European tree and the split of the different branches as it is shown in a number of recent textbooks. Then, I refer to some of the topics that were discussed during the workshop and present the result of the discussions, before I conclude with a short summary and a discussion.

## **1. The Indo-European tree in some recent textbooks**

To be able to provide a brief overview of how the split off of several branches of the Indo-European language tree is presented in today's textbooks in Indo-European studies, I have taken a closer look at four of them. The standard model is the so called Big Bang-model, where no stand is taken on how and when the different branches separated from the stem. It is represented in some earlier works, e.g. Robinson (1992: 11–12). This model has been somewhat modified, as it is commonly held among Indo-Europeanists that the Anatolian branch separated first from the ancestor language and that the other branches separated later. That later model is sometimes called for the revised Big Bang-model and is the one on which there is a broad consensus. Further, there are a number of models with binary branching, but the main issue that remains is the order in which the different branches split off and which post-Indo-European proto-languages can be assumed to have come into existence based on the comparative evidence. Another question is whether one can get help from archeology, DNA-studies and other methods such as computational cladistics to determine the order of the split off. There has long been a consensus on proto-groups such as Indo-Iranian or Balto-Slavic, whereas there is less consensus on whether proto-groups such as Italo-Celtic or Graeco-Armenian can also be assumed to have resulted from the split off.

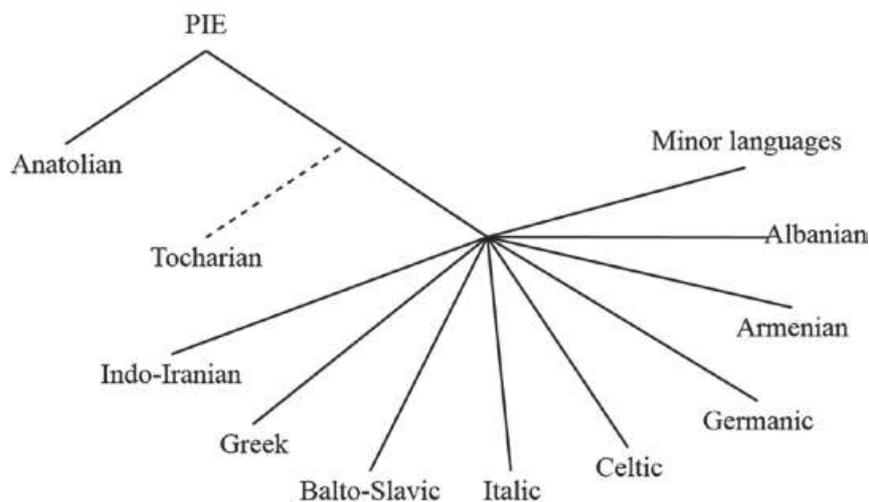
I have thus chosen four different, recent textbooks to examine how they deal with the question of the split off of the Indo-European and the form of the Indo-European tree that they present: (1) an introduction to comparative Indo-European linguistics by Beekes (2011), (2) another introduction to Indo-European language and culture by Fortson (2011), (3) a very recent introduction to Indo-European languages by Kapović (2017) and I have also taken a closer look at (4) a handbook that deals with the transition from Proto-Indo-European to Proto-Germanic (Ringe 2017).

### **1.1. Beekes (2011)**

Beekes lists arguments that have been proposed throughout the history of comparative linguistics for establishing genetic relations between languages relevant for the classification of the Proto-Indo-European evolution (2011: 14–15). He then moves on

