

1. INTRODUCTION

1.1. THE INDO-EUROPEAN LANGUAGE FAMILY

1.1.1. The Indo-European languages are a family of several hundred languages and dialects, including most of the major languages of Europe, as well as many in Asia. Contemporary languages in this family include English, German, French, Spanish, Portuguese, Hindustani (i.e., Hindi and Urdu among other modern dialects), Persian and Russian. It is the largest family of languages in the world today, being spoken by approximately half the world's population as first language. Furthermore, the majority of the other half speaks at least one of them as second language.

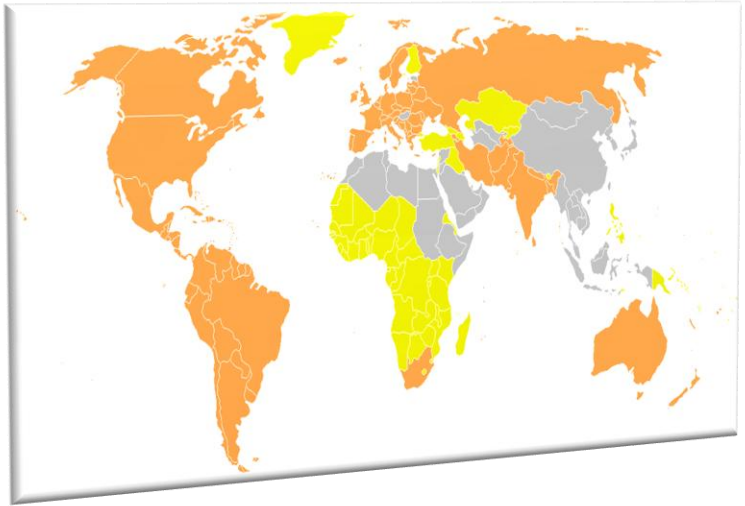


Figure 1. In **dark**, countries with a majority of Indo-European speakers; in **light color**, countries with Indo-European-speaking minorities.

1.1.2. Romans didn't perceive similarities between Latin and Celtic dialects, but they found obvious correspondences with Greek. After Roman Grammarian Sextus Pompeius Festus:

Suppum antiqui dicebant, quem nunc supinum dicimus ex Graeco, videlicet pro adspiratione ponentes <s> litteram, ut idem ὕλας dicunt, et nos silvas; item ἕξ sex, et ἑπτὰ septem.

Such findings are not striking, though, as Rome was believed to have been originally founded by Trojan hero Aeneas and, consequently, Latin was derived from Old Greek.

1.1.3. Florentine merchant Filippo Sassetti travelled to the Indian subcontinent, and was among the first European observers to study the ancient Indian language, Sanskrit. Writing in 1585, he noted some word similarities between Sanskrit and Italian, e.g. *deva/dio*, “God”, *sarpa/serpe*, “snake”, *sapta/sette*, “seven”, *ashta/otto*, “eight”, *nava/nove*, “nine”. This observation is today credited to have foreshadowed the later discovery of the Indo-European language family.

1.1.4. The first proposal of the possibility of a common origin for some of these languages came from Dutch linguist and scholar Marcus Zuerius van Boxhorn in 1647. He discovered the similarities among Indo-European languages, and supposed the existence of a primitive common language which he called “Scythian”. He included in his hypothesis Dutch, Greek, Latin, Persian, and German, adding later Slavic, Celtic and Baltic languages. He excluded languages such as Hebrew from his hypothesis.

However, the suggestions of van Boxborn did not become widely known and did not stimulate further research.

1.1.5. On 1686, German linguist Andreas Jäger published *De Lingua Vetustissima Europae*, where he identified an remote language, possibly spreading from the Caucasus, from which Latin, Greek, Slavic, ‘Scythian’ (i.e., Persian) and Celtic (or ‘Celts-Germanic’) were derived, namely *Scytho-Celtic*.

1.1.6. The hypothesis re-appeared in 1786 when Sir William Jones first lectured on similarities between four of the oldest languages known in his time: Latin, Greek, Sanskrit and Persian:

*“The Sanskrit language, whatever be its antiquity, is of a wonderful structure; more perfect than the Greek, more copious than the Latin, and more exquisitely refined than either, yet bearing to both of them a stronger affinity, both **in the roots of verbs and the forms of grammar**, than could possibly have been produced by accident; so strong indeed, that no philologist could examine them all three, without believing them to have sprung **from some common source**, which, perhaps, no longer exists: there is a similar reason, though not quite so forcible, for supposing that both the Gothic and the Celtic, though blended with a very different idiom, had the same origin with the Sanskrit; and the old Persian might be added to the **same family**”*

1.1.7. Danish Scholar Rasmus Rask was the first to point out the connection between Old Norwegian and Gothic on the one hand, and Lithuanian, Slavonic, Greek and Latin on the other. Systematic comparison of these and other old languages conducted by the young German linguist Franz Bopp supported the theory, and his *Comparative Grammar*, appearing between 1833 and 1852, counts as the starting-point of Indo-European studies as an academic discipline.

1.1.8. The classification of modern Indo-European dialects into ‘languages’ and ‘dialects’ is controversial, as it depends on many factors, such as the pure linguistic ones – most of the times being the least important of them –, and also social, economic, political and historical considerations. However, there are certain common ancestors, and some of them are old well-attested languages (or language systems), such as Classic Latin for modern Romance languages – French, Spanish, Portuguese, Italian, Romanian or Catalan –, Classic Sanskrit for some modern Indo-Aryan languages, or Classic Greek for Modern Greek.

Furthermore, there are some still older IE ‘dialects’, from which these old formal languages were derived and later systematized. They are, following the above examples, *Archaic* or *Old Latin*, *Archaic* or *Vedic Sanskrit* and *Archaic* or *Old Greek*, attested in older compositions, inscriptions and inferred through the study of oral traditions and texts.

And there are also some old related dialects, which help us reconstruct proto-languages, such as Faliscan for *Latino-Faliscan* (and with Osco-Umbrian for an older *Proto-Italic*), the Avestan language for a *Proto-Indo-Iranian* or Mycenaean for an older *Proto-Greek*.

NOTE. Although proto-language groupings for Indo-European languages may vary depending on different criteria, they all have the same common origin, the Proto-Indo-European language, which is generally easier to reconstruct than its dialectal groupings. For example, if we had only some texts of Old French, Old Spanish and Old Portuguese, Mediaeval Italian and Modern Romanian and Catalan, then Vulgar Latin – i.e., the features of the common language spoken by all of them, not the older, artificial, literary Classical Latin – could be easily reconstructed, but the groupings of the derived dialects not. In fact, the actual groupings of the Romance languages are controversial, even knowing well enough Archaic, Classic and Vulgar Latin...

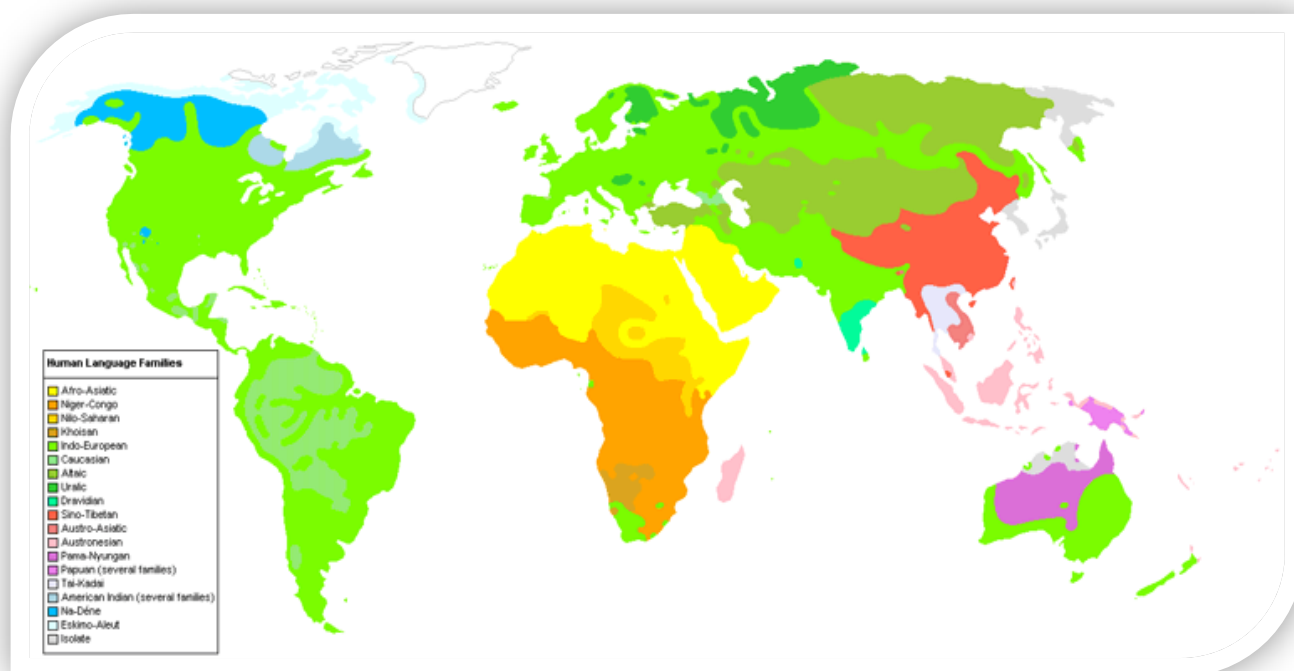


Figure 2. Language families' distribution in the 20th century. In Eurasia and the Americas, Indo-European languages; in Scandinavia, Central Europe and Northern Russia, Uralic languages; in Central Asia, Turkic languages; in Southern India, Dravidian languages; in North Africa, Semitic languages; etc.

1.2. TRADITIONAL VIEWS

1.2.1. In the beginnings of the Indo-European or Indo-Germanic studies using the comparative grammar, the Indo-European proto-language was reconstructed as a unitary language. For Rask, Bopp and other Indo-European scholars, it was a search for *the* Indo-European. Such a language was supposedly spoken in a certain region between Europe and Asia and at one point in time – between ten thousand and four thousand years ago, depending on the individual theories –, and it spread thereafter and evolved into different languages which in turn had different dialects.

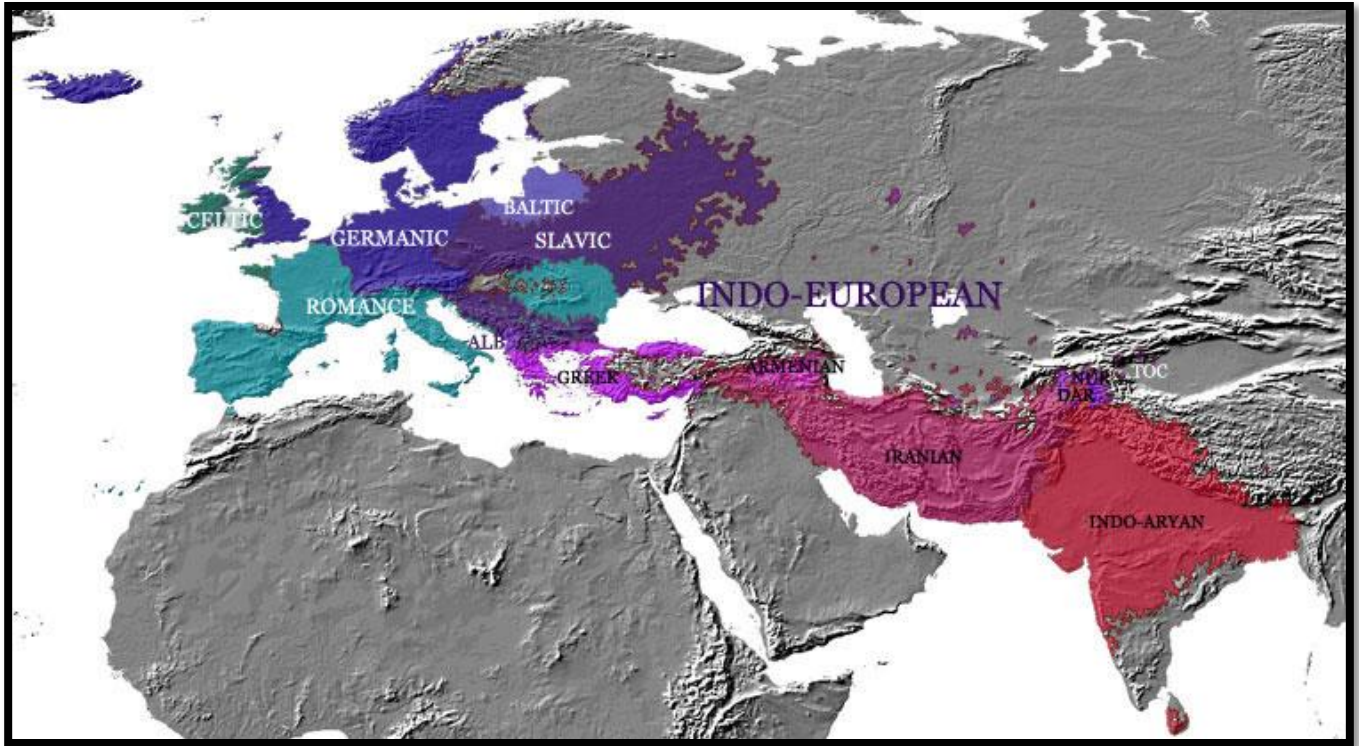


Figure 3. Eurasia ca. 1500 A.D. This map is possibly more or less what the first Indo-Europeanists had in mind when they thought about a common language being spoken by the ancestors of all those Indo-European speakers, a language which should have spread from some precise place and time.

1.2.2. The *Stammbaumtheorie* or Genealogical Tree Theory states that languages split up in other languages, each of them in turn split up in others, and so on, like the branches of a tree. For example, a well known old theory about Indo-European is that, from *the* Indo-European language, two main groups of dialects known as *Centum* and *Satem* separated – so called because of their pronunciation of the gutturals in Latin and Avestan, as in the word **kmtóm**, *hundred*. From these groups others split up, as *Centum* Proto-Germanic, Proto-Italic or Proto-Celtic, and *Satem* Proto-Balto-Slavic, Proto-Indo-Iranian, which developed into present-day Germanic, Romance and Celtic, Baltic, Slavic, Iranian and Indo-Aryan languages.

NOTE. The Centum and Satem isogloss is one of the oldest known phonological differences of IE languages, and is still used by many to classify them in two groups, thus disregarding their relevant morphological and syntactical differences. It is based on a simple vocabulary comparison; as, from PIE **kmtóm** (possibly earlier ***dkmtóm**, from **dékṃ**, *ten*), Satem: O.Ind. *śatám*, Av. *satəm*, Lith. *šimtas*, O.C.S. *sto*, or Centum: Gk. *ἐκατόν*, Lat. *centum*, Goth. *hund*, O.Ir. *cet*, etc.

1.2.3. The *Wellentheorie* or Waves Theory, of J. Schmidt, states that one language is created from another by the spread of innovations, the way water waves spread when a stone hits the water surface. The lines that define the extension of the innovations are called isoglosses. The convergence of different isoglosses over a common territory signals the existence of a new language or dialect. Where isoglosses from different languages coincide, transition zones are formed.

NOTE. Such old theories are based on the hypothesis that there was one common and *static* Proto-Indo-European language, and that all features of modern Indo-European languages can be explained in such unitary scheme, by classifying them either as innovations or as archaisms of that old, rigid proto-language. The language system we propose for the revived Modern Indo-European is based mainly on that traditionally reconstructed Proto-Indo-European, not because we uphold the traditional views, but because we still look for the immediate common ancestor of modern Indo-European languages, and it is that old, unitary Indo-European that scholars had been looking for during the first decades of IE studies.

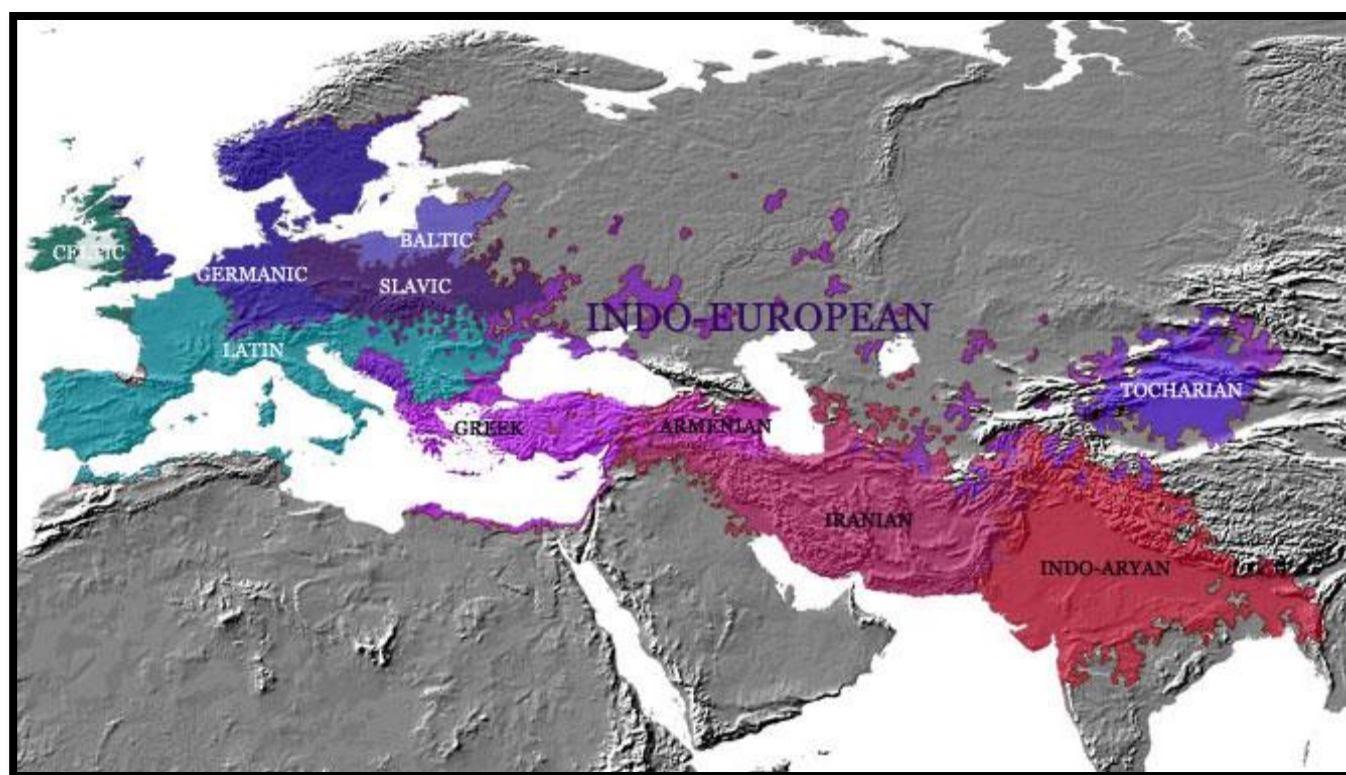


Figure 4. Indo-European dialects' expansion by 500 A.D., after the fall of the Roman Empire.

1.3. THE THEORY OF THE THREE STAGES

1.3.1. Even some of the first Indo-Europeanists had noted in their works the possibility of older origins for the reconstructed (Late) Proto-Indo-European, although they didn't dare to describe those possible older stages of the language.

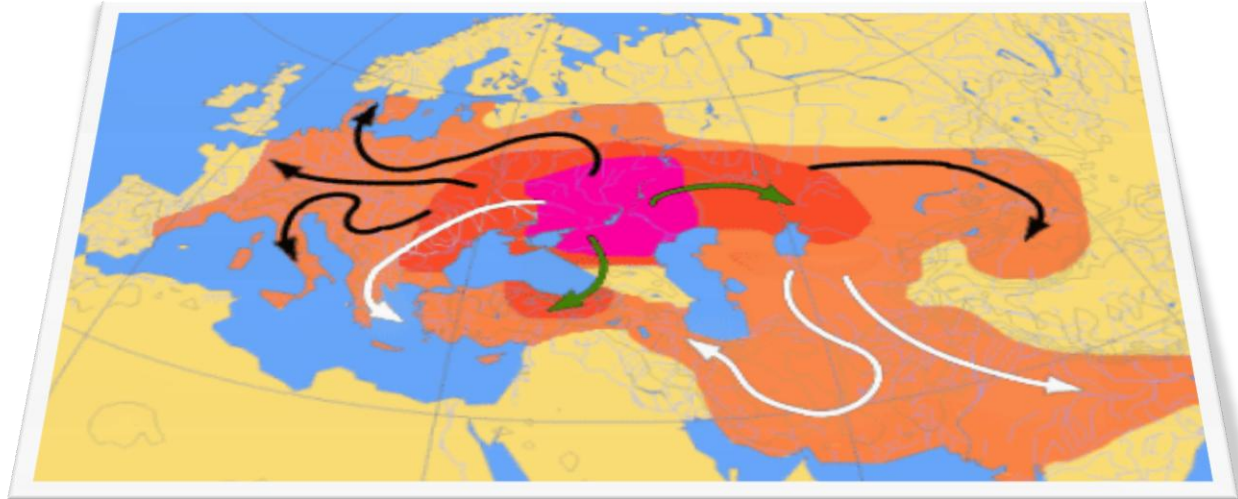


Figure 5. Sample Map of the expansion of Indo-European dialects 4.000-1.000 B.C., according to the Kurgan and Three-Stage hypothesis. Between the Black Sea and the Caspian Sea, the original Yamna culture. In colored areas, expansion of PIE speakers and Proto-Anatolian. After 2.000 BC, black lines indicate the spread of northern IE dialects, while the white ones show the southern or Graeco-Aryan expansion.