to present information about the Indo-European languages and subgroupings, giving a short overview of the different branches of Indo-European in the order of their importance for the reconstruction of Proto-Indo-European (ibid.: 17–30). The order is the following: (1) Indo-Iranian, (2) Greek, (3) Anatolian, (4) Balto-Slavic, (5) Italic, (6) Celtic, (7) Germanic, (8) Tocharian, (9) Armenian, (10) Albanian, and (11) Minor languages which include (a) Phrygian; (b) Messapian; (c) Venetic; (d) Thracian; (e) Macedonian; (f) Illyrian. There are some inscriptions substantiating Phrygian, Messapian, and Venetic; there is hardly anything save names and glosses substantiating Thracian, Macedonian, and Illyrian. Next, he considers the splitting up of Proto-Indo-European into dialects (ibid.: 30–31). Anatolian, represented by Hittite, lacks several forms and categories that are routinely found in the other IE languages. This lack is today often explained through the assumption that some of these missing grammatical categories never existed in Anatolian to begin with, and that the other Indo-European languages jointly created them after the Proto-Anatolians had left the Proto-Indo-European homeland. This shows that Beekes (ibid.: 31) assumes that Anatolian did split off first. It is on the other hand uncertain whether Beekes (ibid.) assumes the split off of Tocharian as the next division in the Indo-European tree. Beekes also discusses (ibid.: 30–31) whether there are any reasons to postulate the existence of a Proto-Balto-Slavo-Germanic subgroup, given some shared features (e.g. the dative and instrumental plural beginning with an *m*-, the adjectival *sk*-suffix and others), but concludes that there is no pressing reason to conclude that Germanic and Balto-Slavic underwent a common development. He still assumes the existence of Proto-Balto-Slavic, given that both language groups share a host of developments in common, especially with respect to the accent (ibid.). Then, according to Beekes, the existence of an Italo-Celtic group (and thus Proto-Italo-Celtic) is much more difficult to prove, as the basis for such a supposition is formed only by a few sound changes and perhaps also morphological development. The Indo-European family tree in the view of Beekes (ibid.) would then have the following form:

Fig. 1. The Indo-European family tree according to Beekes (2011)



## 1.2. Fortson (2011)

Fortson (2011: 9–11) lists the following ten Indo-European branches: (1) Indo-Iranian, (2), Greek, (3) Italic, (4) Celtic, (5) Germanic, (6) Balto-Slavic, (7) Armenian, (8) Albanian, (9) Anatolian, and (10) Tocharian. He states that it is not clear whether a few other clearly Indo-European languages, such as Phrygian, Thracian, and Messapic, belong to any of the ten recognized branches, or constitute separate branches of their own (*ibid.*). Fortson represents the Indo-European family tree as a starburst, where the ancestor language is placed at the top, with the ten branches radiating out from it (*ibid.*: 10–11), see Fig. 2.

Fig. 2. The Indo-European family tree according to Fortson (2011)