1.3.2. Today, a widespread Three-Stage Theory depicts the Proto-Indo-European language evolution into three main historic layers or stages:

1) Indo-European I or **IE I**, also called *Early PIE*, is the hypothetical ancestor of IE II, and probably the oldest stage of the language that comparative linguistics could help reconstruct. There is, however, no common position as to how it was like or where it was spoken.

2) The second stage corresponds to a time before the separation of Proto-Anatolian from the common linguistic community where it coexisted with Pre-IE III. That stage of the language is called Indo-European II or **IE II**, or *Middle PIE*, for some *Indo-Hittite*. This is identified with the early Kurgan cultures in the Kurgan Hypothesis' framework. It is assumed by all Indo-European scholars that Anatolian is the earliest dialect to have separated from PIE, due to its peculiar archaisms, and shows therefore a situation different from that looked for in this Grammar.

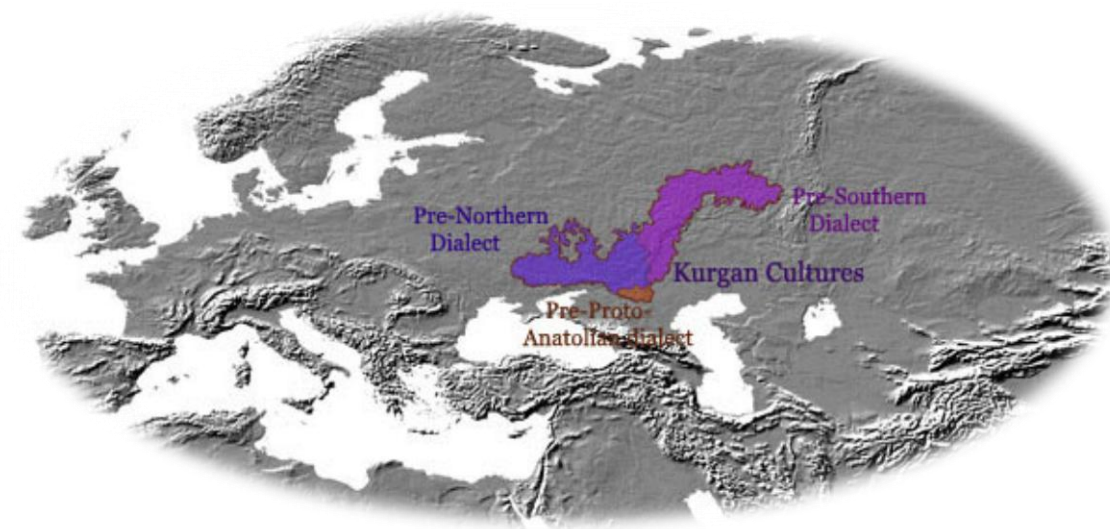


Figure 6. Early Kurgan cultures in ca. 4.000 B.C., showing hypothetical territory where IE II proto-dialects (i.e. pre-IE III and pre-Proto-Anatolian) could have developed.

3) The common immediate ancestor of the early IE proto-languages –more or less the same static PIE searched for since the start of Indo-European studies – is usually called *Late PIE*, also *Indo-European III* or **IE III**, or simply *Proto-Indo-European*. Its prehistoric community of speakers is generally identified with the Yamna or Pit Grave culture (cf. Ukr. *яма*, “pit”), in the Pontic Steppe. Proto-Anatolian speakers are arguably identified with the Maykop cultural community.

NOTE. The development of this theory of three linguistic stages can be traced back to the very origins of Indo-European studies, firstly as a diffused idea of a non-static language, and later widely accepted as a dynamic dialectal evolution, already in the 20th century, after the discovery of the Anatolian scripts.

1.3.3. Another division has to be made, so that the dialectal evolution is properly understood. Late PIE had at least two main dialects, the *Northern* (or IE IIIb) and the *Southern* (or IE IIIa) one. Terms like *Northwestern* or *European* can be found in academic writings referring to the Northern Dialect, but we will use them here to name only the northern dialects of Europe, thus generally excluding Tocharian.

Also, *Graeco-Aryan* is used to refer to the Southern Dialect of PIE. Indo-Iranian is used in this grammar to describe the southern dialectal grouping formed by Indo-Aryan, Iranian and Nuristani dialects, and not – as it is in other texts – to name the southern dialects of Asia as a whole. Thus, unclassified IE dialects like *Cimmerian*, *Scythian* or *Sarmatian* (usually deemed just Iranian dialects) are in this grammar simply some of many southern dialects spoken in Asia in Ancient times.

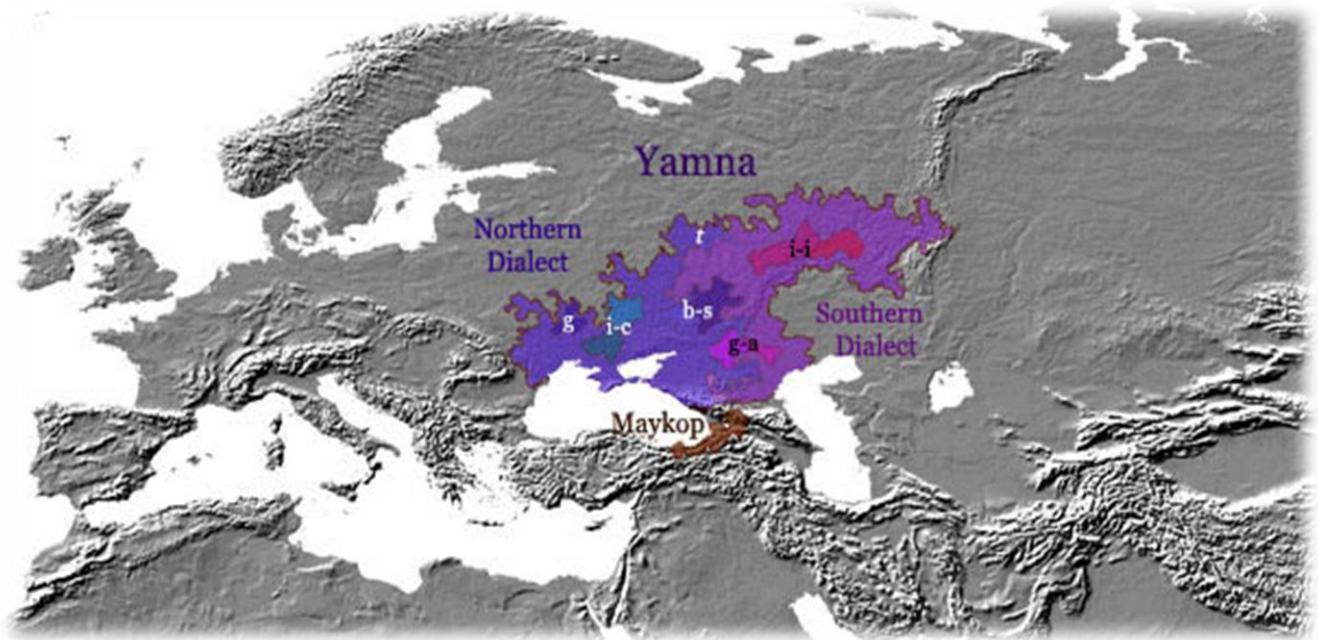


Figure 7. Yamna culture ca. 3000 B.C., probably the time when still a single Proto-Indo-European language was spoken. In two different colors, hypothetical locations of later Northern and Southern Dialects. Other hypothetical groupings are depicted according to their later linguistic and geographical development, i.e. **g**:Germanic, **i-c**:Italo-Celtic, **b-s**:Balto-Slavic, **t**:Tocharian, **g-a**:Graeco-Armenian, **i-i**:Indo-Iranian, among other death and unattested dialects which coexisted necessarily with them.

1.3.4. As far as we know, while speakers of southern dialects (like Proto-Greek, Proto-Indo-Iranian and probably Proto-Armenian) spread in different directions, some speakers of northern dialects remained still in loose contact in Europe, while others (like *Proto-Tocharians*) spread in Asia. Those northern Indo-European dialects of Europe were early Germanic, Celtic, Italic, and probably Balto-Slavic (usually considered transitional with IE IIIa) proto-dialects, as well as other not so well-known dialects like *Proto-Lusitanian*, *Proto-Sicel*, *Proto-Thracian* (maybe *Proto-Daco-Thracian*, for some within a wider *Proto-Graeco-Thracian* group), *pre-Proto-Albanian* (maybe *Proto-Illyrian*), etc.

NOTE. Languages like Venetic, Liburnian, Phrygian, Thracian, Macedonian, Illyrian, Messapic, Lusitanian, etc. are usually called ‘*fragmentary languages*’ (sometimes also ‘*ruinous languages*’), as they are languages we have only fragments from.

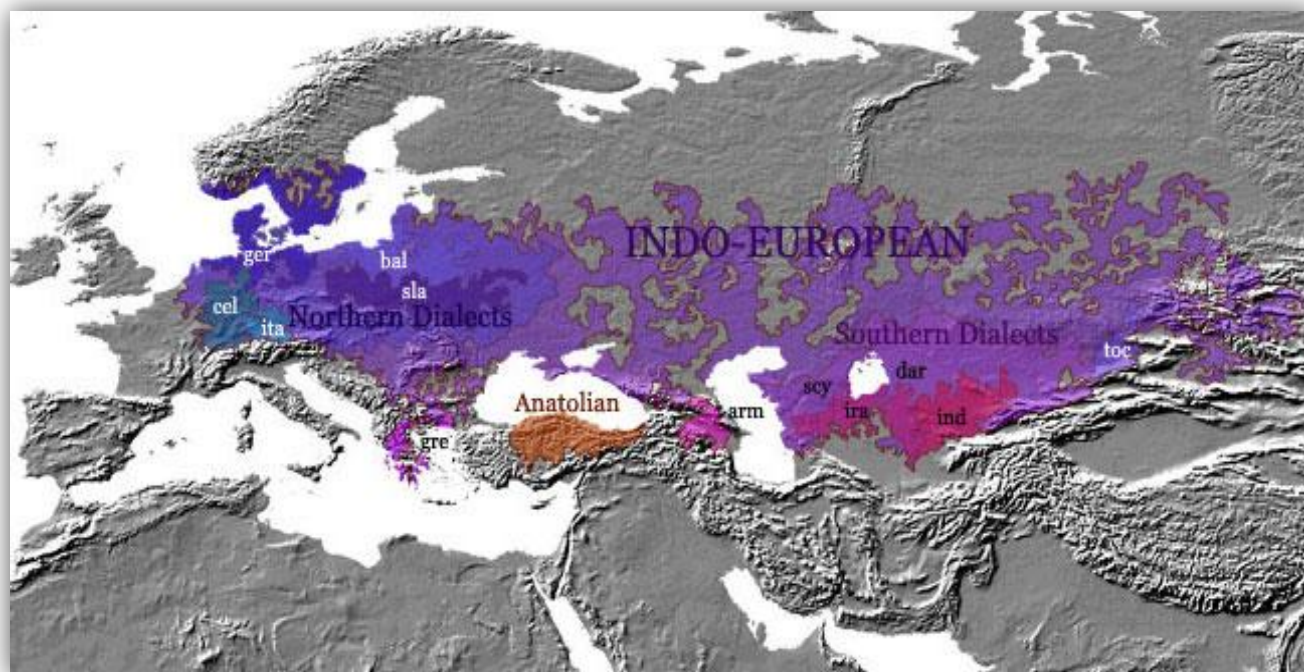


Figure 8. Spread of Late Proto-Indo-European ca. 2000 B.C. At that time, only the European northern dialects remained in contact, allowing the spread of linguistic developments, while the others evolved more or less independently. Anatolian dialects as Hittite and Luwian attested since 1900 B.C., and Proto-Greek Mycenaean dialect attested in 16th century B.C.

Other Indo-European dialects attested in Europe which remain unclassified are Paleo-Balkan languages like Thracian, Dacian, Illyrian (some group them into *Graeco-Thracian*, *Daco-Thracian* or *Thraco-Illyrian*), Paionian, Venetic, Messapian, Liburnian, Phrygian and maybe also Ancient Macedonian and Ligurian.

The European dialects have some common features, as a general reduction of the 8-case paradigm into a five- or six-case noun inflection system, the *-r* endings of the middle voice, as well as the lack of satemization. The southern dialects, in turn, show a generalized Augment in *é-*, a general Aorist formation and an 8-case system (also apparently in Proto-Greek).

NOTE. Balto-Slavic (and, to some extent, Italic) dialects, either because of their original situation within the PIE dialectal territories, or because they remained in contact with Southern Indo-European dialects after the first PIE split (e.g. through the Scythian or Iranian expansions) present features usually identified with Indo-Iranian, as an 8-case noun declension and phonetic satemization, and at the same time morphological features common to Germanic and Celtic dialects, as the verbal system.

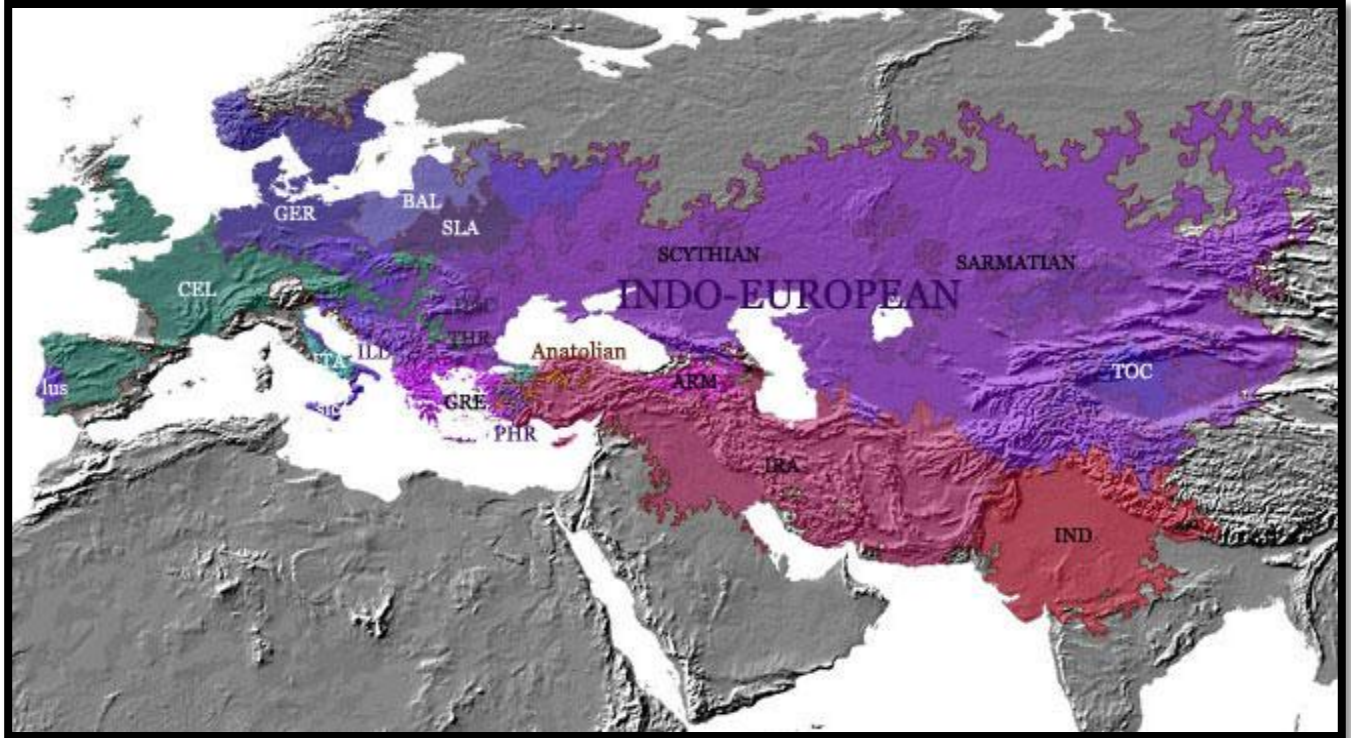


Figure 9. Eurasia ca. 500 B.C. The spread of Scythians allow renewed linguistic contact between Indo-Iranian and Slavic languages, whilst Armenian- and Greek-speaking communities are again in close contact with southern IE dialects, due to the Persian expansion. Italo-Celtic speakers spread and drive other northern dialects (as Lusitanian or Sicul) further south. Later Anatolian dialects, as Lycian, Lydian and Carian, are still spoken.

NOTE. The term *Indo-European* itself now current in English literature, was coined in 1813 by the British scholar Sir Thomas Young, although at that time, there was no consensus as to the naming of the recently discovered language family. Among the names suggested were *indo-germanique* (C. Malte-Brun, 1810), *Indoeuropean* (Th. Young, 1813), *japetisk* (Rasmus C. Rask, 1815), *indisch-teutsch* (F. Schmitthenner, 1826), *sanskritisch* (Wilhelm von Humboldt, 1827), *indokeltisch* (A. F. Pott, 1840), *arioeuropeo* (G. I. Ascoli, 1854), *Aryan* (F. M. Müller, 1861), *aryaque* (H. Chavée, 1867).

In English, *Indo-German* was used by J. C. Prichard in 1826 although he preferred *Indo-European*. In French, use of *indo-européen* was established by A. Pictet (1836). In German literature, *Indo-Europäisch* was used by Franz Bopp since 1835, while the term *Indo-Germanisch* had already been introduced by Julius von Klapproth in 1823, intending to include the northernmost and the southernmost of the family's branches, as it were as an abbreviation of the full listing of involved languages that had been common in earlier literature, opening the doors to ensuing fruitless discussions whether it should not be *Indo-Celtic*, or even *Tocharo-Celtic*.

1.4. THE PROTO-INDO-EUROPEAN *URHEIMAT* OR ‘HOMELAND’

1.4.1. The search for the *Urheimat* or ‘Homeland’ of the prehistoric community who spoke Early Proto-Indo-European has developed as an archaeological quest along with the linguistic research looking for the reconstruction of that proto-language.

1.4.2. The **Kurgan hypothesis** was introduced by Marija Gimbutas in 1956 in order to combine archaeology with linguistics in locating the origins of the Proto-Indo-Europeans. She named the set of cultures in question “Kurgan” after their distinctive burial mounds and traced their diffusion into Europe. According to her hypothesis (1970: “*Proto-Indoeuropean culture: the Kurgan culture during the 5th to the 3rd Millennium*



Figure 10. Photo of a Kurgan from the *Archaeology Magazine*.

B.C.”, *Indo-European and Indo-Europeans*, Philadelphia, 155-198), PIE speakers were probably located in the Pontic Steppe. This location combines the expansion of the Northern and Southern dialects, whilst agreeing at the same time with the four successive stages of the Kurgan cultures.

1.4.3. Gimbutas' original suggestion identifies four successive stages of the Kurgan culture and three successive “waves” of expansion.

1. **Kurgan I**, Dnieper/Volga region, earlier half of the 4th millennium BC. Apparently evolving from cultures of the Volga basin, subgroups include the Samara and Seroglazovo cultures.

2. **Kurgan II–III**, latter half of the 4th millennium BC. Includes the Sredny Stog culture and the Maykop culture of the northern Caucasus. Stone circles, early two-wheeled chariots, anthropomorphic stone stelae of deities.

3. **Kurgan IV** or Pit Grave culture, first half of the 3rd millennium BC, encompassing the entire steppe region from the Ural to Romania.

➤ **Wave 1**, predating Kurgan I, expansion from the lower Volga to the Dnieper, leading to coexistence of Kurgan I and the Cucuteni culture. Repercussions of the migrations extend as far as the Balkans and along the Danube to the Vinča and Lengyel cultures in Hungary.

➤ **Wave 2**, mid 4th millennium BC, originating in the Maykop culture and resulting in advances of “**kurganized**” hybrid cultures into northern Europe around 3000 BC – Globular Amphora culture, Baden culture, and ultimately Corded Ware culture. In the belief of Gimbutas, this corresponds to the first intrusion of IE dialects into western and northern Europe.

➤ **Wave 3**, 3000–2800 BC, expansion of the Pit Grave culture beyond the steppes, with the appearance of the characteristic pit graves as far as the areas of modern Romania, Bulgaria and eastern Hungary.

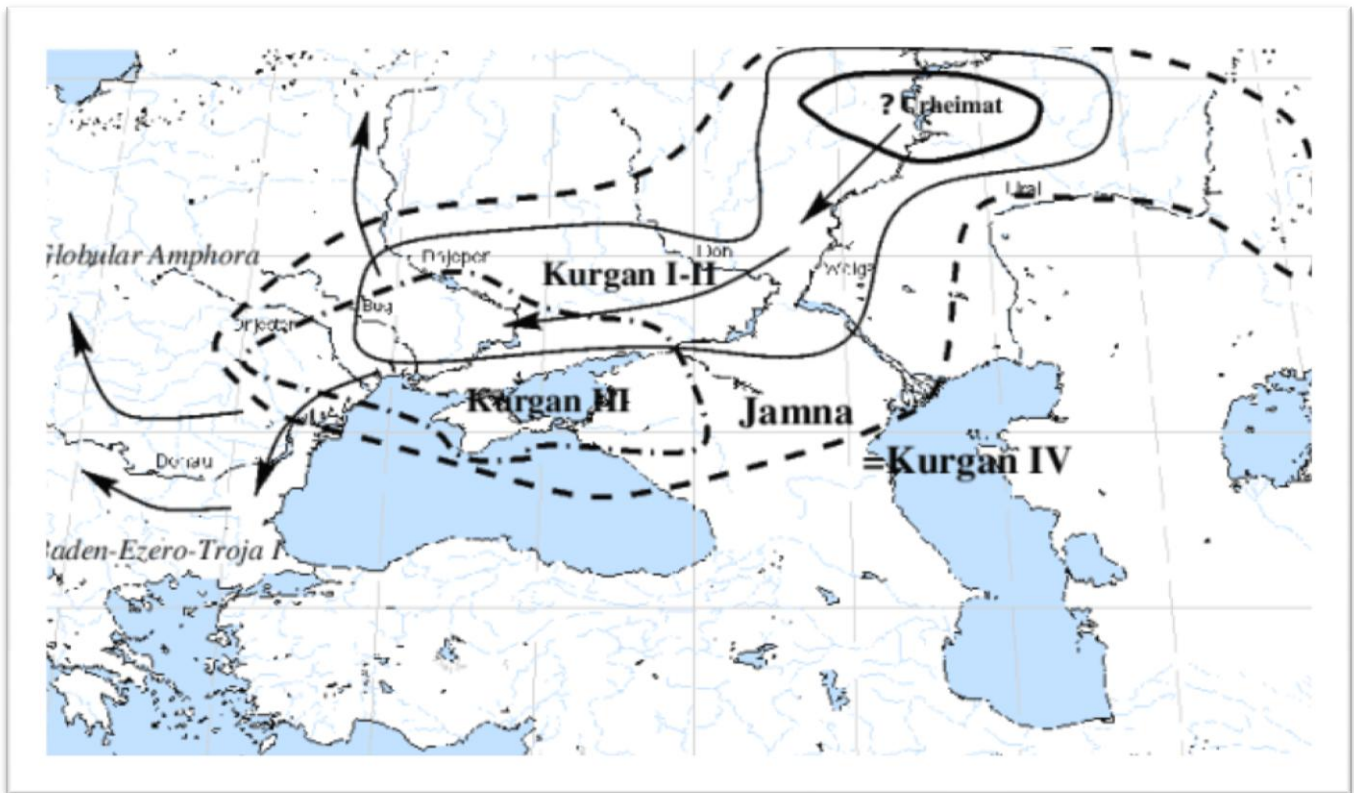


Figure 11. Hypothetical Homeland or Urheimat of the first PIE speakers, from 4,500 BC onwards. The Yamnaya or Jamna (Pit Grave) culture lasted from ca. 3,600 till 2,200. In this time the first wagons appeared. People were buried with their legs flexed, a position which remained typical for the Indo-Europeans for a long time. The burials were covered with a mound, a kurgan. During this period, from 3,600 till 3,000 IE II split up into IE III and Anatolian. From ca. 3,000 B.C on, IE III dialects began to differentiate and spread by 2,500 west- and southward (European Dialects, Armenian) and eastward (Indo-Iranian, Tocharian). By 2,000 the dialectal breach is complete.

1.4.3. The European or northwestern dialects, i.e. Celtic, Germanic, Italic, Baltic and Slavic, have developed together in the European Subcontinent but, because of the different migrations and settlements, they have undergone independent linguistic changes. Their original common location is usually traced back to some place to the East of the Rhine, to the North of the Alps and the Carpathian Mountains, to the South of Scandinavia and to the East of the Eastern European Lowlands or Russian Plain, not beyond Moscow.

This linguistic theory is usually mixed with archaeological findings:

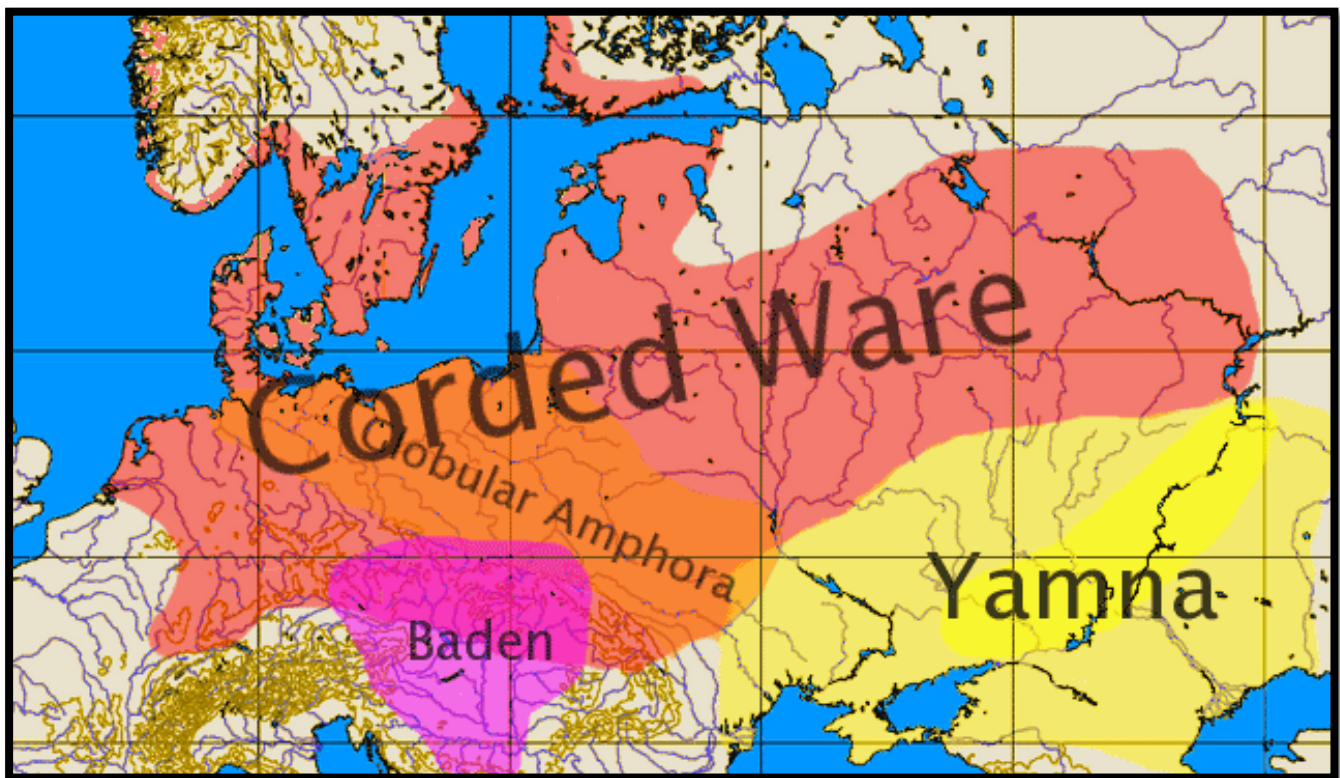


Figure 15. ca 2.000 B.C. The Corded Ware complex of cultures traditionally represents for many scholars the arrival of the first speakers of Northern Dialects in central Europe, coming from the Yamna culture. The complex dates from about 3.000-2.000. The Globular Amphorae culture may be slightly earlier, but the relation between these two cultures is unclear. Denmark and southern Scandinavia are supposed to have been the Germanic homeland, while present-day West Germany would have been the Celtic (and possibly Italic) homeland; the east zone, then, corresponds to the Balto-Slavic homeland. Their proto-languages certainly developed closely (if they weren't the same) until 2.000 B.C.

Kurgan Hypothesis & Proto-Indo-European reconstruction

ARCHAEOLOGY (Kurgan Hypothesis)	LINGUISTICS (Three-Stage Theory)
ca. 4500-4000. <i>Sredny Stog, Dnieper-Donets and Sarama cultures, domestication of the horse.</i>	Early PIE is spoken, probably somewhere in the Pontic-Caspian Steppe.
ca. 4000-3500. <i>The Yamna culture, the kurgan builders, emerges in the steppe, and the Maykop culture in northern Caucasus.</i>	Middle PIE or IE II split up in two different communities, the Proto-Anatolian and the Pre-IE III.
ca. 3500-3000. <i>The Yamna culture is at its peak, with stone idols, two-wheeled proto-chariots, animal husbandry, permanent settlements and hillforts, subsisting on agriculture and fishing, along rivers. Contact of the Yamna culture with late Neolithic Europe cultures results in kurganized Globular Amphora and Baden cultures. The Maykop culture shows the earliest evidence of the beginning Bronze Age, and bronze weapons and artifacts are introduced.</i>	Late Proto-Indo-European or IE III and Proto-Anatolian evolve in different communities. Anatolian is isolated south of the Caucasus, and have no more contacts with the linguistic innovations of IE III.
3000-2500. <i>The Yamna culture extends over the entire Pontic steppe. The Corded Ware culture extends from the Rhine to the Volga, corresponding to the latest phase of Indo-European unity. Different cultures disintegrate, still in loose contact, enabling the spread of technology.</i>	IE III disintegrates into various dialects corresponding to different cultures, at least a Southern and a Northern one. They remain still in contact, enabling the spread of phonetic (like the Satem isogloss) and morphological innovations, as well as early loan words.
2500-2000. <i>The Bronze Age reaches Central Europe with the Beaker culture of Northern Indo-Europeans. Indo-Iranians settle north of the Caspian in the Sintashta-Petrovka and later the Andronovo culture.</i>	The breakup of the southern IE dialects is complete. Proto-Greek spoken in the Balkans and a distinct Proto-Indo-Iranian dialect. Some northern dialects develop in Northern Europe, still in loose contact.
2000-1500. <i>The chariot is invented, leading to the split and rapid spread of Iranians and other peoples from the Andronovo culture and the Bactria-Margiana Complex over much of Central Asia, Northern India, Iran and Eastern Anatolia. Greek Darg Ages and flourishing of the Hittite Empire. Pre-Celtics Unetice culture has an active metal industry.</i>	Indo-Iranian splits up in two main dialects, Indo-Aryan and Iranian. European proto-dialects like Germanic, Celtic, Italic, Baltic and Slavic differentiate from each other. A Proto-Greek dialect, Mycenaean, is already written in Linear B script. Anatolian languages like Hittite and Luwian are also written.
1500-1000. <i>The Nordic Bronze Age sees the rise of the Germanic Urnfield and the Celtic Hallstatt cultures in Central Europe, introducing the Iron Age. Italic peoples move to the Italian Peninsula. Rigveda is composed. The Hittite Kingdoms and the Mycenaean civilization decline.</i>	Germanic, Celtic, Italic, Baltic and Slavic are already different proto-languages , developing in turn different dialects. Iranian and other related southern dialects expand through military conquest, and Indo-Aryan spreads in the form of its sacred language, Sanskrit.
1000-500. <i>Northern Europe enters the Pre-Roman Iron Age. Early Indo-European Kingdoms and Empires in Eurasia. In Europe, Classical Antiquity begins with the flourishing of the Greek peoples. Foundation of Rome.</i>	Celtic dialects spread over Europe. Osco-Umbrian and Latin-Faliscan attested in the Italian Peninsula. Greek and Old Italic alphabets appear. Late Anatolian dialects. Cimmerian, Scythian and Sarmatian in Asia, Paleo-Balkan languages in the Balkans.

1.5. OTHER LINGUISTIC AND ARCHAEOLOGICAL THEORIES

1.5.1. A common development of new theories about Indo-European has been to revise the Three-Stage assumption. It is actually not something new, but only the come back to more traditional views, by reinterpreting the new findings of the Hittite scripts, trying to insert the Anatolian features into the old, static PIE concept.

1.5.2. The most known new alternative theory concerning PIE is the **Glottalic theory**. It assumes that Proto-Indo-European was pronounced more or less like Armenian, i.e. instead of PIE **p**, **b**, **bh**, the pronunciation would have been ***p'**, ***p**, ***b**, and the same with the other two voiceless-voiced-voiced aspirated series of consonants. The Indo-European *Urheimat* would have been then located in the surroundings of Anatolia, especially near Lake Urmia, in northern Iran, near present-day Armenia and Azerbaijan, hence the archaism of Anatolian dialects and the glottalics still found in Armenian.

NOTE. Such linguistic findings are supported by Th. Gamkredlize-V. Ivanov (1990: "*The early history of Indo-European languages*", *Scientiphic American*, where early Indo-European vocabulary deemed "of southern regions" is examined, and similarities with Semitic and Kartvelian languages are also brought to light. Also, the mainly archaeological findings of Colin Renfrew (1989: *The puzzle of Indoeuropean origins*, Cambridge-New York), supported by the archaism of Anatolian dialects, may indicate a possible origin of Early PIE speakers in Anatolia, which, after Renfrew's model, would have then migrated into southern Europe.

1.5.3. Other alternative theories concerning Proto-Indo-European are as follows:

I. The **European Homeland thesis** maintains that the common origin of the Indo-European languages lies in Europe. These thesis have usually a nationalistic flavour, more or less driven by Archeological or Linguistic theories.

NOTE. It has been traditionally located in 1) **Lithuania** and the surrounding areas, by R.G. Latham (1851) and Th. Poesche (1878: *Die Arier. Ein Beitrag zur historischen Anthropologie*, Jena); 2) **Scandinavia**, by K. Penka (1883: *Origines ariacae*, Viena); 3) **Central Europe**, by G. Kossinna (1902: "*Die Indogermanische Frage archäologisch beantwortet*", *Zeitschrift für Ethnologie*, 34, pp. 161-222), P. Giles (1922: *The Aryans*, New York), and by linguist/archaeologist G. Childe (1926: *The Aryans. A Study of Indo-European Origins*, London).

a. The *Old European* or **Alteuropäisch Theory** compares some old European vocabulary (especially river names), which would be older than the spread of Late PIE through Europe. It points out the possibility of an older, pre-IE III spread of IE, either of IE II or I or maybe their ancestor.

b. This is, in turn, related with the theories of a **Neolithic revolution** causing the peacefully spreading of an older Indo-European language into Europe from Asia Minor from around 7000 BC, with the advance of farming. Accordingly, more or less all of Neolithic Europe would have been Indo-European speaking, and the Northern IE III Dialects would have replaced older IE dialects, from IE II or Early Proto-Indo-European.

c. There is also a ***Paleolithic Continuity Theory***, which derives Proto-Indo-European from the European Paleolithic cultures, with some research papers available online at the researchers' website, <http://www.continuitas.com/>.

NOTE. Such Paleolithic Continuity could in turn be connected with Frederik Kortlandt's Indo-Uralic and Altaic studies (<http://kortlandt.nl/publications/>) – although they could also be inserted in Gimbutas' early framework.

II. Another hypothesis, contrary to the European ones, also mainly driven today by a nationalistic view, traces back the origin of PIE to Vedic Sanskrit, postulating that it is very *pure*, and that the origin can thus be traced back to the Indus valley civilization of ca. 3000 BC.

NOTE. Such Pan-Sanskritism was common among early Indo-Europeanists, as Schlegel, Young, A. Pictet (1877: *Les origines indoeuropéens*, Paris) or Schmidt (who preferred Babylonia), but are now mainly supported by those who consider Sanskrit almost equal to Late Proto-Indo-European. For more on this, see S. Misra (1992: *The Aryan Problem: A Linguistic Approach*, Delhi), Elst's *Update on the Aryan Invasion Debate* (1999), followed up by S.G. Talageri's *The Rigveda: A Historical Analysis* (2000), both part of "Indigenous Indo-Aryan" viewpoint by N. Kazanas, the so-called "Out of India" theory, with a framework dating back to the times of the Indus Valley Civilization, deeming PIE simply a hypothesis (<http://www.omilosmeleton.gr/english/documents/SPIE.pdf>).

III. Finally, the Black Sea deluge theory dates the origins of the IE dialects expansion in the genesis of the Sea of Azov, ca. 5600 BC, which in turn would be related to the Bible Noah's flood, as it would have remained in oral tales until its writing down in the Hebrew *Tanakh*. This date is generally considered as rather early for the PIE spread.

NOTE. W. Ryan and W. Pitman published evidence that a massive flood through the Bosphorus occurred about 5600 BC, when the rising Mediterranean spilled over a rocky sill at the Bosphorus. The event flooded 155,000 km² of land and significantly expanded the Black Sea shoreline to the north and west. This has been connected with the fact that some Early Modern scholars based on Genesis 10:5 have assumed that the 'Japhetite' languages (instead of the 'Semitic' ones) are rather the direct descendants of the Adamic language, having separated before the confusion of tongues, by which also Hebrew was affected. That was claimed by *Blessed Anne Catherine Emmerich* (18th c.), who stated in her private revelations that most direct descendants of the *Adamic language* were Bactrian, Zend and Indian languages, related to her Low German dialect. It is claimed that Emmerich identified this way *Adamic language* as Early PIE.

1.6. RELATIONSHIP TO OTHER LANGUAGES

1.6.1. Many higher-level relationships between PIE and other language families have been proposed. But these speculative connections are highly controversial. Perhaps the most widely accepted proposal is of an Indo-Uralic family, encompassing PIE and Proto-Uralic. The evidence usually cited in favor of this is the proximity of the proposed Urheimaten of the two proto-languages, the typological similarity between the two languages, and a number of apparent shared morphemes.

NOTE. Other proposals, further back in time (and correspondingly less accepted), model PIE as a branch of Indo-Uralic with a Caucasian substratum; link PIE and Uralic with Altaic and certain other families in Asia, such as Korean, Japanese, Chukotko-Kamchatkan and Eskimo-Aleut (representative proposals are Nostratic and Joseph Greenberg's Eurasiatic); or link some or all of these to Afro-Asiatic, Dravidian, etc., and ultimately to a single Proto-World family (nowadays mostly associated with Merritt Ruhlen). Various proposals, with varying levels of skepticism, also exist that join some subset of the putative Eurasiatic language families and/or some of the Caucasian language families, such as Uralo-Siberian, Ural-Altaic (once widely accepted but now largely discredited), Proto-Pontic, and so on.

1.6.2. **Indo-Uralic** is a hypothetical language family consisting of Indo-European and Uralic (i.e. Finno-Ugric and Samoyedic). Most linguists still consider this theory speculative and its evidence insufficient to conclusively prove genetic affiliation.

1.6.3. Dutch linguist Frederik Kortlandt supports a model of Indo-Uralic in which the original Indo-Uralic speakers lived north of the Caspian Sea, and the Proto-Indo-European speakers began as a group that branched off westward from there to come into geographic proximity with the Northwest Caucasian languages, absorbing a Northwest Caucasian lexical blending before moving farther westward to a region north of the Black Sea where their language settled into canonical Proto-Indo-European.