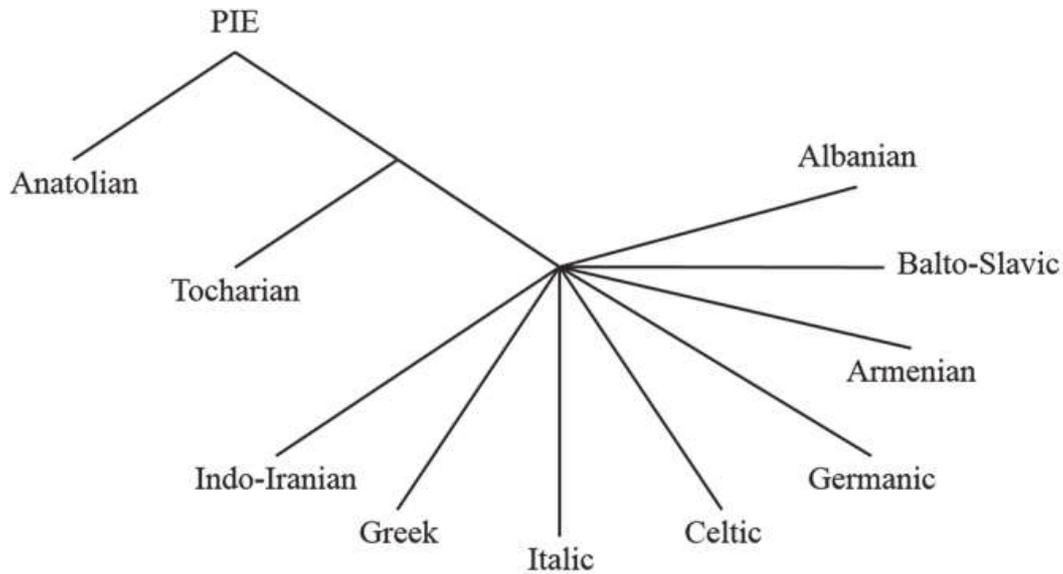
Implicit in this diagram is the notion of a more or less simultaneous “breakup” of the proto-language into ten or more dialect areas (resulting later in branches). Fortson (*ibid.*: 11) then mentions that this notion has often come under attack. For one thing, he says, the trees of most other language families tend to have a binary branching structure. In addition, it has been speculated for a long time that certain branches of IE should be grouped together into what are known as subgroups. For example, it has been proposed that Indo-Iranian, Greek, and Armenian all descend from one common dialect area of late Proto-Indo-European; and the same has been proposed for Italic and Celtic, for which an Italo-Celtic subgroup has been suggested. Importantly, Fortson notes that although many pieces of evidence have been presented over the years in defense of one or another subgrouping model, unfortunately the evidence is of varying quality and open to multiple interpretations. One example he mentions is the difficulty in separating an innovation of one branch that spreads by diffusion to the neighboring ones from a linguistic feature that is passed on from

an earlier stage ancestor to geographically contiguous branches (*ibid.*). Another difficulty with trying to determine the inner subgrouping of the family tree is that, by current estimates, the earliest documentation of IE languages (from the mid-second millennium BC) is still over two millennia later than the date of Proto-Indo-European itself (*ibid.*: 12). Fortson (*ibid.*) also mentions that some scholars claim that some innovations are due to influences of substrate languages spoken by the original non-IE-speaking populations of the territories into which IE speakers migrated, and not due to the usual processes of language-internal change.

### 1.3. Kapović (2017)

Like Fortson (2011), Kapović (2017: 3–5) lists the ten principal IE branches: (1) Indo-Iranian, (2), Greek, (3) Italic, (4) Celtic, (5) Germanic, (6) Balto-Slavic, (7) Armenian, (8) Albanian, (9) Anatolian, and (10) Tocharian. Furthermore, he divides fragmentarily attested IE languages into two basic groups: those that have at least one attested text/inscription and those that are attested only through onomastics or individual words in texts written in other languages. Kapović notes that, when reconstructing Proto-Indo-European, it is very difficult to separate different layers and determine whether an element should be reconstructed for Proto-Indo-European in general, dialectal Proto-Indo-European, post-Anatolian Proto-Indo-European (i.e. Proto-Indo-European after the early split of Anatolian), early Proto-Indo-European (i.e. Proto-Indo-European including Anatolian), post-Proto-Indo-European time, etc. (*ibid.*: 6). Other factors like sociolinguistic variables, dialect, gender, social status etc. also complicate the reconstruction. Kapović also considers the probability that Proto-Indo-European had dialects, since it likely covered a substantial ground. The dialectal picture of Proto-Indo-European is not completely clear because it is sometimes difficult to distinguish Proto-Indo-European dialectal features from post-Proto-Indo-European shared innovations and accidental independent innovations, he writes. Furthermore, he notes that it is now widely accepted that Anatolian was the first branch to separate from the rest of Proto-Indo-European – and adds in a parenthesis that Tocharian was the next (*ibid.*). He does not present a cladistic tree, but we can assume that such tree would have binary branching of Anatolian and Tocharian, as well as the Indo-Iranian and the Balto-Slavic subgroups. The tree as seen by Kapović (*ibid.*) is rendered in Fig. 3.

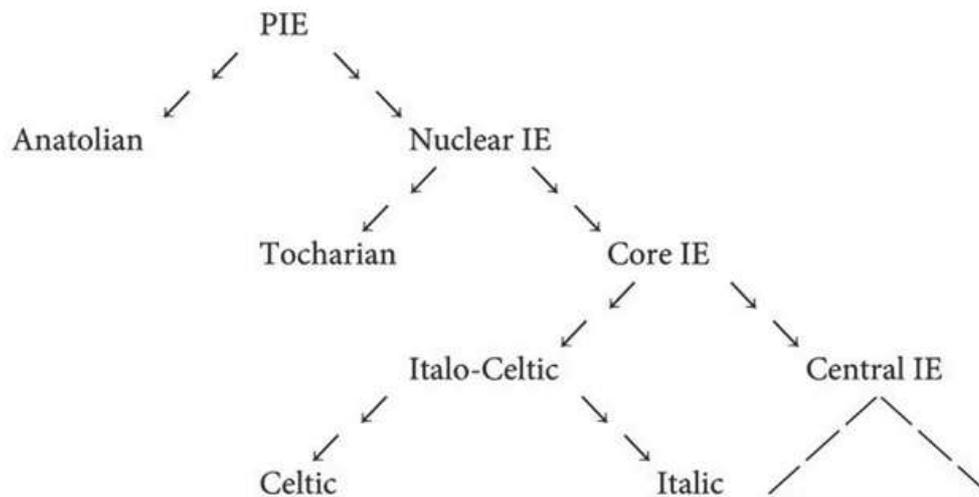
Fig. 3. The Indo-European family tree according to Kapović (2017)



#### 1.4. Ringe (2017)

Ringe gives a short introduction to Proto-Indo-European, stating that there is strong evidence pointing to the river valleys of Ukraine in the centuries between 5000 and 4000 BC (2017: 5). He states that there is currently a consensus among Indo-Europeanists that the Anatolian subfamily is, in effect, one-half of the IE family, all the other subgroups together forming the other half. Furthermore, it is beginning to appear that within the non-Anatolian subgroup, Tocharian is the outlier against all other subgroups (ibid.). He gives a probable cladistic tree of the IE family, rendered below in Fig. 4.

Fig. 4. The Indo-European family tree according to Ringe (2017: 7)



The Central subgroup contains the other branches, i.e. Germanic, Balto-Slavic, Indo-Iranian, Armenian, Greek, and probably Albanian, but Ringe is cautious about subgrouping these. He notes, however, that “it seems possible that Indo-Iranian, Balto-Slavic and Germanic were parts of a dialect chain at a very early date” (ibid.: 6).

### 1.5. Summary

There is also wide consensus about Anatolian being the first branch to split off, followed by Tocharian. As for the rest, some of the above-mentioned authors assume clustering of some language families, normally Indo-Iranian and Balto-Slavic, whereas others, like Ringe (ibid.), also propose an Italo-Celtic proto-group. Furthermore, the possible connection between Greek, Armenian and Indo-Iranian is mentioned by Fortson (2011) and another link between Indo-Iranian, Balto-Slavic and Germanic by Ringe (2017).

## 2. (Some) topics discussed at the workshop

The topics discussed can be divided into two subgroups: the first one is the question of whether computational cladistics is a useful tool for Indo-European studies (and diachronic linguistics in general), the second one the position of different languages or language groups: Anatolian, Tocharian, Italo-Celtic, Germanic, Greek and Armenian, Indo-Iranian, Albanian, and Balto-Slavic. As many arguments have been advanced against the use of cladistics in diachronic linguistics – one of the major obstacles being the irregularity of changes in terms of time and borrowings between languages, e.g. Evans, Ringe and Warnow (2006) and Chang et al. (2015) – I will focus on the problems of split off of different Indo-European branches and the subgrouping of these.