1.6.4. The most common arguments in favour of a relationship between Indo-European and Uralic are based on seemingly common elements of morphology, such as the pronominal roots (**m-* for first person; **t-* for second person; **i-* for third person), case markings (accusative **-m*; ablative/partitive **-ta*), interrogative/relative pronouns (**kw-* 'who?, which?'; **j-* 'who, which' to signal relative clauses) and a common SOV word order. Other, less obvious correspondences are suggested, such as the Indo-European plural marker **-es* (or **-s* in the accusative plural **-m_s-s*) and its Uralic counterpart **-t*. This same word-final assibilation of **-t* to **-s* may also be present in Indo-European second-person singular **-s* in comparison with Uralic second-person singular **-t*. Compare, within Indo-European itself, **-s* second-person singular injunctive, **-si* second-person singular present indicative, **-tHa* second-person singular perfect, **-te* second-person plural present indicative, **tu* 'you' (singular) nominative, **tei* 'to you' (singular) enclitic pronoun. These forms suggest that the underlying second-person marker in Indo-European may be **t* and that the **u* found in forms such as **tu* was originally an affixal particle.

A second type of evidence advanced in favor of an Indo-Uralic family is lexical. Numerous words in Indo-European and Uralic resemble each other. The problem is to weed out words due to borrowing. Uralic languages have been in contact with a succession of Indo-European languages for millenia. As a result, many words have been borrowed between them, most often from Indo-European languages into Uralic ones.

Proto-Indo-European and Proto-Uralic side by side

Meaning	Proto-Indo-European	Proto-Uralic
<i>I, me</i>	* me 'me' [acc], * mene 'my' [gen]	* <i>mVnV 'I'</i>
<i>you</i> (sg)	* tu [nom], * twe [obj], * tewe 'your' [gen]	* <i>tun</i>
[demonstrative]	* so 'this, he/she' [animate nom]	* <i>ša</i> [3ps]
<i>who?</i> [animate interrogative pronoun]	* k^wi- 'who?, what?' * k^wo- 'who?, what?'	* <i>ken</i> 'who?' * <i>ku-</i> 'who?'
[relative pronoun]	* jo-	* <i>-ja</i> [nomen agentis]
[definite accusative]	* -m	* <i>-m</i>
[ablative/partitive]	* -od	* <i>-ta</i>
[dual]	* -h₁	* <i>-k</i>
[Nom./Acc. plural]	* -es [nom.pl], * -m̥-s [acc.pl]	* <i>-k</i>
[Obl. plural]	* -i [pronominal plural] (as in * we-i- 'we', * to-i- 'those')	* <i>-i</i>
[1ps]	* -m [1ps active]	* <i>-m</i>
[2ps]	* -s [2ps active]	* <i>-t</i>
[stative]	* -s- [aorist], * -es- [stative substantive], * -t [stative substantive]	* <i>-ta</i>
[negative]	* nei * ne	* <i>ei-</i> [negative verb]
<i>to give</i>	* deh₃-	* <i>toHi-</i>
<i>to moisten,</i> <i>water</i>	* wed- 'to wet', * wód_ṛ 'water'	* <i>weti</i> 'water'
<i>to assign,</i> <i>name</i>	nem- 'to assign, to allot', * h₁nom_ṇ 'name'	* <i>nimi</i> 'name'

1.7. INDO-EUROPEAN DIALECTS OF EUROPE

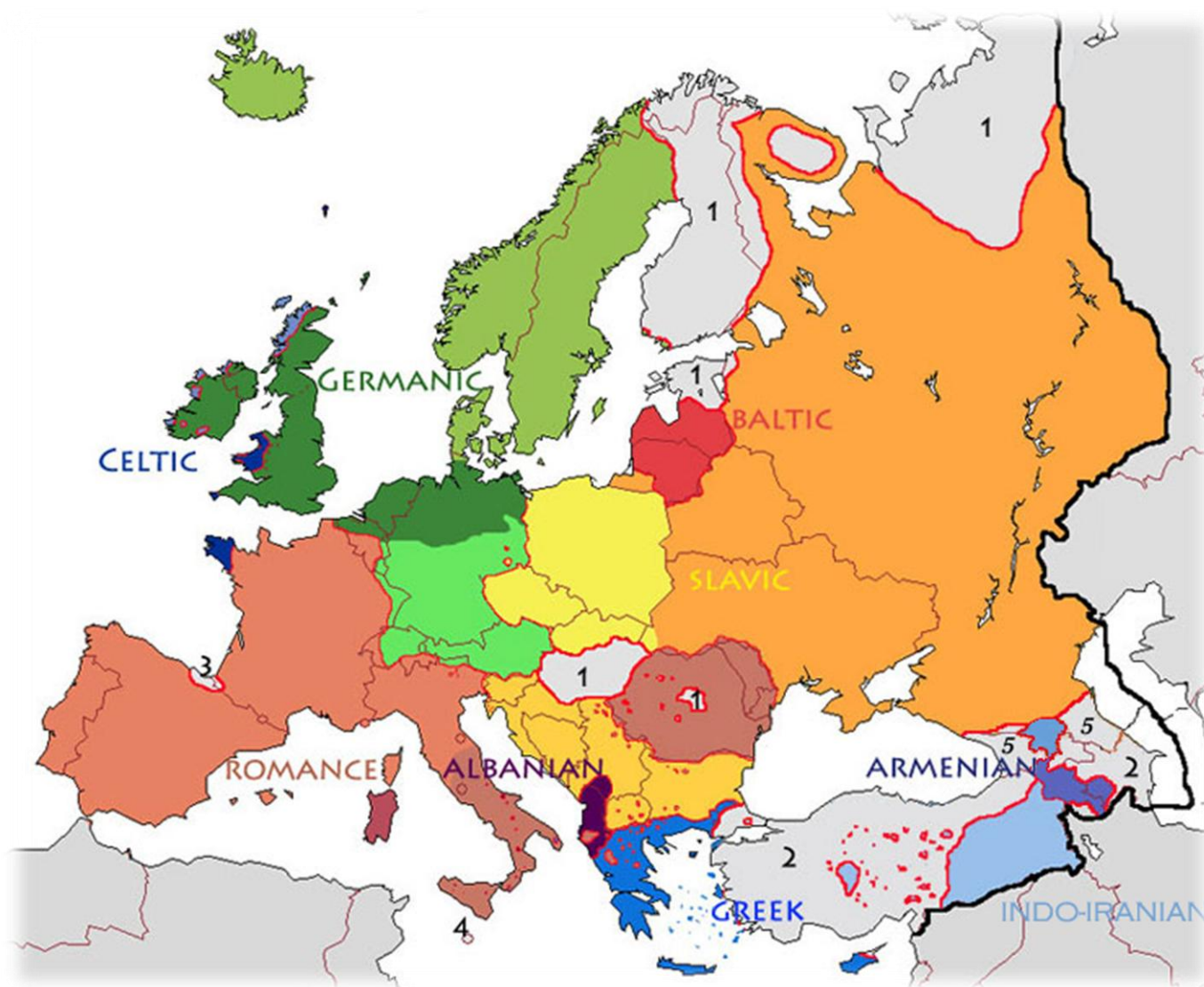


Figure 16. European languages. The **black line** divides the zones traditionally (or politically) considered inside the European subcontinent. Northern dialects are all but Greek and Kurdish (Iranian); Armenian is usually considered a Graeco-Aryan dialect, while Albanian is usually classified as a Northern one. Numbered inside the map, non-Indo-European languages: 1) Uralic languages; 2) Turkic languages; 3) Basque; 4) Maltese; 5) Caucasian languages.

SCHLEICHER'S FABLE: FROM PROTO-INDO-EUROPEAN TO MODERN ENGLISH

« *The Sheep and the Horses. A sheep that had no wool saw horses, one pulling a heavy wagon, one carrying a big load, and one carrying a man quickly. The sheep said to the horses: "My heart pains me, seeing a man driving horses". The horses said: "Listen, sheep, our hearts pain us when we see this: a man, the master, makes the wool of the sheep into a warm garment for himself. And the sheep has no wool". Having heard this, the sheep fled into the plain.* »

IE III, ca. 3000 BC: H₃ouis h₁ékʷo(s)es-qe. H₃ouis, k^wesjo ʷHneh₂ ne h₁est, h₁ékʷoms spekét, h₁óinom g^wh₃um wóghom wéghontm, h₁óinom-k^we mégeh₂m b^hórom, h₁óinom-k^we d^hHgh^mónm h₁oh₁ku b^hérontm. H₃owis nu h₁ékʷob^hios ʷeuk^wét: kerd h₂ég^hnutoi h₁moí h₁ékʷoms h₂égontm wiHrom wídn^htei. H₁ékʷo(s)es tu ʷeuk^wónt: Klúd^hi, h₃ówi! kerd h₂ég^hnutoi nsméi wídn^htb^hios: H₂ner, pótis, h₃oujom-r ʷHneh₂m suéb^hi g^whermóm ʷéstrom k^wrneúti. Neg^hi h₃oujom ʷHneh₂ h₁ésti. Tod kékluos h₃ouis h₂égrom b^hugét.

IE IIIb, ca. 2.000 BC (as MIE, with Latin script): Ówis ékwōs-qe. Ówis, qésio wġnā ne est, ékwoms spekét, óinom (ghe) crum wóghom wéghontm, óinom-qe mégām bhórom, óinom-qe dhghmónm óku bhérontm. Ówis nu ékwobh(i)os wewqét: krd ághnutoi moí, ékwoms ágontm wírom wídn^htei. Ékwōs tu wewqónt: Klúd^hi, ówi! krd ághnutoi nsméi wídn^htbh(i)os: anér, pótis, ówjom-r wġnām séb^hi chermóm wéstrom q^rneúti. Ówjom-qe wġnā ne ésti. Tod kékluwos ówis ágrom bhugét.

IE IIIa, ca. 1.500 BC (Proto-Indo-Iranian dialect): Avis ak'vasas-ka. Avis, jasmin varnā na āst, dadark'a ak'vans, tam, garum vāgham vaghantam, tam, magham bhāram, tam manum āku bharantam. Avis ak'vabhjas avavakat; k'ard aghnutai mai vidanti manum ak'vans ag'antam. Ak'vāsas avavakant: k'rudhi avai, kard aghnutai vividvant-svas: manus patis varnām avisāns karnauti svabhjam gharmam vastram avibhjas-ka varnā na asti. Tat k'uk'ruvants avis ag'ram abhugat.

Proto-Italic, ca. 1.000 BC	Proto-Germanic, ca. 500 BC	Proto-Balto-Slavic, ca. 1 AD
Ouis ekuoi-kue	Awiz ehwaz-uh	Avis asvas(-ke)
ouis, kuesio ulana ne est,	awiz, hwesja wulno ne ist,	avis, kesjo vŭlna ne est,
speket ekuos,	spehet ehwanz,	spek'et asvās,
oinum brum uogum ueguntum,	ainan krun wagan wegantun,	inam gŭrō vezam vezantŭ,
oinum-kue megam forum,	ainan-uh mekon boran,	inam(-ke) még'am bóram,
oinum-kue humonum oku ferontum.	ainan-uh gumonun ahu berontun.	inam(-ke) zemenam jasu berantŭ.
Ouis nu ekuobus uokuet:	Awiz nu ehwamaz weuhet:	Avis nu asvamas vjauket:
kord áhnutor mihi uiduntei,	hert agnutai meke witantei,	sėrd aznutė me vėdėti,
ekuos aguntum uirum.	ehwans akantun weran.	asvās azantŭ viram.
Ekuos uokuont: Kludi, oui!	Ehwaz weuhant: hludi, awi!	Asvas vjaukant: sludi, awi!
kord ahnutor nos uiduntbos:	kert aknutai uns wituntmaz:	sėrd aznutė nas vėdŭntmas:
ner, potis, ulanam ouium	mannaz, fothiz, wulnon awjan	māg, pat', vŭlnam avjam
kurneuti sibi fermum uestrum.	hwurneuti sebi warman wistran.	karnjauti sebi g'armam vastram.
Ouium-kue ulana ne esti.	Awjan-uh wulno ne isti.	Avjam(-ke) vŭlna ne esti.
Tod kekluos ouis agrum fugit	That hehluwaz awiz akran buketh.	Tod sesluvas avis ak'ram buget.

1.7.1. NORTHERN INDO-EUROPEAN DIALECTS

A. GERMANIC

1.2.1. The **Germanic languages** form one of the branches of the Indo-European language family. The largest Germanic languages are English and German, with ca. 340 and some 120 million native speakers, respectively. Other significant languages include a number Low Germanic dialects (like Dutch) and the Scandinavian languages, Danish, Norwegian and Swedish.

Their common ancestor is **Proto-Germanic**, probably still spoken in the mid-1st millennium B.C. in Iron Age Northern Europe, since its separation from the Proto-Indo-European language around 2.000 BC. Germanic, and all its descendants, is characterized by a number of unique linguistic features, most famously the consonant change known as *Grimm's Law*. Early Germanic dialects enter history with the Germanic peoples who settled in northern Europe along the borders of the Roman Empire from the 2nd century.

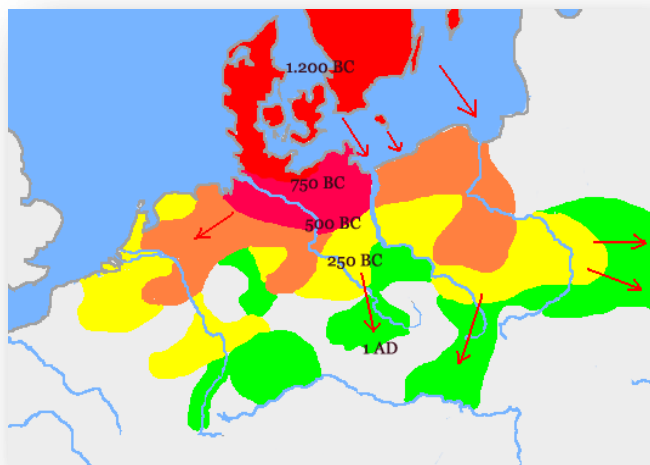


Figure 17. Expansion of Germanic tribes 1.200 B.C. – 1 A.D.

NOTE. **Grimm's law** (also known as the First Germanic Sound Shift) is a set of statements describing the inherited Proto-Indo-European stops as they developed in Proto-Germanic some time in the 1st millennium BC. It establishes a set of regular correspondences between early Germanic stops and fricatives and the stop consonants of certain other Indo-European languages (Grimm used mostly Latin and Greek for illustration). As it is presently formulated, Grimm's Law consists of three parts, which must be thought of as three consecutive phases in the sense of a chain shift:

- a. Proto-Indo-European voiceless stops change into voiceless fricatives.
- b. Proto-Indo-European voiced stops become voiceless.
- c. Proto-Indo-European voiced aspirated stops lose their aspiration and change into plain voiced stops.

The 'sound law' was discovered by Friedrich von Schlegel in 1806 and Rasmus Christian Rask in 1818, and later elaborated (i.e. extended to include standard German) in 1822 by Jacob Grimm in his book *Deutsche Grammatik*.

The earliest evidence of the Germanic branch is recorded from names in the 1st century by Tacitus, and in a single instance in the 2nd century BC, on the *Negau helmet*. From roughly the 2nd century AD, some speakers of early Germanic dialects developed the *Elder Futhark*. Early runic inscriptions are also largely limited to personal names, and difficult to interpret. The Gothic language was written in the

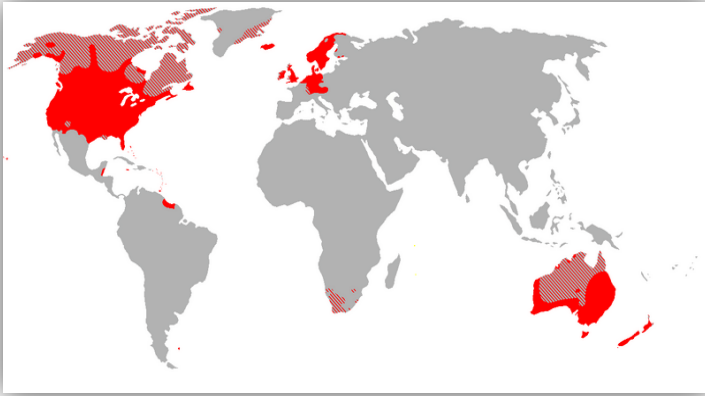


Figure 18. Spread of Germanic languages

Germanic languages use a variety of accent marks and extra letters, including umlaut, the ß (*Eszett*), IJ, Æ, Å, Ð, and Þ, from runes. Historic printed German is frequently set in blackletter typefaces.

Effects of the Grimm's Law in examples:

<i>IE-Gmc</i>	<i>Germanic (shifted) examples</i>	<i>Non-Germanic (unshifted)</i>
p → <i>f</i>	Eng. <i>foot</i> , Du. <i>voet</i> , Ger. <i>Fuß</i> , Goth. <i>fōtus</i> , Ice. <i>fótur</i> , Da. <i>fod</i> , Nor., Swe. <i>fot</i>	O.Gk. <i>πούς</i> (<i>pūs</i>), Lat. <i>pēs</i> , <i>pedis</i> , Skr. <i>pāda</i> , Russ. <i>pod</i> , Lith. <i>pėda</i>
t → <i>p</i>	Eng. <i>third</i> , O.H.G. <i>thritto</i> , Goth. <i>þridja</i> , Ice. <i>þriðji</i>	O.Gk. <i>τρίτος</i> (<i>tritōs</i>), Lat. <i>tertius</i> , Gae. <i>treas</i> , Skr. <i>treta</i> , Russ. <i>tretij</i> , Lith. <i>trys</i>
k → <i>h</i>	Eng. <i>hound</i> , Du. <i>hond</i> , Ger. <i>Hund</i> , Goth. <i>hunds</i> , Ice. <i>hundur</i> , Sca. <i>hund</i>	O.Gk. <i>κύων</i> (<i>kýōn</i>), Lat. <i>canis</i> , Gae. <i>cú</i> , Skr. <i>svan-</i> , Russ. <i>sobaka</i>
k^w → <i>hw</i>	Eng. <i>what</i> , Du. <i>wat</i> , Ger. <i>was</i> , Goth. <i>hva</i> , Da. <i>hvad</i> , Ice. <i>hvað</i>	Lat. <i>quod</i> , Gae. <i>ciod</i> , Skr. <i>ka-</i> , <i>kiṃ</i> , Russ. <i>ko-</i>
b → <i>p</i>	Eng. <i>peg</i>	Lat. <i>baculum</i>
d → <i>t</i>	Eng. <i>ten</i> , Du. <i>tien</i> , Goth. <i>taíhun</i> , Ice. <i>tíu</i> , Da., Nor.: <i>tí</i> , Swe. <i>tio</i>	Lat. <i>decem</i> , Gk. <i>δέκα</i> (<i>déka</i>), Gae. <i>deich</i> , Skr. <i>daśan</i> , Russ. <i>des'at'</i>
g → <i>k</i>	Eng. <i>cold</i> , Du. <i>koud</i> , Ger. <i>kalt</i>	Lat. <i>gelū</i>
g^w → <i>kw</i>	Eng. <i>quick</i> , Du. <i>kwiek</i> , Ger. <i>keck</i> , Goth. <i>qius</i> , O.N. <i>kvikr</i> , Swe. <i>kveck</i>	Lat. <i>vivus</i> , Gk. <i>βίος</i> (<i>bios</i>), Gae. <i>beò</i> , Lith. <i>gyvas</i>
b^h → <i>b</i>	Eng. <i>brother</i> , Du. <i>broeder</i> , Ger. <i>Bruder</i> , Goth. <i>broþar</i> , Sca. <i>broder</i>	Lat. <i>frāter</i> , O.Gk. <i>φρατήρ</i> (<i>phrātēr</i>), Skr. <i>bhrātā</i> , Lith. <i>brolis</i> , O.C.S. <i>bratru</i>
d^h → <i>d</i>	Eng. <i>door</i> , Fris. <i>doar</i> , Du. <i>deur</i> , Goth. <i>daúr</i> , Ice. <i>dyr</i> , Da., Nor. <i>dør</i> , Swe. <i>dörr</i>	O.Gk. <i>θύρα</i> (<i>thýra</i>), Skr. <i>dwār</i> , Russ. <i>dver'</i> , Lith. <i>durys</i>

g^h→g	Eng. <i>goose</i> , Fris. <i>goes</i> , Du. <i>gans</i> , Ger. <i>Gans</i> , Ice. <i>gæs</i> , Nor.,Swe. <i>gås</i>	Lat. <i>anser</i> < * <i>hanser</i> , O.Gk. <i>χῆν</i> (<i>khēn</i>), Skr. <i>hansa</i> , Russ. <i>gus'</i>
g^{wh}→gw	Eng. <i>wife</i> , O.E. <i>wif</i> , Du. <i>wijf</i> , O.H.G. <i>wib</i> , O.N. <i>vif</i> , Fae.: <i>vív</i> , Sca. <i>viv</i>	Tocharian B: <i>kwípe</i> , Tocharian A: <i>kip</i>

A known exception is that the voiceless stops did not become fricatives if they were preceded by IE **s**.

PIE	Germanic examples	Non-Germanic examples
sp	Eng. <i>spew</i> , Goth. <i>speiwan</i> , Du. <i>spuien</i> , Ger. <i>speien</i> , Swe. <i>spy</i>	Lat. <i>spuere</i>
st	Eng. <i>stand</i> , Du. <i>staan</i> , Ger. <i>stehen</i> , Ice. <i>standa</i> , Nor.,Swe. <i>stá</i>	Lat. <i>stāre</i> , Skr. <i>sta</i> Russian: <i>stat'</i>
sk	Eng. <i>short</i> , O.N. <i>skorta</i> , O.H.G. <i>scurz</i> , Du. <i>kort</i>	Skr. <i>krdhuh</i> , Lat. <i>curtus</i> , Lith. <i>skurdus</i>
sk^w	Eng. <i>scold</i> , O.N. <i>skáld</i> , Ice. <i>skáld</i> , Du. <i>Schelden</i>	Proto-Indo-European: <i>sk^wetlo</i>

Similarly, PIE **t** did not become a fricative if it was preceded by **p**, **k**, or **k^w**. This is sometimes treated separately under the Germanic *spirant law*:

Change	Germanic examples	Non-Germanic examples
pt→ft	Goth. <i>hliftus</i> “thief”	O.Gk. <i>κλέπτης</i> (<i>kleptēs</i>)
kt→ht	Eng. <i>eight</i> , Du. <i>acht</i> , Fris. <i>acht</i> , Ger. <i>acht</i> , Goth. <i>ahtáu</i> , Ice. <i>átta</i>	O.Gk. <i>οκτώ</i> (<i>oktō</i>), Lat. <i>octō</i> , Skr. <i>aṣṭan</i>
k^wt→h(w)t	Eng. <i>night</i> , O.H.G. <i>naht</i> , Du.,Ger. <i>nacht</i> , Goth. <i>nahts</i> , Ice. <i>nótt</i>	Gk. <i>nuks</i> , <i>nukt-</i> , Lat. <i>nox</i> , <i>noct-</i> , Skr. <i>naktam</i> , Russ. <i>noch</i> , Lith. <i>naktis</i>

The Germanic “sound laws”, allow one to define the expected sound correspondences between Germanic and the other branches of the family, as well as for Proto-Indo-European. For example, Germanic (word-initial) *b-* corresponds regularly to Italic *f-*, Greek *p^h-*, Indo-Aryan *b^h-*, Balto-Slavic and Celtic *b-*, etc., while Germanic **f-* corresponds to Latin, Greek, Sanskrit, Slavic and Baltic *p-* and to zero (no initial consonant) in Celtic. The former set goes back to PIE [**b^h**] (reflected in Sanskrit and modified in various ways elsewhere), and the latter set to an original PIE [**p**] – shifted in Germanic, lost in Celtic, but preserved in the other groups mentioned here.

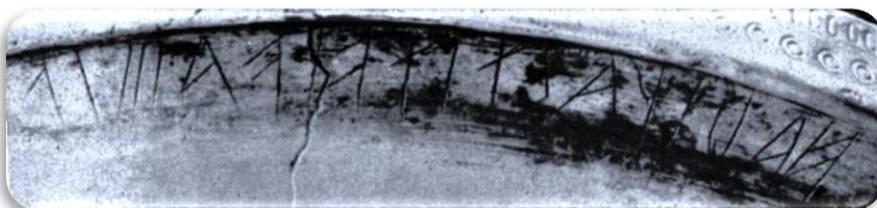


Figure 19 The Negau helmet (found in Negova, Slovenia), ca. 400 BC, contains the earliest attested Germanic inscription (read from right to left). It reads **harikastiteiva**\\ip, translated as “Harigast the priest”, and it was added probably ca. 200 BC.

B. ROMANCE

The **Romance languages**, a major branch of the Indo-European language family, comprise all languages that descended from Latin, the language of the Roman Empire. Romance languages have some 800 million native speakers worldwide, mainly in the Americas, Europe, and Africa, as well as in many smaller regions scattered through the world. The

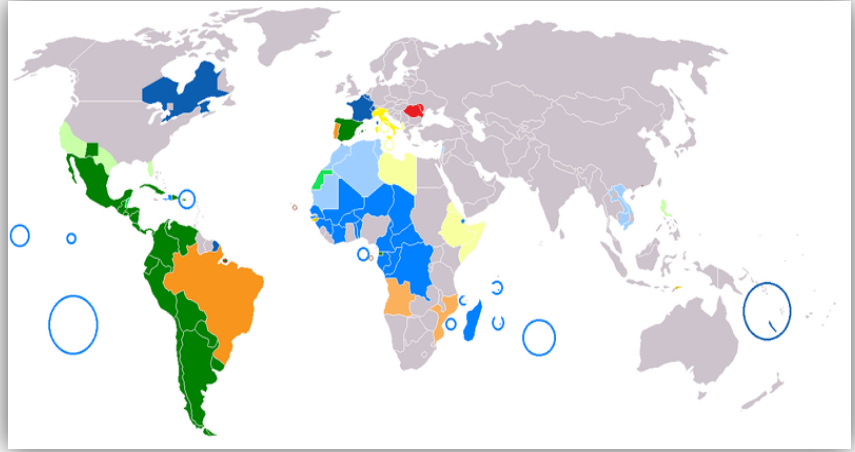


Figure 20. Regions where Romance languages are spoken, either as mother tongue or as second language.

largest languages are Spanish and Portuguese, with about 400 and 200 million mother tongue speakers respectively, most of them outside Europe. Within Europe, French (with 80 million) and Italian (70 million) are the largest ones. All Romance languages descend from Vulgar Latin, the language of soldiers, settlers, and slaves of the Roman Empire, which was substantially different from the Classical Latin of the Roman *literati*. Between 200 BC and 100 AD, the expansion of the Empire, coupled with administrative and educational policies of Rome, made Vulgar Latin the dominant native language over a wide area spanning from the Iberian Peninsula to the Western coast of the Black Sea. During the Empire's decadence and after its collapse and fragmentation in the 5th century, Vulgar Latin evolved independently within each local area, and eventually diverged into dozens of distinct languages. The overseas empires established by Spain, Portugal and France after the 15th century then spread Romance to

the other continents — to such an extent that about 2/3 of all Romance speakers are now outside Europe.

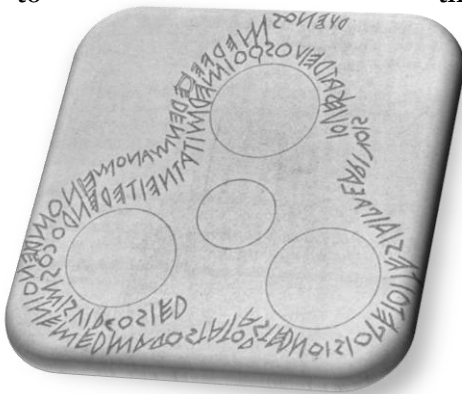


Figure 21. The 'Duenos' (Lat. 'buenus') Inscription in Old Latin, ca. 6th century BC.

Latin is usually classified, along with Faliscan, as another Italic dialect. The Italic speakers were not native to Italy, but migrated into the Italian Peninsula in the course of the 2nd millennium BC, and were apparently related to the Celtic tribes that roamed over a large part of Western Europe at the time. Archaeologically, the Apennine culture of inhumations enters the Italian Peninsula from ca. 1350 BC, east to west; the Iron Age reaches Italy from ca. 1100 BC, with the Villanovan culture (cremating), intruding north to

south. Before the Italic arrival, Italy was populated primarily by non-Indo-European groups (perhaps including the Etruscans). The first settlement on the Palatine hill dates to ca. 750 BC, settlements on the Quirinal to 720 BC, both related to the Founding of Rome.

The ancient Venetic language, as revealed by its inscriptions (including complete sentences), was also closely related to the Italic languages and is sometimes even classified as Italic. However, since it also shares similarities with other Western Indo-European branches (particularly Germanic), some linguists prefer to consider it an independent Indo-European language.

Italic is usually divided into:

❖ **Sabellic**, including:

- **Oscan**, spoken in south-central Italy.
- **Umbrian** group:
 - **Umbrian**
 - **Volscian**
 - **Aequian**
 - **Marsian**,
 - **South Picene**

❖ **Latino-Faliscan**, including:

- **Faliscan**, which was spoken in the area around *Falerii Veteres* (modern Civita Castellana) north of the city of Rome and possibly Sardinia
- **Latin**, which was spoken in west-central Italy. The Roman conquests eventually spread it throughout the Roman Empire and beyond.



Figure 22. Iron Age Italy. In central Italy, Italic languages. In southern and north-western Italy, other Indo-European languages. Venetic, Sicanian and Sicel were possibly also languages of the IE family.

Phonetic changes from PIE to Latin: **b^h** > f, **d^h** > f, **g^h** > h/f, **g^w** > v/g, **k^w** > kw (qu)/k (c), **p** > p/ qu.



Figure 23. The Masiliana tablet abecedarium, ca. 700 BC, read right to left: ABGDEVZHΘIKLMN[Ε]OPŚQRSTUXΦΨ.

The Italic languages are first attested in writing from Umbrian and Faliscan inscriptions dating to the 7th century BC. The alphabets used are based on the Old Italic alphabet, which is itself based on the Greek alphabet. The Italic languages themselves show minor influence from the Etruscan and somewhat more from the Ancient Greek languages.

Oscan had much in common with Latin, though there are also some differences, and many common word-groups in Latin were represented by different forms; as, Latin *uolo, uelle, uolui*, and other such forms from PIE **wel**, *will*, were represented by words derived from **gher**, *desire*, cf. Oscan *herest*, “he wants, desires” as opposed to Latin *uult* (id.). Latin *locus*, “place” was absent and represented by *slaagid*.

In phonology, Oscan also shows a different evolution, as Oscan 'p' instead of Latin 'qu' (cf. Osc. *pis*, Lat. *quis*); 'b' instead of Latin 'v'; medial 'f' in contrast to Latin 'b' or 'd' (cf. Osc. *mefiai*, Lat. *mediae*), etc.

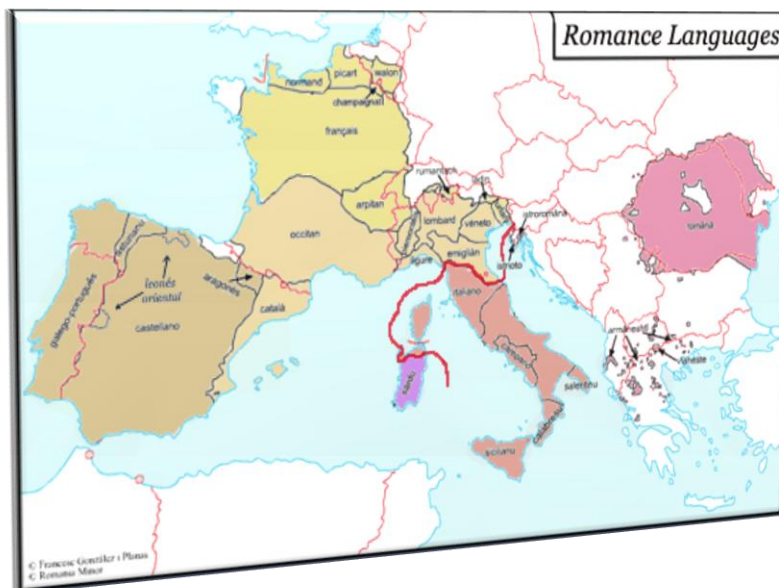
Up to 8 cases are found; apart from the 6 cases of Classic Latin (i.e. N-V-A-G-D-Ab), there was a Locative (cf. Lat. *proximae viciniae, domī, carthagini*, Osc. *aasai* ‘in ārā’ etc.) and an Instrumental (cf. Columna Rostrata Lat. *pugnandod, marid, naualid*, etc, Osc. *cadeis amnud*, ‘inimicitiae causae’, *preiuatud* ‘prīuātō’, etc.). About forms different from original Genitives and Datives, compare Genitive (Lapis Satricanus:) *popliosio valesiosio* (the type in -ī is also very old, Segomaros -ī), and Dative (Praeneste Fibula:) *numasioi*, (Lucius Cornelius Scipio Epitaph:) *quoiei*.

As Rome extended its political dominion over the whole of the Italian Peninsula, so too did Latin become dominant over the other Italic languages, which ceased to be spoken perhaps sometime in the 1st century AD.

Figure 25. Romance Languages Today. The Red line divides Western from Eastern (and Insular) Romance.



Figure 24. Forum inscription in Latin, written boustrophedon



C. SLAVIC

The **Slavic languages** (also called **Slavonic** languages), a group of closely related languages of the Slavic peoples and a subgroup of the Indo-European language family, have speakers in most of Eastern Europe, in much of the Balkans, in parts of Central Europe, and in the northern part of Asia. The largest languages are Russian and Polish, with 165 and some 47 million speakers, respectively. The oldest Slavic literary language was Old Church Slavonic, which later evolved into Church Slavonic.



Figure 26. Distribution of Slavic languages in Europe now and in the past (in stripes).

There is much debate whether pre-Proto-Slavic branched off directly from Proto-Indo-European, or whether it passed through a **Proto-Balto-Slavic** stage which split apart before 1000BC.



Figure 27. Historical distribution of the Slavic languages. The larger shaded area is the Prague-Penkov-Kolochin complex of cultures of the sixth to seventh centuries, likely corresponding to the spread of Slavic-speaking tribes of the time. The smaller shaded area indicates the core area of Slavic river names.

NOTE. For example the Freising monuments show a language which contains some phonetic and lexical elements peculiar to Slovenian dialects (e.g. rhotacism, the word *krilatec*).

In the second half of the ninth century, the dialect spoken north of Thessaloniki became the basis for the first written Slavic language, created by the brothers Cyril and Methodius who translated portions of the Bible and other church books. The language they recorded is known as Old Church Slavonic. Old Church Slavonic is not identical to Proto-Slavic, having been recorded at least two centuries after the breakup of Proto-Slavic, and it shows features that clearly distinguish it from Proto-Slavic. However, it is still reasonably close, and the mutual intelligibility between Old Church Slavonic and other Slavic dialects of those days was proved by Cyril's and Methodius' mission to Great Moravia and Pannonia. There, their early South Slavic dialect used for the translations was clearly understandable to the local population which spoke an early West Slavic dialect.

As part of the preparation for the mission, the Glagolitic alphabet was created in 862 and the most important prayers and liturgical books, including the Aprakos Evangeliar – a Gospel Book lectionary containing only feast-day and Sunday readings –, the Psalter, and Acts of the Apostles, were translated. The language and the alphabet were taught at the Great Moravian Academy (O.C.S. *Velkomoravské učilište*) and were used for government and religious documents and books. In 885, the use of the Old Church Slavonic in Great Moravia was prohibited by the Pope in favour of Latin. Students of the two apostles, who were expelled from Great Moravia in 886, brought the Glagolitic alphabet and the Old

The original homeland of the speakers of Proto-Slavic remains controversial too. The most ancient recognizably Slavic hydronyms (river names) are to be found in northern and western Ukraine and southern Belarus. It has also been noted that Proto-Slavic seemingly lacked a maritime vocabulary.

The Proto-Slavic language existed approximately to the middle of the first millennium AD. By the 7th century, it had broken apart into large dialectal zones. Linguistic differentiation received impetus from the dispersion of the Slavic peoples over a large territory – which in Central Europe exceeded the current extent of Slavic-speaking territories. Written documents of the 9th, 10th & 11th centuries already show some local linguistic features.

Church Slavonic language to the Bulgarian Empire, where it was taught and Cyrillic alphabet developed in the Preslav Literary School.

Vowel changes from PIE to Proto-Slavic:

- $i_1 < \text{PIE } \bar{i}, \text{ei};$
- $i_2 < \text{reduced } *ai \text{ } (*\bar{a}i / *ui) < \text{PIE } \textbf{ai}, \textbf{oi};$
- $\text{ь} < *i < \text{PIE } \textbf{i};$
- $e < \text{PIE } \textbf{e};$
- $\text{ę} < \text{PIE } \textbf{en}, \textbf{em};$
- $\check{e}_1 < \text{PIE } *e,$
- $\check{e}_2 < *ai < \text{PIE } \textbf{ai}, \textbf{oi};$
- $a < *\bar{a} < \text{PIE } \bar{a}, \bar{o};$
- $o < *a < \text{PIE } \textbf{a}, \textbf{o}, *a;$
- $\text{о} < *an, *am < \text{PIE } \textbf{an}, \textbf{on}, \textbf{am}, \textbf{om};$
- $\text{ѳ} < *u < \text{PIE } \textbf{u};$
- $y < \text{PIE } \bar{u};$
- $u < *au < \text{PIE } \textbf{au}, \textbf{ou}.$

NOTE 1. Apart from this simplified equivalences, other evolutions appear:

- The vowels i_2 , \check{e}_2 developed later than i_1 , \check{e}_1 . In Late Proto-Slavic there were no differences in pronunciation between i_1 and i_2 as well as between \check{e}_1 and \check{e}_2 . They had caused, however, different changes of preceding velars, see below.

- Late Proto-Slavic yers ь , ѳ < earlier i , u developed also from

reduced PIE **e**, **o** respectively. The reduction was probably a morphologic process rather than phonetic.

- We can observe similar reduction of $*\bar{a}$ into $*\bar{u}$ (and finally y) in some endings, especially in closed syllables.
- The development of the Sla. i_2 was also a morphologic phenomenon, originating only in some endings.
- Another source of the Proto-Slavic y is $*\bar{o}$ in Germanic loanwords – the borrowings took place when Proto-Slavic no longer had \bar{o} in native words, as PIE \bar{o} had already changed into $*\bar{a}$.
- PIE $*\bar{a}$ disappeared without traces when in a non-initial syllable.
- PIE **eu** probably developed into $*jau$ in Early Proto-Slavic (or: during the Balto-Slavic epoch), and eventually into Proto-Slavic $*ju$.

- According to some authors, PIE long diphthongs **ēi**, **āi**, **ōi**, **ēu**, **āu**, **ōu** had twofold development in Early Proto-Slavic, namely they shortened in endings into simple $*ei$, $*ai$, $*oi$, $*eu$, $*au$, $*ou$ but they lost their second element elsewhere and changed into $*\bar{e}$, $*\bar{a}$, $*\bar{o}$ with further development like above.

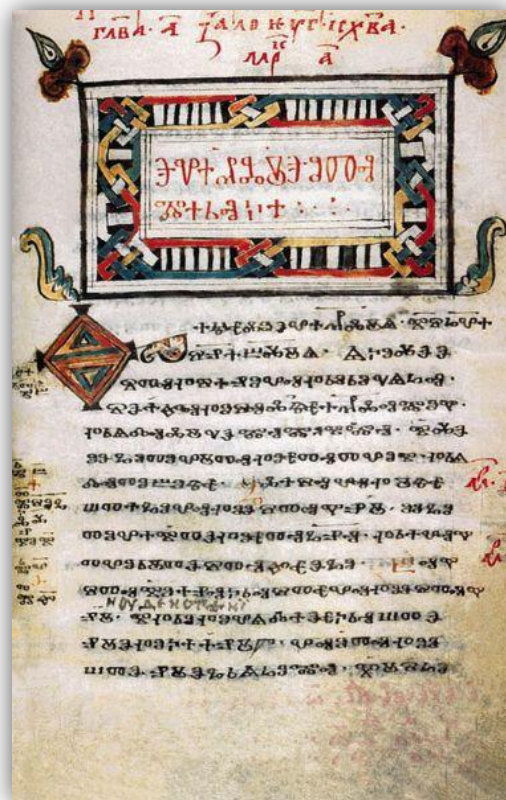


Figure 28. A page from the 10th-11th century Codex Zographensis found in the Zograf Monastery in 1843. It is written in Old Church Slavonic, in the Glagolitic alphabet designed by brothers St Cyril and St Methodius.

NOTE 2. Other vocalic changes from Proto-Slavic include *jo, *jъ, *jy changed into *je, *jъ, *ji; *o, *ъ, *y also changed into *e, *ъ, *i after *c, *ѣ, *s' which developed as the result of the 3rd palatalization; *e, *ě changed into *o, *a after *č, *ѣ, *š, *ž in some contexts or words; a similar change of *ě into *a after *j seems to have occurred in Proto-Slavic but next it can have been modified by analogy.