Two of the main arguments for assuming that Anatolian split off first are (1) the lack of a number of grammatical categories in Anatolian, which are present in the rest of Indo-European (especially in the tense system of the verb, but also in the nominal inflection), and (2) retained archaisms such as the presence of the Proto-Indo-European laryngals,  $h_2$  and  $h_3$ , in Hittite and Luvian. As for Tocharian, one argument for reckoning that the group did split off after Anatolian could be the lack of s-aorist in the group (e.g. Jasanoff 2003: 175), at the same time as Tocharian is much less archaic compared to Anatolian (the PIE laryngals are lost, for example, and the case system in Proto-Tocharian is strongly reduced compared to the PIE system<sup>2</sup>).

---

2 I speak here of the four cases in Tocharian inherited from PIE, the so-called *primary cases*, not of the later developed *secondary cases*.

The rest of the groups, Armenian, Albanian, Balto-Slavic, Celtic, Germanic, Greek, Indo-Iranian, and Italic, can be clustered together because of the fact that a coherent, inherited verb system (the *Cowgill-Rix verb*) can be reconstructed for these (Ringe 2017: 6). This inherited verb system cannot be reconstructed for Hittite and Tocharian (Jasanoff 2003).

As for the Italo-Celtic, many arguments in favor of assuming this subgroup were put forth by Schrijver (2017) during the workshop. The arguments come from common phonetic and morphologic development in Italic and Celtic, as e.g. Dybo's rule (shortening of long vowels before resonant followed by a stressed vowel), development of word-initial stress (also present in Germanic), development of \*CHC into \*CaC (cf. lat. *animus* 'soul, spirit' < \**anam-* < PIE \**h<sub>2</sub>enh<sub>1</sub>m-* > PCelt. \**anamū* > Middle Breton *eneff* 'soul'), creating superlative of adjectives and nouns with the suffix \*-*ismHo-*, genitive singular of o-stems on \*-*osjo*, and many others. Schrijver (ibid.) suggests that the Italo-Celtic group split into Proto-Italic and Proto-Celtic and that Proto-Celtic is thus a sister group of Proto-Italic, especially given the development of syllabic nasals, the long vowel system, and the PIE mediae aspiratae. Also archeological evidence points in this direction, as Schrijver (ibid.) views the Canegrate culture as intrusive and connected to the Urnfields cultures in what today is Western Italy, Southeastern France and Northwestern Switzerland. He concludes thus that the Celtic homeland was located most probably south of the Alps, not north of it.

The Germanic branch is difficult to place together with any other branch of Indo-European. There are many lexical similarities between Germanic and Balto-Slavic<sup>3</sup>, and some morphological similarities: the above-mentioned m-case, i.e. dative and instrumental plural beginning with an m-, and the adjectival *sk*-suffix, changing of accusative case into genitive case in the presence of sentential negation (Streitberg 1905/06) and others, but these can also be considered areal features, while shared developments are not sufficient to determine the exact place of Germanic within the tree. Still, the centum-reflexes of PIE palatovelars before a syllable containing a resonant in Balto-Slavic<sup>4</sup> (Kortlandt 2013) may be one of possible arguments in favor of grouping Germanic together with Balto-Slavic and Indo-Iranian (cf. Ringe 2017: 6, mentioned above).

Another question discussed at the workshop is whether one could assume Graeco-Armenian. The idea dates back to the beginning of the 20th century, but it has not gained support among the majority of Indo-Europeanists. At the workshop, Olsen (2017) advocated this view. Many of the similarities between Greek and Armenian

3 Mostly loans: many Germanic loans into Balto-Slavic, but few Balto-Slavic loans into Germanic, the Gothic verb *plinsjan* 'dance' (< OCS *plęsati*) being one of these few exceptions (Lehmann 1986: 273).