On the origin of Proto-Slavic consonants, the following relationships are regularly found:

- *p* < PIE *p*;
- *b* < PIE *b*, *b^h*;
- *t* < PIE *t*;
- *d* < PIE *d*, *d^h*;
- *k* < PIE *k*, *k^w*;
 - *s* < PIE **k^s*;
- *g* < PIE *g*, *g^h*, *g^w*, *g^{wh}*;
 - *z* < PIE **g^j*, **g^{j^h}*;
- *s* < PIE *s*;
 - *z* < PIE *s* [z] before a voiced consonant;
 - *x* < PIE *s* before a vowel when after *r*, *u*, *k*, *i*, probably also after *l*;
- *m* < PIE *m*;
- *n* < PIE *n*;
- *l* < PIE *l*;
- *r* < PIE *r*;
- *v* < PIE *w*;
- *j* < PIE *j*.

In some words the Proto-Slavic *x* developed from other PIE phonemes, like *k^h*, *k^s*, *sk*.

About the common changes of Slavic dialects, compare:

1) In the 1st palatalization,

- *k, *g, *x > *č, *ѣ, *š before *i1, *ě1, *e, *ę, *ъ;
- next *ѣ* changed into *ž* everywhere except after *z*;
- *kt, *gt > *tj before *i1, *ě1, *e, *ę, *ъ (there are only examples for *kti).



Figure 29. Page from the Spiridon Psalter in Church Slavonic, a language derived from Old Church Slavonic by adapting pronunciation and orthography, and replacing some old and obscure words and expressions by their vernacular counterparts.

2) In the 2nd palatalization (which apparently didn't occur in old northern Russian dialects)

- $*k, *g, *x > *c, *z, *s'$ before $*i_2, *ě_2$;
- $*s'$ mixed with s or $š$ in individual Slavic dialects;
- $*z$ simplified into z , except Polish;
- also $*kv, *gv, *xv > *cv, *zv, *s'v$ before $*i_2, *ě_2$ in some dialects (not in West Slavic and probably not in East Slavic – Russian examples may be of South Slavic origin);

3) The third palatalization

- $*k, *g, *x > *c, *z, *s'$ after front vowels ($*i, *b, *ě, *e, *e$) and $*br (= *r')$, before a vowel;
- it was progressive contrary to the 1st and the 2nd palatalization;
- it occurred inconsistently, only in certain words, and sometimes it was limited to some Proto-Slavic dialects;

sometimes a palatalized form and a non-palatalized one existed side-by-side even within the same dialect (e.g. O.C.S. *sikъ* || *sicъ* 'such');

In fact, no examples are known for the 3rd palatalization after $*ě, *e$, and (few) examples after $*r'$ are limited to Old Church Slavonic.

In Consonants + **j**

- $*sj, *zj > *š, *ž$;
- $*stj, *zdj > *šč, *žž$;
- $*kj, *gj, *xj > *č, *ž, *š$ (next $*ž > *ž$);
- $*skj, *zgj > *šč, *žž$;
- $*tj, *dj$ had been preserved and developed variously in individual Slavic dialects;
- $*rj, *lj, *nj$ were preserved until the end of Proto-Slavic, next developed into palatalized $*r', *l', *n'$;
- $*pj, *bj, *vj, *mj$ had been preserved until the end of the Proto-Slavic epoch, next developed into $*p', *b', *v', *m'$ in most Slavic dialects, except Western Slavic.

D. BALTIC



Figure 30. Distribution of Baltic languages today and in the past (in stripes)

The **Baltic languages** are a group of related languages belonging to the Indo-European language family and spoken mainly in areas extending east and southeast of the Baltic Sea in Northern Europe.

The language group is sometimes divided into two sub-groups: Western Baltic, containing only extinct languages as Prussian or Galindan, and Eastern Baltic, containing both extinct and the two living languages in the group, Lithuanian and Latvian – including literary Latvian and Latgalian. While related, the Lithuanian, the Latvian, and particularly the Old Prussian vocabularies differ substantially from each other and are not mutually intelligible. The now extinct Old Prussian language has been considered the most archaic of the Baltic languages.

Baltic and Slavic share more close similarities, phonological, lexical, and morpho-syntactic, than any other language groups within the Indo-European language family. Many linguists, following the lead of such notable Indo-Europeanists as August Schleicher and Oswald Szemerényi, take these to indicate that the two groups separated from a common ancestor, the **Proto-Balto-Slavic** language, only well after the breakup of Indo-European.

The first evidence was that many words are common in their form and meaning to Baltic and Slavic, as “run” (cf. Lith. *bėgu*, O.Pruss. *bīgtwei*, Sla. *běgŏ*, Russ. *begu*, Pol. *biegnę*), “tilia” (cf. Lith. *liepa*, Ltv. *liepa*, O.Pruss. *līpa*, Sla. *lipa*, Russ. *lipa*, Pol. *lipa*), etc.

NOTE. The amount of shared words might be explained either by existence of common Balto-Slavic language in the past or by their close geographical, political and cultural contact throughout history.

Until Meillet's *Dialectes indo-européens* of 1908, Balto-Slavic unity was undisputed among linguists – as he notes himself at the beginning of the *Le Balto-Slave* chapter, “*L'unité linguistique balto-slave est l'une de celles que personne ne conteste*” (“Balto-Slavic linguistic unity is one of those that no one contests”). Meillet's critique of Balto-Slavic confined itself to the seven characteristics listed by Karl Brugmann in 1903, attempting to show that no single one of these is sufficient to prove genetic unity.

Szemerényi in his 1957 re-examination of Meillet's results concludes that the Balts and Slavs did, in fact, share a “*period of common language and life*”, and were probably separated due to the incursion of Germanic tribes along the Vistula and the Dnepr roughly at the beginning of the Common Era. Szemerényi notes fourteen points that he judges cannot be ascribed to chance or parallel innovation:

- phonological palatalization
- the development of *i* and *u* before PIE resonants
- ruki Sound law (v.i.)
- accentual innovations
- the definite adjective
- participle inflection in *-yo-*
- the genitive singular of thematic stems in *-ā(t)-*
- the comparative formation
- the oblique 1st singular *men-*, 1st plural *nōsom*
- *tos/tā* for PIE *so/sā* pronoun
- the agreement of the irregular athematic verb (Lithuanian *dúoti*, Slavic *datъ*)
- the preterite in *ē/ā*
- verbs in Baltic *-āju*, Sla. *-uję*
- the strong correspondence of vocabulary not observed between any other pair of branches of the Indo-European languages.
- lengthening of a short vowel before a voiced plosive (Winter)



Figure 31 Baltic Tribes c. 1200 AD.

NOTE. ‘Ruki’ is the term for a sound law which is followed especially in Balto-Slavic and Indo-Iranian dialects. The name of the term comes from the sounds which cause the phonetic change, i.e. PIE *s* > *š* / *r*, *u*, *k*, *i* (it associates with a Slavic word which means ‘hands’ or ‘arms’). A sibilant [s] is retracted to [ʃ] after *i*, *u*, *r*, and after velars (i.e. *k* which may have developed from earlier *k*, *g*, *gh*). Due to the character of the retraction, it was probably an apical sibilant (as in Spanish), rather than the dorsal of English. The first phase (*s* > *š*) seems to be universal, the later retroflexion (in Sanskrit and probably in Proto-Slavic as well) is due to levelling of the sibilant system, and so is the third phase - the retraction to velar [x] in Slavic and also in some Middle Indian languages, with parallels in e.g. Spanish. This rule was first formulated for the Indo-European by Holger Pedersen, and it is known sometimes as the “*Pedersen law*”.

E. CELTIC

The **Celtic languages** are the languages descended from Proto-Celtic, or “Common Celtic”, a dialect of Proto-Indo-European.

During the 1st millennium BC, especially between the 5th and 2nd centuries BC they were spoken across Europe, from the southwest of the Iberian Peninsula and the North Sea, up the Rhine and down the Danube to the Black Sea and the Upper Balkan Peninsula, and into Asia Minor (Galatia). Today, Celtic languages are now limited to a few enclaves in the British Isles and on the peninsula of Brittany in France.

The distinction of Celtic into different sub-families probably occurred about 1000 BC. The early Celts are commonly associated with the archaeological Urnfield culture, the La Tène culture, and the Hallstatt culture.

Scholarly handling of the Celtic languages has been rather argumentative owing to lack of primary source data. Some scholars distinguish Continental and Insular Celtic, arguing that the differences between the Goidelic and Brythonic languages arose after these split off from the Continental Celtic languages. Other scholars distinguish P-Celtic from Q-Celtic, putting most of the Continental Celtic languages in the former group – except for Celtiberian, which is Q-Celtic.

There are two competing schemata of categorization. One scheme, argued for by Schmidt (1988) among others, links Gaulish with Brythonic in a **P-Celtic** node, leaving Goidelic as **Q-Celtic**. The difference between P and Q languages is the treatment of PIE ***kʷ***, which became ****p*** in the P-Celtic languages but ****k*** in Goidelic. An example is the Proto-Celtic verbal root ****kʷrin-*** “to buy”, which became ***pryn-*** in Welsh but ***cren-*** in Old Irish.

The other scheme links Goidelic and Brythonic together as an Insular Celtic branch, while Gaulish and Celtiberian are referred to as Continental Celtic. According to this theory, the ‘P-Celtic’ sound change of [kʷ] to [p] occurred independently or areally. The proponents of the Insular Celtic hypothesis point to other shared innovations among Insular Celtic languages, including inflected prepositions, VSO word order, and the lenition of intervocalic [m] to [β], a nasalized voiced bilabial fricative (an extremely rare

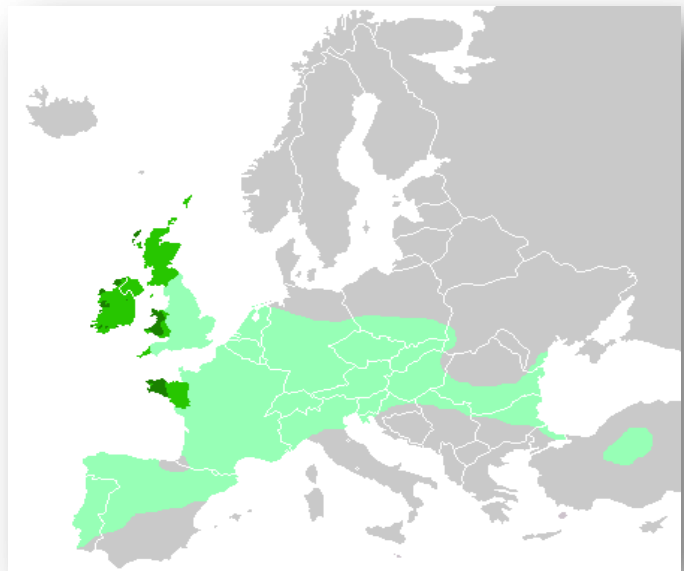


Figure 32. Distribution of Celtic languages in Europe, at its greatest expansion in 500 B.C. in **lighter color**, the so-called ‘Celtic Nations’ in **darker color**, and today’s Celtic-speaking populations in the **darkest color**.

sound), etc. There is, however, no assumption that the Continental Celtic languages descend from a common “Proto-Continental Celtic” ancestor. Rather, the Insular/Continental schemata usually consider Celtiberian the first branch to split from Proto-Celtic, and the remaining group would later have split into Gaulish and Insular Celtic. Known PIE evolutions into Proto-Celtic:

- **p** > Ø in initial and intervocalic positions
- **l** > /li/
- **r** > /ri/
- **g^{wh}** > /g/
- **g^w** > /b/
- **ō** > /ā/, /ū/

NOTE. Later evolution of Celtic languages: **ē** > /ī/; Thematic genitive ***ōd**/***ī**; Aspirated Voiced > Voiced; Specialized Passive in **-r**.



Figure 33. Inscription *CEΓOMAPOC OYIAAONEOC TOOYTIOYC NAMAYCATIC EIΩPOY BHΛH CAMI COCIN NEMHTON*, translated as “Segomaros, son of Uillo, toutious (tribe leader) of Namasos, dedicated this sanctuary to Belesama”.

Italo-Celtic refers to the hypothesis that Italic and Celtic dialects are descended from a common ancestor, Proto-Italo-Celtic, at a stage post-dating Proto-Indo-European. Since both Proto-Celtic and Proto-Italic date to the early Iron Age (say, the centuries on either side of 1000 BC), a probable time frame for the assumed period of language contact would be the late Bronze Age, the early to mid 2nd millennium BC. Such grouping is supported among others by Meillet (1890), and Kortlandt (2007).

One argument for Italo-Celtic was the thematic Genitive in *i* (*dominus, domini*). Both in Italic (*Popliosio Valesiosio*, Lapis Satricanus) and in Celtic (Leponitic, Celtiberian *-o*), however, traces of the *-osyo* Genitive of Proto-Indo-European have been discovered, so that the spread of the *i*-Genitive could have occurred in the two groups independently, or by areal diffusion. The community of *-ī* in Italic and Celtic may be then attributable to early contact, rather than to an original unity. The *i*-Genitive has been compared to the so-called Cvi formation in Sanskrit, but that too is probably a comparatively late development. The phenomenon is probably related to the Indo-European feminine long *i* stems and the Luwian *i*-mutation.

Another argument was the *ā*-subjunctive. Both Italic and Celtic have a subjunctive descended from an earlier optative in *-ā-*. Such an optative is not known from other languages, but the suffix occurs in Balto-Slavic and Tocharian past tense formations, and possibly in Hittite *-ahh-*.

Both Celtic and Italic have collapsed the PIE Aorist and Perfect into a single past tense.

F. FRAGMENTARY DIALECTS

MESSAPIAN

Messapian (also known as *Messapic*) is an extinct Indo-European language of south-eastern Italy, once spoken in the regions of Apulia and Calabria. It was spoken by the three Iapygian tribes of the region: the Messapians, the Daunii and the Peucetii. The language, a centum dialect, has been preserved in about 260 inscriptions dating from the 6th to the 1st century BC.

There is a hypothesis that Messapian was an Illyrian language. The Illyrian languages were spoken mainly on the other side of the Adriatic Sea. The link between Messapian and Illyrian is based mostly on personal names found on tomb inscriptions and on classical references, since hardly any traces of the Illyrian language are left.

The Messapian language became extinct after the Roman Empire conquered the region and assimilated the inhabitants.

Some phonetic characteristics of the language may be regarded as quite certain:

- the change of PIE short **-o-** to **-a-**, as in the last syllable of the genitive *kalatoras*.
- of final **-m** to **-n**, as in *aran*.
- of **-ni-** to **-nn-**, as in the Messapian praenomen *Dazohonnes* vs. the Illyrian praenomen *Dazonius*; the Messapian genitive *Dazohonnihi* vs. Illyrian genitive *Dasonii*, etc.
- of **-ti-** to **-tth-**, as in the Messapian praenomen *Dazetthes* vs. Illyrian *Dazetius*; the Messapian genitive *Dazetthihi* vs. the Illyrian genitive *Dazetii*; from a *Dazet-* stem common in Illyrian and Messapian.
- of **-si-** to **-ss-**, as in Messapian *Vallasso* for *Vallasio*, a derivative from the shorter name *Valla*.
- the loss of final **-d**, as in *tepiše*, and probably of final **-t**, as in *-des*, perhaps meaning “set”, from PIE **dhe-**, “set, put”.
- the change of voiced aspirates in Proto-Indo-European to plain voiced consonants: PIE **dh-** or **dh-** to **d-** or **-d-**, as Mes. *anda* (< PIE **en-dha-** < PIE **en-**, “in”, compare Gk. *entha*), and PIE **bh-** or **bh-** to **b-** or **-b-**, as Mes. *beran* (< PIE **bher-**, “to bear”).
- **-au-** before (at least some) consonants becomes **-ā-**: *Bāsta*, from *Bausta*
- the form *penkaheh* – which Torp very probably identifies with the Oscan stem *pompaio* – a derivative of the Proto-Indo-European numeral **penqe-**, “five”.

If this last identification be correct it would show, that in Messapian (just as in Venetic and Ligurian) the original labiovelars (**k^w**, **g^w**, **g^{wh}**) were retained as gutturals and not converted into labials. The change of *o* to *a* is exceedingly interesting, being associated with the northern branches of Indo-

European such as Gothic, Albanian and Lithuanian, and not appearing in any other southern dialect hitherto known. The Greek *Aphrodite* appears in the form *Aprodita* (Dat. Sg., *fem.*).

The use of double consonants which has been already pointed out in the Messapian inscriptions has been very acutely connected by Deecke with the tradition that the same practice was introduced at Rome by the poet Ennius who came from the Messapian town Rudiae (Festus, p. 293 M).

VENETIC

Venetic is an Indo-European language that was spoken in ancient times in the Veneto region of Italy, between the Po River delta and the southern fringe of the Alps.

The language is attested by over 300 short inscriptions dating between the 6th century BC and 1st century. Its speakers are identified with the ancient people called *Veneti* by the Romans and *Enetoi* by the Greek. It became extinct around the 1st century when the local inhabitants were assimilated into the Roman sphere.

Venetic was a centum dialect. The inscriptions use a variety of the Northern Italic alphabet, similar to the Old Italic alphabet.

The exact relationship of Venetic to other Indo-European languages is still being investigated, but the majority of scholars agree that Venetic, aside from Liburnian, was closest to the Italic languages. Venetic may also have been related to the Illyrian languages, though the theory that Illyrian and Venetic were closely related is debated by current scholarship.

Some important parallels with the Germanic languages have also been noted, especially in pronominal forms:

Ven. *ego*, “I”, acc. *mego*, “me”; Goth. *ik*, acc. *mik*; Lat. *ego*, acc. *me*.

Ven. *sselboisselboi*, “to oneself”; O.H.G. *selb selbo*; Lat. *sibi ipsi*.

Venetic had about six or even seven noun cases and four conjugations (similar to Latin). About 60 words are known, but some were borrowed from Latin (*liber.tos.* < *libertus*) or Etruscan. Many of them show a clear Indo-European origin, such as Ven. *vhraterei* < PIE *bhraterei*, “to the brother”.

In Venetic, PIE stops **bh**, **dh** and **gh** developed to /f/, /f/ and /h/, respectively, in word-initial position (as in Latin and Osco-Umbrian), but to /b/, /d/ and /g/, respectively, in word-internal intervocalic position, as in Latin. For Venetic, at least the developments of **bh** and **dh** are clearly attested. Faliscan and Osco-Umbrian preserve internal /f/, /f/ and /h/.

There are also indications of the developments of PIE **g^w**- > **w**-, PIE **k^w** > ***kv** and PIE ***g^{wh}**- > **f**- in Venetic, all of which are parallel to Latin, as well as the regressive assimilation of PIE sequence **p...k^w**... > **k^w...k^w**..., a feature also found in Italic and Celtic (Lejeune 1974).

LIGURIAN

The **Ligurian language** was spoken in pre-Roman times and into the Roman era by an ancient people of north-western Italy and south-eastern France known as the Ligures. Very little is known about this language (mainly place names and personal names remain) which is generally believed to have been Indo-European; it appears to have adopted significantly from other Indo-European languages, primarily Celtic (Gaulish) and Italic (Latin).

Strabo states *“As for the Alps... Many tribes (éthnê) occupy these mountains, all Celtic (Keltikà) except the Ligurians; but while these Ligurians belong to a different people (hetero-ethneis), still they are similar to the Celts in their modes of life (bíois).”*

LIBURNIAN

The **Liburnian language** is an extinct language which was spoken by the ancient Liburnians, who occupied Liburnia in classical times. The Liburnian language is reckoned as an Indo-European language, usually classified within the Centum group. It appears to have been on the same Indo-European branch as the Venetic language; indeed, the Liburnian tongue may well have been a Venetic dialect.

No writings in Liburnian are known however. The grouping of Liburnian with Venetic is based on the Liburnian onomastics. In particular, Liburnian anthroponyms show strong Venetic affinities, with many common or similar names and a number of common roots, such as *Vols-*, *Volt-*, and *Host-* (<PIE **ghos-ti-**, “*stranger, guest, host*”). Liburnian and Venetic names also share suffixes in common, such as *-icus* and *-ocus*.

These features set Liburnian and Venetic apart from the Illyrian onomastic province, though this does not preclude the possibility that Venetic-Liburnian and Illyrian may have been closely related, belonging to the same Indo-European branch. In fact, a number of linguists argue that this is the case, based on similar phonetic features and names in common between Venetic-Liburnian on the one hand and Illyrian on the other.

The Liburnians were conquered by the Romans in 35 BC. The Liburnian language eventually was replaced by Latin, undergoing language death probably very early in the Common era.

LUSITANIAN

Lusitanian (so named after the Lusitani or Lusitanians) was a paleo-Iberian Indo-European language known by only five inscriptions and numerous toponyms and theonyms. The language was spoken before the Roman conquest of Lusitania, in the territory inhabited by Lusitanian tribes, from Douro to the Tagus rivers in the Iberian Peninsula.

The Lusitanians were the most numerous people in the western area of the Iberian peninsula, and there are those who consider that they came from the Alps; others believe the Lusitanians were a native Iberian tribe. In any event, it is known that they were established in the area before the 6th century BC.

Lusitanian appears to have been an Indo-European language which was quite different from the languages spoken in the centre of the Iberian Peninsula. It would be more archaic than the Celtiberian language.

The affiliation of the Lusitanian language is still in debate. There are those who endorse that it is a Celtic language. This Celtic theory is largely based upon the historical fact that the only Indo-European tribes that are known to have existed in Portugal at that time were Celtic tribes. The apparent Celtic character of most of the lexicon —anthroponyms and toponyms — may also support a Celtic affiliation.

There is a substantial problem in the Celtic theory however: the preservation of initial /p/, as in Lusitanian *pater* or *porcom*, meaning “father” and “pig”, respectively. The Celtic languages had lost that initial /p/ in their evolution; compare Lat. *pater*, Gaul. *ater*, and Lat. *porcum*, O.Ir. *orc*. However, the presence of this /p/ does not necessarily preclude the possibility of Lusitanian being Celtic, because it could have split off from Proto-Celtic before the loss of /p/, or when /p/ had become /ϕ/ (before shifting to /h/ and then being lost); the letter *p* could have been used to represent either sound.

A second theory, defended by Francisco Villar and Rosa Pedrero, relates Lusitanian with the Italic languages. The theory is based on parallels in the names of deities, as Lat. *Consus*, Lus. *Cossue*, Lat. *Seia*, Lus. *Segia*, or Marrucian *Iovia*, Lus. *Iovea(i)*, etc. and other lexical items, as Umb. *gomia*, Lus. *comaiam*, with some other grammatical elements.

Inscriptions have been found in Spain in Arroyo de la Luz (Cáceres), and in Portugal in Cabeço das Fragas (Guarda) and in Moledo (Viseu).

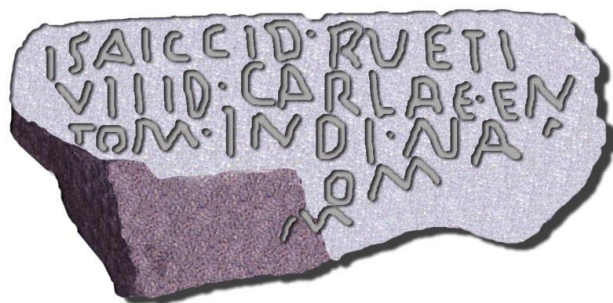


Figure 34. Arroyo de la Luz (Cáceres)
Inscription: ISAICCID. RVETI. PVPPID. CARLAE.
EN ETOM. INDI. NA(.) (....) CE. IOM. M

G. NORTHERN INDO-EUROPEAN IN ASIA: TOCHARIAN

Tocharian or **Tokharian** is one of the most obscure branches of the group of Indo-European languages. The name of the language is taken from people known to the Greek historians (Ptolemy VI, 11, 6) as the Tocharians (Greek Τόχαροι, “*Tokharoi*”). These are sometimes identified with the Yuezhi and the Kushans, while the term *Tokharistan* usually refers to 1st millennium Bactria. A

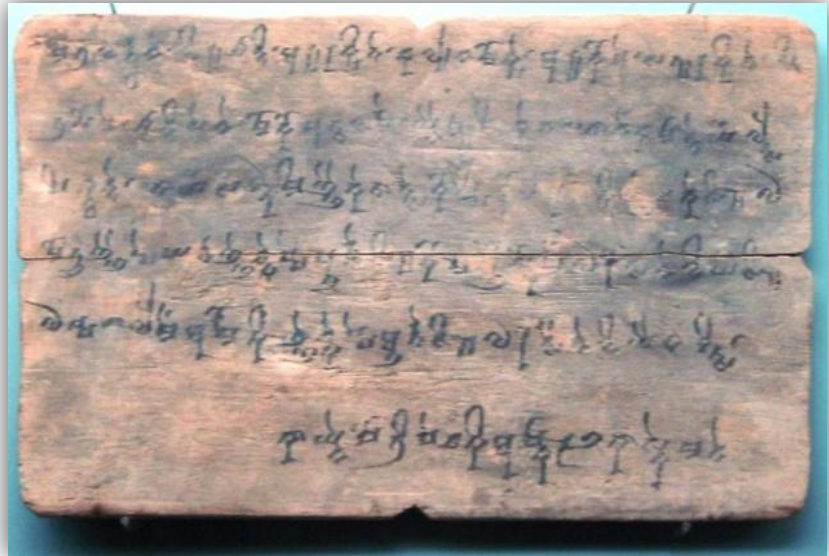


Figure 35. Wooden plate with inscriptions in Tocharian.
Kucha, China, 5th-8th century.

Turkic text refers to the Turfanian language (Tocharian A) as *twqry*. Interpretation is difficult, but F. W. K. Müller has associated this with the name of the Bactrian *Tokharoi*. In Tocharian, the language is referred to as *arish-käna* and the Tocharians as *arya*.

Tocharian consisted of two languages; **Tocharian A** (Turfanian, Arsi, or East Tocharian) and **Tocharian B** (Kuchean or West Tocharian). These languages were spoken roughly from the 6th to 9th century centuries; before they became extinct, their speakers were absorbed into the expanding Uyghur tribes. Both languages were once spoken in the Tarim Basin in Central Asia, now the Xinjiang Autonomous Region of China.

Tocharian is documented in manuscript fragments, mostly from the 8th century (with a few earlier ones) that were written on palm leaves, wooden tablets and Chinese paper, preserved by the extremely dry climate of the Tarim Basin. Samples of the language have been discovered at sites in Kucha and Karasahr, including many mural inscriptions.

Tocharian A and B are not intercomprehensible. Properly speaking, based on the tentative interpretation of *twqry* as related to *Tokharoi*, only Tocharian A may be referred to as *Tocharian*, while Tocharian B could be called *Kuchean* (its native name may have been *kuśiññe*), but since their grammars are usually treated together in scholarly works, the terms A and B have proven useful. The common Proto-Tocharian language must precede the attested languages by several centuries, probably dating to the 1st millennium BC.

1.7.2. SOUTHERN INDO-EUROPEAN DIALECTS

A. GREEK

Greek (Gk. Ελληνικά, “*Hellenic*”) is an Indo-European branch with a documented history of 3,500 years. Today, Modern Greek is spoken by 15 million people in Greece, Cyprus, the former Yugoslavia, particularly the former Yugoslav Republic of Macedonia, Bulgaria, Albania and Turkey.

Greek has been written in the Greek alphabet, the first true alphabet, since the 9th century B.C. and before that, in Linear B and the Cypriot syllabaries. Greek literature has a long and rich tradition.



Figure 36. Location of Ancient Greek dialects by 400 BC.

Greek has been spoken in the Balkan Peninsula since the 2nd millennium BC. The earliest evidence of this is found in the Linear B tablets dating from 1500 BC. The later Greek alphabet is unrelated to Linear B, and was derived from the Phoenician alphabet; with minor modifications, it is still used today.

Mycenaean is the most ancient attested form of the Greek branch, spoken on mainland Greece and on Crete in the 16th to 11th centuries BC, before the Dorian invasion. It is preserved in inscriptions in Linear B, a script invented on Crete before the 14th century BC. Most instances of these inscriptions are on clay tablets found in Knossos and in Pylos. The language is named after Mycenae, the first of the palaces to be excavated.

The tablets remained long undeciphered, and every conceivable language was suggested for them, until Michael Ventris deciphered the script in 1952 and proved the language to be an early form of Greek or closely related to the Greek branch of Indo-European.

The texts on the tablets are mostly lists and inventories. No prose narrative survives, much less myth or poetry. Still, much may be glimpsed from these records about the people who produced them, and about the Mycenaean period at the eve of the so-called Greek Dark Ages.

Unlike later varieties of Greek, Mycenaean Greek probably had seven grammatical cases, the nominative, the genitive, the accusative, the dative, the instrumental, the locative, and the vocative. The instrumental and the locative however gradually fell out of use.

NOTE. For the Locative in **-ei**, compare *di-da-ka-re*, ‘*didaskalei*’, *e-pi-ko-e*, ‘*Epikóhei*’, etc (in Greek there are syntactic compounds like *puloi-genēs*, ‘born in Pylos’); also, for remains of an Ablative case in **-ōd**, compare (months’ names) *ka-ra-e-ri-jo-me-no*, *wo-de-wi-jo-me-no*, etc.

Proto-Greek, a Centum dialect within the southern IE dialectal group (very close to Mycenaean), does appear to have been affected by the general trend of palatalization characteristic of the Satem group, evidenced for example by the (post-Mycenaean) change of labiovelars into dentals before *e* (e.g. *k^we* > *te* “and”).

The primary sound changes from PIE to Proto-Greek include

- Aspiration of /s/ -> /h/ intervocalic
- De-voicing of voiced aspirates.
- Dissimilation of aspirates (Grassmann's law), possibly post-Mycenaean.
- word-initial *j*- (not *Hj*-) is strengthened to *dj*- (later ζ-)

The loss of prevocalic ***s** was not completed entirely, famously evidenced by *sus* “sow”, *dasus* “dense”; *sun* “with” is another example, sometimes considered contaminated with PIE **kom** (Latin *cum*, Proto-Greek **kon*) to Homeric / Old Attic *ksun*, although probably consequence of Gk. psi-substrate (Villar).

Sound changes between Proto-Greek and Mycenaean include:

- Loss of final stop consonants; final /m/ -> /n/.
- Syllabic /m/ and /n/ -> /am/, /an/ before resonants; otherwise /a/.
- Vocalization of laryngeals between vowels and initially before consonants to /e/, /a/, /o/ from *h*₁, *h*₂, *h*₃ respectively.

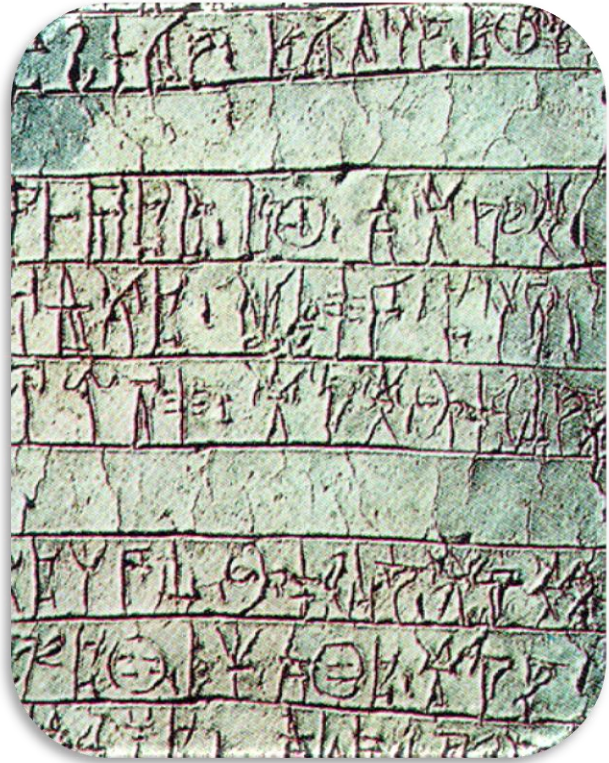


Figure 37 Linear B has roughly 200 signs, divided into syllabic signs with phonetic values and logograms (or ideograms) with semantic values

- The sequence CRHC (C = consonant, R = resonant, H = laryngeal) becomes CRēC, CRāC, CRōC from H = **h*₁, **h*₂, **h*₃, respectively.

- The sequence CRHV (C = consonant, R = resonant, H = laryngeal, V = vowel) becomes CaRV.

- loss of *s* in consonant clusters, with supplementary lengthening, *esmi* -> *ēmi*

- creation of secondary *s* from clusters, *ntia* -> *nsa*. Assibilation *ti* -> *si* only in southern dialects.

The PIE dative, instrumental and locative cases are syncretized into a single dative case. Some desinences are innovated, as e.g. dative plural *-si* from locative plural *-su*.

Nominative plural *-oi*, *-ai* replaces late PIE *-ōs*, *-ās*.

The superlative on *-tatos* (PIE *-tm-to-s*) becomes productive.

The peculiar oblique stem *gunaik-* “women”, attested from the Thebes tablets is probably Proto-Greek; it appears, at least as *gunai-* also in Armenian.

The pronouns *houtos*, *ekeinos* and *autos* are created. Use of *ho*, *hā*, *ton* as articles is post-Mycenaean.

An isogloss between Greek and the closely related Phrygian is the absence of *r*-endings in the Middle in Greek, apparently already lost in Proto-Greek.

Proto-Greek inherited the augment, a prefix *é-* to verbal forms expressing past tense. This feature it shares only with Indo-Iranian and Phrygian (and to some extent, Armenian), lending support to a Southern or Graeco-Aryan Dialect.

The first person middle verbal desinences *-mai*, *-mān* replace *-ai*, *-a*. The third singular *pherei* is an analogical innovation, replacing expected Doric **phereti*, Ionic **pheresi* (from PIE *bhéreti*).

The future tense is created, including a future passive, as well as an aorist passive.

The suffix *-ka-* is attached to some perfects and aorists.

Infinitives in *-ehen*, *-enai* and *-men* are created.

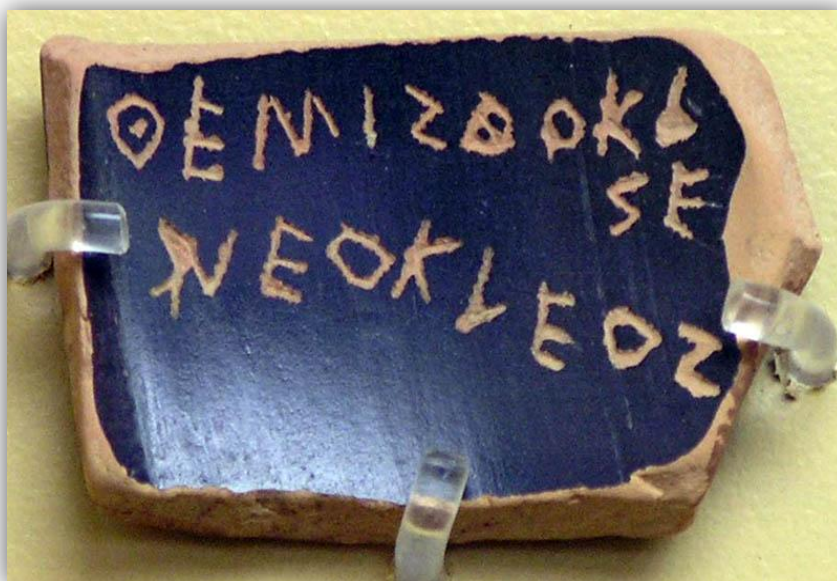


Figure 38. A ballot voting for Themistocles, son of Neocles, under the Athenian Democracy, ca. 470 BC.

B. ARMENIAN



Figure 39. Distribution of Armenian speakers in the 20th Century.

Armenian is an Indo-European language spoken in the Armenian Republic and also used by Armenians in the Diaspora. It constitutes an independent branch of the Indo-European language family.

Armenian is regarded as a close relative of Phrygian. From the modern languages Greek seems to be the most closely related to Armenian, sharing major isoglosses with it. Some linguists have proposed that the linguistic ancestors of the Armenians and Greeks were either identical or in a close contact relation.

The earliest testimony of the Armenian language dates to the 5th century AD, the Bible translation of Mesrob Mashtots. The earlier history

of the language is unclear and the subject of much speculation. It is clear that Armenian is an Indo-European language, but its development is opaque. The Graeco-Armenian hypothesis proposes a close relationship to the Greek language, putting both in the larger context of Paleo-Balkans languages – notably including Phrygian, which is widely accepted as an Indo-European language particularly close to Greek, and sometimes Ancient Macedonian –, consistent with Herodotus' recording of the Armenians as descending from colonists of the Phrygians.

In any case, Armenian has many layers of loanwords, and shows traces of long language contact with Hurro-Urartian, Greek and Iranian.

The Proto-Armenian sound-laws are varied and eccentric, such as **dw-* yielding *erk-*, and in many cases still uncertain.

PIE voiceless stops are aspirated in Proto-Armenian, a circumstance that gave rise to the Glottalic theory, which postulates that this aspiration may have been sub-phonematic already in PIE. In certain contexts, these aspirated stops are further reduced to *w*, *h* or zero in Armenian (as IE **pods**, supposed PIE **pots*, into Armenian *otn*, Greek *pous* “foot”; PIE *treis*, Armenian *erek*, Greek *treis* “three”).

The reconstruction of Proto-Armenian being very uncertain, there is no general consensus on the date range when it might have been alive. If Herodotus is correct in deriving Armenians from Phrygian stock, the Armenian-Phrygian split would probably date to between roughly the 12th and 7th centuries

BC, but the individual sound-laws leading to Proto-Armenian may have occurred at any time preceding the 5th century AD. The various layers of Persian and Greek loanwords were likely acquired over the course of centuries, during Urartian (pre-6th century BC) Achaemenid (6th to 4th c. BC; Old Persian), Hellenistic (4th to 2nd c. BC Koine Greek) and Parthian (2nd c. BC to 3rd c. AD; Middle Persian) times.

The Armenians according to Diakonoff, are then an amalgam of the Hurrian (and Urartians), Luvians and the Proto-Armenian Mushki who carried their IE language eastwards across Anatolia. After arriving in its historical territory, Proto-Armenian would appear to have undergone massive influence on part the languages

it eventually replaced. Armenian phonology, for instance, appears to have been greatly affected by Urartian, which may suggest a long period of bilingualism.

Grammatically, early forms of Armenian had much in common with classical Greek and Latin, but the modern language (like Modern Greek) has undergone many transformations. Interestingly enough, it shares with Italic dialects the secondary IE suffix *-tio(n)*, extended from *-ti*, cf. Arm թյուն (*t'youn*).



Figure 40 Armenian manuscript,
ca. 5th-6th AD

C. INDO-IRANIAN

The **Indo-Iranian language** group constitutes the easternmost extant branch of the Indo-European family of languages. It consists of four language groups: the Indo-Aryan, Iranian, Nuristani, and Dardic – sometimes classified within the Indic subgroup. The term **Aryan languages** is also traditionally used to refer to the Indo-Iranian languages.

The contemporary Indo-Iranian languages form the largest sub-branch of Indo-European, with more than one billion speakers in total, stretching from Europe (Romani) and the Caucasus (Ossetian) to East India (Bengali and Assamese). A 2005 estimate counts a total of 308 varieties, the largest in terms of native speakers being Hindustani (Hindi and Urdu, ca. 540 million), Bengali (ca. 200 million), Punjabi (ca. 100 million), Marathi and Persian (ca. 70 million each), Gujarati (ca. 45 million), Pashto (40 million), Oriya (ca. 30 million), Kurdish and Sindhi (ca. 20 million each).

The speakers of the Proto-Indo-Iranian language, the Proto-Indo-Iranians, are usually associated with the late 3rd millennium BC Sintashta-Petrovka culture of Central Asia. Their expansion is believed to have been connected with the invention of the chariot.

The main change separating Proto-Indo-Iranian from Late PIE, apart from the satemization, is the collapse of the ablauting vowels *e*, *o*, *a* into a single vowel, Ind.-Ira. **a* (but see Brugmann's law in Appendix II). Grassmann's law, Bartholomae's law, and the Ruki sound law were also complete in Proto-Indo-Iranian. Among the sound changes from Proto-Indo-Iranian to Indo-Aryan is the loss of the voiced sibilant **z*, among those to Iranian is the de-aspiration of the PIE voiced aspirates.



Figure 41. Current distribution of Indo-Iranian dialects in Asia.

Proto-Indo-Iranian	Old Iranian	Vedic Sanskrit
<i>*açva</i> ("horse")	Av., O.Pers. <i>aspa</i>	<i>aśva</i>
<i>*bhag-</i>	O.Pers. <i>baj-</i> (<i>bāji</i> ; "tribute")	<i>bhag-</i> (<i>bhaga</i>)
<i>*bhrātr-</i> ("brother")	O.Pers. <i>brātar</i>	<i>bhrātṛ</i>
<i>*bhūmī</i> ("earth", "land")	O.Pers. <i>būmi</i>	<i>bhūmī</i>
<i>*martya</i> ("mortal", "man")	O.Pers. <i>martya</i>	<i>martya</i>
<i>*māsa</i> ("moon")	O.Pers. <i>māha</i>	<i>māsa</i>
<i>*vāsara</i> ("early")	O.Pers. <i>vāhara</i> ("spring")	<i>vāsara</i> ("morning")
<i>*arta</i> ("truth")	Av. <i>aša</i> , O.Pers. <i>arta</i>	<i>ṛta</i>
<i>*draugh-</i> ("falsehood")	Av. <i>druj</i> , O.Pers. <i>draug-</i>	<i>druh-</i>
<i>*sauma</i> "pressed (juice)"	Av. <i>haoma</i>	<i>soma</i>

I. IRANIAN

KURDISH

The **Kurdish language** (*Kurdî* in Kurdish) is spoken in the region loosely called Kurdistan, including Kurdish populations in parts of Iran, Iraq, Syria and Turkey. Kurdish is an official language in Iraq while it is banned in Syria. The number of speakers in Turkey is deemed to be more than 15 million.