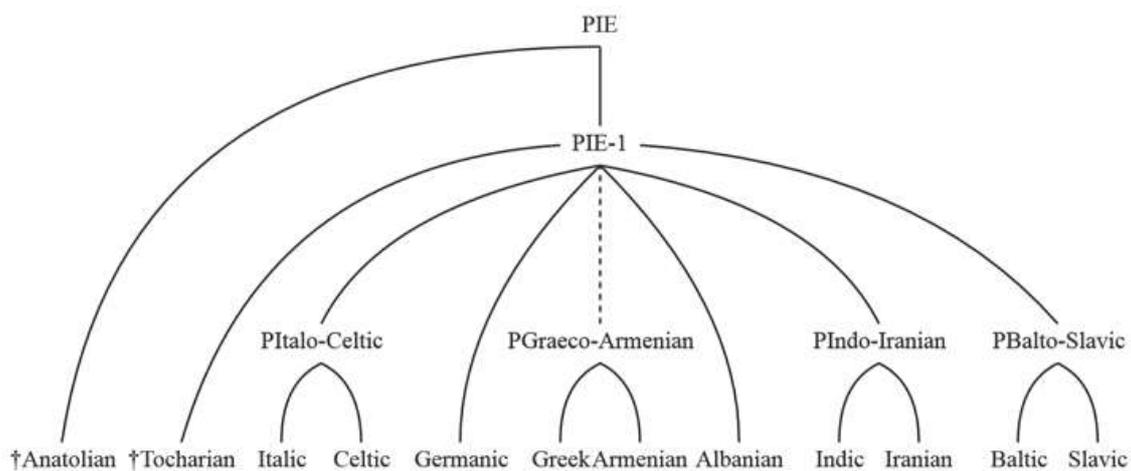
4 Polish *krowa* 'cow' < PIE \**korh<sub>2</sub>-ueh<sub>2</sub>* (with a centum-reflex of PIE \**k*) and Polish *sarna* 'row' < < PIE \**krh<sub>2</sub>-neh<sub>2</sub>* (with a satem-reflex of PIE \**k*), both etymologies according to Derksen (2008: 236, 485).

may point to a common development between these groups, but some may also be considered areal phenomena. To the discussed genetic similarities belong for instance palatalization of labiovelars, i-epenthesis, Greek iteratives in *-/k/-* and Armenian “weak aorists” in  $c^c < *s\hat{k}-, *m\grave{n}t-$  stems, the suffix *\*-ōdēs*, and others, especially many lexical-morphological correspondences.

The concluding discussion resulted in positive and negative results. On the positive side, there was a general agreement that Anatolian must have been the first branch to split off and that there are relatively good reasons to assume Italo-Celtic. The negative results are the general conviction that computational cladistics is not very useful, that Graeco-Armenian is not self-evident, and that any binary-branching tree after split-off of Anatolian rests on little evidence. Further, the assembly agreed that Tocharian, Germanic, Indo-Iranian, and Balto-Slavic are difficult to place and that the position of Albanian within the tree is uncertain.

If one would like to sketch an Indo-European family tree after the workshop, the model would be the following: Anatolian is the first branch to split off from the proto-language, thereafter a post-Anatolian Proto-Indo-European (PIE-1) is proposed, from which seven lines are drawn (from left to right): (i) Tocharian; (ii) Proto-Italo-Celtic divided further into Italic and Celtic; (iii) Germanic; (iv) possibly Proto-Graeco-Armenian divided further into Greek and Armenian; (v) Albanian; (vi) Proto-Indo-Iranian divided further into Indic and Iranian; (vii) Proto-Balto-Slavic divided further into Baltic and Slavic. Importantly, the model is based on shared innovations. This is shown in Fig. 5 below.

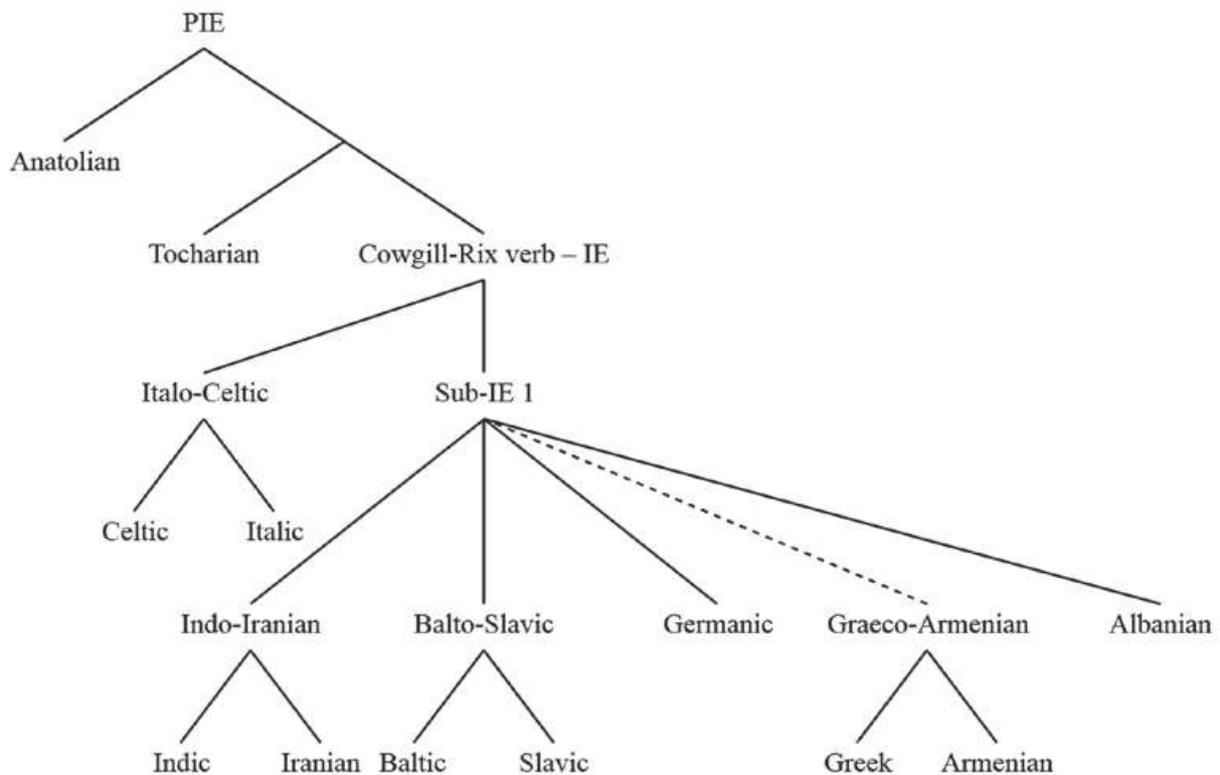
Fig. 5. The Indo-European family tree as agreed upon during the workshop



If one would like to merge the tree illustrated above with the line of argument and the tree presented by Ringe (2017), the result would be slightly different. Most importantly, the Italo-Celtic subgroup would split off after Anatolian and Tocharian, but before the rest of branches. Indo-Iranian, Balto-Slavic and Germanic would

be placed adjacent to each other, in order to take into consideration the idea of dialectal chain (ibid.: 6), but without forming of any subgroup. The last two groups would be Graeco-Armenian (possibly) and Albanian. The result would be as follows, see Fig. 6.

Fig. 6. The Indo-European family tree given the view of Ringe (2017) and the results of the workshop



It is important to keep in mind that one has arrived at the results shown above almost exclusively on the basis on language-internal evidence. However, the newest tendencies in the Indo-European studies try to make greater use of both archaeology and of DNA-studies of human remains (e.g. Allentoft et al. 2015 and Jones et al. 2015). If one can argue convincingly that there is a plausible historical connection between the DNA-profile and the language spoken by the individual(s), the DNA-analysis together with archaeological evidence and the methods of diachronic linguistics could give impressive results in the future. Such attempts have already been made, one recent example being Haak et al. (2015). The study can show that a major migration to Europe from the steppes north of the Black and Caspian Seas, connected to Yamnaya-culture and Corded ware-culture took place about 4,500 years ago. However, the authors are conscious of the fact that “the findings from ancient DNA are silent on the question of the languages spoken by preliterate populations” (ibid.: 210), but they also see a clear parallel between this migration and the supposed spread of the Indo-European languages assumed on the base of language-internal

data. Therefore, this may be the direction in which research is going, although this road is not always devoid of traps along the way.

## References

- ALLENTOFT M.E. et al., 2015, *Population Genomics of Bronze Age Eurasia*, “Nature” 522, p. 167–172, [on-line:] <http://dx.doi.org/10.1038/nature14507>.
- BEEKES R.S.P., 2011, *Comparative Indo-European Linguistics. An Introduction*, Amsterdam.
- CHANG W. et al., 2015, *Ancestry-Constrained Phylogenetic Analysis Supports the Indo-European Steppe Hypothesis*, “Language” 91(1), p. 194–244, [on-line:] <https://doi.org/10.1353/lan.2015.0005>.
- DERKSEN R., 2008, *Etymological Dictionary of the Slavic Inherited Lexicon*, Leiden.
- EVANS S.N., RINGE D., WARNOW T., 2006, *Inference of Divergence Times as a Statistical Inverse Problem*, [in:] P. Forster, C. Renfrew (eds), *Phylogenetic Methods and the Prehistory of Languages*, Cambridge, p. 119–129.
- FORTSON B.W., 2011, *Indo-European Language and Culture. An Introduction*, Hoboken.
- HAAK W. et al., 2015, *Massive Migration from the Steppe Was a Source for Indo-European Languages in Europe*, “Nature” 522, p. 207–211, [on-line:] <http://dx.doi.org/10.1038/nature14317>.
- JASANOFF J., 2003, *Hittite and the Indo-European Verb*, Oxford.
- JONES E.R. et al., 2015, *Upper Palaeolithic Genomes Reveal Deep Roots of Modern Eurasians*, “Nature Communications” 6, [on-line:] <http://dx.doi.org/10.1038/ncomms9912>.
- KAPOVIĆ M., 2017, *Indo-European Languages – Introduction*, [in:] M. Kapović, A.G. Ramat, P. Ramat (eds), *The Indo-European Languages*, New York, p. 1–9.
- KORTLANDT F.H.H., 2013, *Palatovelars Before Syllabic Resonants: Another Look*, “Baltistica” 48(1), p. 13–17, [on-line:] <http://dx.doi.org/10.15388/baltistica.48.1.2146>.
- LEHMANN W.P., 1986, *A Gothic Etymological Dictionary*, Leiden.
- OLSEN B.A., 2017, *The Greek-Armenian Connection*, Part I, Handout, *The Indo-European family tree*, Copenhagen, 15–17 February.
- RINGE D., 2017, *From Proto-Indo-European to Proto-Germanic*, Oxford.
- ROBINSON O.W., 1992, *Old English and its Closest Relatives. A Survey of the Earliest Germanic Languages*, London – New York.
- SCHRIJVER P., 2017, *Italo-Celtic and the Question of i[t]s Homeland*, Handout, *The Indo-European family tree*, Copenhagen, 15–17 February.
- STREITBERG W., 1905/06, *Gotica. 2. Got. ni waithts*, “Indogermanische Forschungen” 18, p. 401–404.

**The Indo-European family tree and its split –  
a report from the workshop *The Indo-European family tree*  
Summary**

The paper provides an overview of a number of recent presentations of the Indo-European family tree and the split off its branches, as found in standard textbooks on the topic. These views are contrasted against the results of a workshop, *The Indo-European family tree* (University of Copenhagen, 15–17 Feb. 2017). Our account specifically addresses whether there are reasons to assume the existence of Italo-Celtic, Graeco-Armenian and what is the position of the Germanic branch in the tree. The use of new archaeological methods, computational cladistics and DNA-studies and their possible importance for diachronic linguistics are also mentioned.