The original language of the people in the area of Kurdistan was Hurrian, a non-IE language belonging to the Caucasian family. This older language was replaced by an Iranian dialect around 850 BC, with the arrival of the Medes. Nevertheless, Hurrian influence on Kurdish is still evident in its ergative grammatical structure and in its toponyms.



Figure 42. Current distribution of Kurdish-speaking population in the Near East.

OSSETIC

Ossetic or **Ossetian** (Ossetic Ирон æвзæг, *Iron ævzhağ* or Иронæй, *Ironau*) is an Iranian language spoken in Ossetia, a region on the slopes of the Caucasus Mountains, on the borders of the Russian Federation and Georgia.

The Russian area is known as *North Ossetia-Alania*, while the area in Georgia is called *South Ossetia* or *Samachablo*. Ossetian speakers number about 700.000, sixty percent of whom live in Alania, and twenty percent in South Ossetia

Ossetian, together with Kurdish, Tati and Talyshi, is one of the main Iranian languages with a sizeable community of speakers in the Caucasus. It is descended from Alanic, the language of the Alans, medieval tribes emerging from the earlier Sarmatians. It is believed to be the only surviving descendant of a Sarmatian language. The closest genetically related language is the Yaghnobi language of Tajikistan, the only other living member of the Northeastern Iranian branch. Ossetic has a plural formed by the suffix *-ta*, a feature it shares with Yaghnobi, Sarmatian and the now-extinct Sogdian; this is taken as evidence of a formerly wide-ranging Iranian-language dialect continuum on the Central Asian steppe. The Greek-derived names of ancient Iranian tribes in fact reflect this special plural, e.g. Saromatae (Σαρομάται) and Masagetae (Μασαγέται).

II. INDO-ARYAN

ROMANY LANGUAGES

Romany (or Romani) is the term used for the Indo-European languages of the European Roma and Sinti. These Indo-Aryan languages should not be confused with either Romanian or Romansh, both of which are Romance languages.

The **Roma** people, often referred to as Gypsies, are an ethnic group who live primarily in Europe. They are believed to be descended from nomadic peoples from northwestern India and Pakistan who began a Diaspora from the eastern end of the Iranian Plateau into Europe and North Africa about 1.000 years ago. **Sinte** or Sinti is the name some communities of the nomadic people usually called Gypsies in English prefer for themselves. This includes communities known in German and Dutch as *Zigeuner* and in Italian as *Zingari*. They are closely related to, and are usually considered to be a subgroup of, the Roma people. Roma and Sinte do not form a majority in any state.

Today's dialects of Romany are differentiated by the vocabulary accumulated since their departure from Anatolia, as well as through divergent phonemic evolutions and grammatical features. Many Roma no longer speak the language or speak various new contact languages from the local language with the addition of Romany vocabulary.

There are independent groups currently working toward standardizing the language, including groups in Romania, Serbia, Montenegro, the United States, and Sweden. A standardized form of Romani is used in Serbia, and in Serbia's autonomous province of Vojvodina Romani is one of the officially recognized languages of minorities having its own radio stations and news broadcasts.

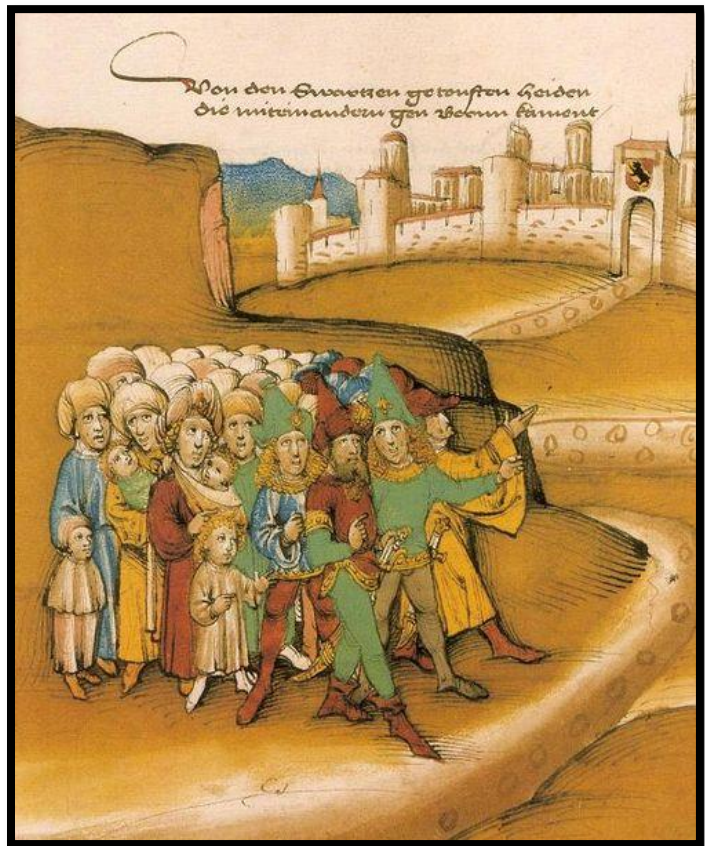


Figure 43. First arrival of the Roma outside Berne in the 15th century, described by the chronicler as *getoufte heiden* "baptized heathens" and drawn with dark skin and wearing Saracen-style clothing and weapons (Spiezer Schilling, p. 749).

A long-standing common categorization was a division between the *Vlax* (from *Vlach*) from *non-Vlax* dialects. *Vlax* are those Roma who lived many centuries in the territory of Romania. The main distinction between the two groups is the degree to which their vocabulary is borrowed from Romanian. *Vlax*-speaking groups include the great number of speakers, between half and two-thirds of all Romani speakers. Bernard Gillad Smith first made this distinction, and coined the term *Vlax* in 1915 in the book *The Report on the Gypsy tribes of North East Bulgaria*. Subsequently, other groups of dialects were recognized, primarily based on geographical and vocabulary criteria, including:

- Balkan Romani: in Albania, Bulgaria, Greece, Macedonia, Moldova, Montenegro, Serbia, Romania, Turkey and Ukraine.
- Romani of Wales.
- Romani of Finland.
- Sinte: in Austria, Croatia, the Czech Republic, France, Germany, Italy, the Netherlands, Poland, Serbia, Montenegro, Slovenia, and Switzerland.
- Carpathian Romani: in the Czech Republic, Poland (particularly in the south), Slovakia, Hungary, Romania, and Ukraine.
- Baltic Romani: in Estonia, Latvia, Lithuania, Poland, Belarus, Ukraine and Russia.
- Turkish dialects:
 - Rumeli (Thrace) dialect (Thrace, Uskudar, a district on the Anatolian side of the Bosphorus): most loanwords are from Greek.
 - Anatolian dialect. Most loanwords are from Turkish, Kurdish and Persian.
 - Posha dialect, Armenian Gypsies from eastern Anatolia mostly nomads although some have settled in the region of Van, Turkey. The Kurds call them *Mytryp* (settled ones).

Some Roma have developed Creole languages or mixed languages, including:

- *Caló* or Iberian-Romani, which uses the Romani lexicon and Spanish grammar (the *Calé*).
- Romungro.
- Lomavren or Armenian-Romani.
- Angloromani or English-Romani.
- Scandoromani (Norwegian-Traveller Romani or Swedish-Traveller Romani).
- Romano-Greek or Greek-Romani.
- Romano-Serbian or Serbian-Romani.
- Boyash, a dialect of Romanian with Hungarian and Romani loanwords.
- Sinti-Manouche-Sinti (Romani with German grammar).

1.7.3. OTHER INDO-EUROPEAN DIALECTS OF EUROPE

A. ALBANIAN

Albanian (*gjuha shqipe*) is a language spoken by over 8 million people primarily in Albania, Kosovo, and the Former Yugoslav Republic of Macedonia, but also by smaller numbers of ethnic Albanians in other parts of the Balkans, along the eastern coast of Italy and in Sicily, as well other emigrant groups. The language forms its own distinct branch of the Indo-European languages.

The Albanian language has no living close relatives among the modern languages. There is no scholarly consensus over its origin and dialectal classification. Some scholars maintain that it derives from the Illyrian language, and others claim that it derives from Thracian.



Figure 44. Albanian language and its dialects Gheg, Tosk (also Arbëreshë and Arvanitika)

While it is considered established that the Albanians originated in the Balkans, the exact location from which they spread out is hard to pinpoint. Despite varied claims, the Albanians probably came from farther north and inland than would suggest the present borders of Albania, with a homeland concentrated in the mountains.

Given the overwhelming amount of shepherding and mountaineering vocabulary as well as the extensive influence of Latin, it is more likely the Albanians come from north of the Jireček line, on the Latin-speaking side, perhaps in part from the late Roman province of *Dardania* from the western Balkans. However, archaeology has more convincingly pointed to the early Byzantine province of *Praevitana* (modern northern Albania) which shows an area where a primarily shepherding, transhumance population of Illyrians retained their culture.

The period in which Proto-Albanian and Latin interacted was protracted and drawn out over six centuries, 1st c. AD to 6th or 7th c. AD. This is born out into roughly three layers of borrowings, the largest number belonging to the second layer. The first, with the fewest borrowings, was a time of less important interaction. The final period, probably preceding the Slavic or Germanic invasions, also has a notably smaller amount of borrowings. Each layer is characterized by a different treatment of most vowels, the first layer having several that follow the evolution of Early Proto-Albanian into Albanian; later layers reflect vowel changes endemic to Late Latin and presumably Proto-Romance. Other

formative changes include the syncretism of several noun case endings, especially in the plural, as well as a large scale palatalization.

A brief period followed, between 7th c. AD and 9th c. AD, that was marked by heavy borrowings from Southern Slavic, some of which predate the “o-a” shift common to the modern forms of this language group. Starting in the latter 9th c. AD, a period followed of protracted contact with the Proto-Romanians, or Vlachs, though lexical borrowing seems to have been mostly one sided – from Albanian into Romanian. Such a borrowing indicates that the Romanians migrated from an area where the majority was Slavic (i.e. Middle Bulgarian) to an area with a majority of Albanian speakers, i.e. Dardania, where Vlachs are recorded in the 10th c. AD. This fact places the Albanians at a rather early date in the Western or Central Balkans, most likely in the region of Kosovo and Northern Albania.

References to the existence of Albanian as a distinct language survive from the 1300s, but without recording any specific words. The oldest surviving documents written in Albanian are the *Formula e Pagëzimit* (Baptismal formula), *Un'te paghesont' pr'emenit t'Atit e t'Birit e t'Spirit Senit*, “I baptize thee in the name of the Father, and the Son, and the Holy Spirit”, recorded by Pal Engjelli, Bishop of Durres in 1462 in the Gheg dialect, and some New Testament verses from that period.

B. PALEO-BALKAN LANGUAGES

PHRYGIAN

The **Phrygian language** was the Indo-European language spoken by the Phrygians, a people that settled in Asia Minor during the Bronze Age.

Phrygian is attested by two corpora, one, Paleo-Phrygian, from around 800 BC and later, and another after a period of several centuries, Neo-Phrygian, from around the beginning of the Common Era. The Palaeo-Phrygian corpus is further divided (geographically) into inscriptions of **Midas-city**, Gordion, Central, Bithynia, Pteria, Tyana, Daskyleion, Bayindir, and “various” (*documents divers*). The **Mysian** inscriptions show a language classified as a separate Phrygian



Figure 45. Traditional Phrygian region and expanded Kingdom.

dialect, written in an alphabet with an additional letter, the “Mysian s”. We can reconstruct some words with the help of some inscriptions written with a script similar to the Greek one.

The language survived probably into the sixth century AD, when it was replaced by Greek.

Ancient historians and myths sometimes did associate Phrygian with Thracian and maybe even Armenian, on grounds of classical sources. Herodotus recorded the Macedonian account that Phrygians emigrated into Asia Minor from Thrace (7.73). Later in the text (7.73), Herodotus states that the Armenians were colonists of the Phrygians, still considered the same in the time of Xerxes I. The earliest mention of Phrygian in Greek sources, in the *Homeric Hymn to Aphrodite*, depicts it as different from Trojan: in the hymn, Aphrodite, disguising herself as a mortal to seduce the Trojan prince Anchises, tells him

“Otreus of famous name is my father, if so be you have heard of him, and he reigns over all Phrygia rich in fortresses. But I know your speech well beside my own, for a Trojan nurse brought me up at home”. Of Trojan, unfortunately, nothing is known.

Its structure, what can be recovered from it, was typically Indo-European, with nouns declined for case (at least four), gender (three) and number (singular and plural), while the verbs are conjugated for tense, voice, mood, person and number. No single word is attested in all its inflectional forms.

Many words in Phrygian are very similar to the reconstructed Proto-Indo-European forms. Phrygian seems to exhibit an augment, like Greek and Armenian, c.f. *eberet*, probably corresponding to PIE **é-bher-e-t* (Greek *epheret*).

A sizable body of Phrygian words are theoretically known; however, the meaning and etymologies and even correct forms of many Phrygian words (mostly extracted from inscriptions) are still being debated.

A famous Phrygian word is *bekos*, meaning “bread”. According to Herodotus (*Histories* 2.9) Pharaoh Psammetichus I wanted to establish the original language. For this purpose, he ordered two children to be reared by a shepherd, forbidding him to let them hear a single word, and charging him to report the children's first utterance. After two years, the shepherd reported that on entering their chamber, the children came up to him, extending their hands, calling *bekos*. Upon enquiry, the pharaoh discovered that this was the Phrygian word for “wheat bread”, after which the Egyptians conceded that the Phrygian nation was older than theirs. The word *bekos* is also attested several times in Palaeo-Phrygian inscriptions on funerary stelae. It was suggested that it is cognate to English *bake*, from PIE **bheh₃g*; cf. Greek *phōgō*, “to roast”, Latin *focus*, “fireplace”, Armenian *bosor*, “red”, and *bots* “flame”, Irish *goba* “smith”, and so on.



Figure 46. Phrygian inscription in Midas City.

Bedu according to Clement of Alexandria's *Stromata*, quoting one Neanthus of Cyzicus means “water” (PIE **wed*). The Macedonians are said to have worshiped a god called *Bedu*, which they interpreted as “air”. The god appears also in Orphic ritual.

Other Phrygian words include:

- *anar*, 'husband', from PIE **ner-* 'man'; cf. Gk. *anēr* (ανήρ) “man, husband”, O.Ind. *nara*, *nṛ*, Av. *nā/nar-*, Osc. *ner-um*, Lat. *Nero*, Welsh *ner*, Alb. *ñjeri* “man, person”.
- *attagos*, 'goat'; cf. Gk. *tragos* (τράγος) “goat”, Ger. *Ziege* “goat”, Alb. *dhi* “she-goat”.
- *balaïos*, 'large, fast', from PIE **bel-* 'strong'; cognate to Gk. *belteros* (βέλτερος) “better”, Rus. *bol'shói* “large, great”, Welsh *balch* “proud”.
- *belte*, 'swamp', from PIE **b^hel-*, 'to gleam'; cf. Gk. *baltos* (βάλτος) “swamp”, Alb. *baltë*, “silt, mud”, Bulg. *blato* (O.Bulg. *balta*) “swamp”, Lith. *baltas* “white”, Russ. *bledny*, Bulg. *bleden* “pale”.
- *brater*, 'brother', from PIE **b^hrater-*, 'brother';
- *daket*, 'does, causes', PIE **d^he-k-*, 'to set, put';
- *germe*, 'warm', PIE **g^{wh}er-*, 'warm'; cf. Gk. *thermos* (θερμός) “warm”, Pers. *gar-me* “warm”, Arm. *ǰerm* “warm”, Alb. *zjarm* “warm”.
- *kakon*, 'harm, ill', PIE **kaka-*, 'harm'; cf. Gk. *kakós* (κακός) “bad”, Alb. *keq* “bad, evil”, Lith. *keñti* “to be evil”.
- *knoumane*, 'grave', maybe from PIE **knu-*, 'to scratch'; cf. Gk. *knaō* (κνάω) “to scratch”, Alb. *krromë* “scurf, scabies”, O.H.G. *hnuo* “notch, groove”, *nuoen* “to smooth out with a scraper”, Lith. *knisti* “to dig”.
- *manka*, 'stela'.
- *mater*, 'mother', from PIE **mater-*, 'mother';
- *meka*, 'great', from PIE **meg-*, 'great';
- *zamelon*, 'slave', PIE **d^hg^hom-*, 'earth'; cf. Gk. *chamelos* (χαμηλός) “adj. on the ground, low”, Sr.-Cr. *zèmlja* and Bul. *zèmya/zèmlishte* “earth/land”, Lat. *humilis* “low”.

THRACIAN

Excluding Dacian, whose status as a Thracian language is disputed, Thracian was spoken in substantial numbers in what is now southern Bulgaria, parts of Serbia, the Republic of Macedonia, Northern Greece – especially prior to Ancient Macedonian expansion –, throughout Thrace (including European Turkey) and in parts of Bithynia (North-Western Asiatic Turkey).

As an extinct language with only a few short inscriptions attributed to it (v.i.), there is little known about the Thracian language, but a number of features are agreed upon. A number of probable Thracian

words are found in inscriptions – most of them written with Greek script – on buildings, coins, and other artifacts.

Thracian words in the Ancient Greek lexicon are also proposed. Greek lexical elements may derive from Thracian, such as *balios*, “dappled” (< PIE **bhel-*, “to shine”, Pokorny also cites Illyrian as a possible source), *bounos*, “hill, mound”, etc.

Most of the Thracians were eventually Hellenized – in the province of Thrace – or Romanized – in Moesia, Dacia, etc. –, with the last remnants surviving in remote areas until the 5th century.

DACIAN

The **Dacian language** was an Indo-European language spoken by the ancient people of Dacia. It is often considered to have been a northern variant of the Thracian language or closely related to it.

There are almost no written documents in Dacian. Dacian used to be one of the major languages of South-Eastern Europe, stretching from what is now Eastern Hungary to the Black Sea shore. Based on archaeological findings, the origins of the Dacian culture are believed to be in Moldavia, being identified as an evolution of the Iron Age Basarabi culture.

It is unclear exactly when the Dacian language became extinct, or even whether it has a living descendant. The initial Roman conquest of part of Dacia did not put an end to the language, as Free Dacian tribes such as the Carpi may have continued to speak Dacian in Moldavia and adjacent regions as late as the 6th or 7th century AD, still capable of leaving some influences in the forming Slavic languages.

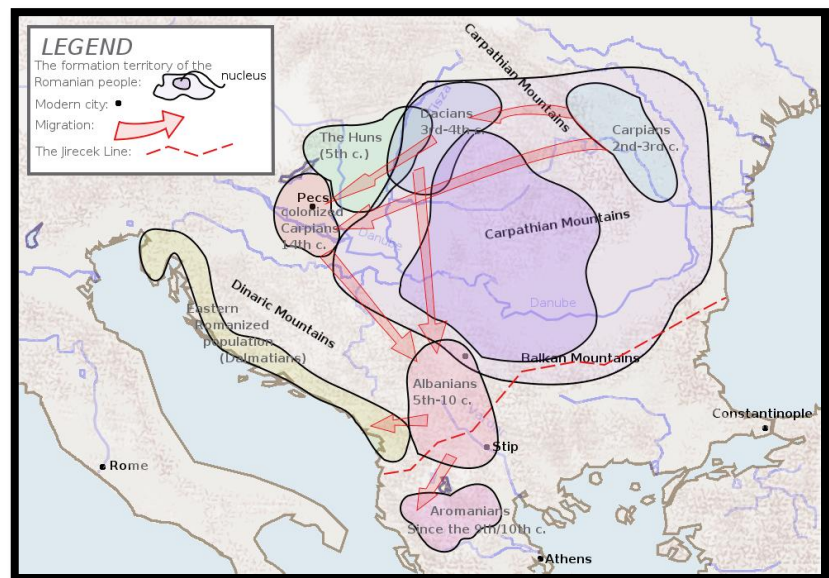


Figure 47. Theoretical scenario: the Albanians as a migrant Dacian people

- According to one hypothesis, a branch of Dacian continued as the Albanian language (Hasdeu, 1901);
- Another hypothesis considers Albanian to be a Daco-Moesian Dialect that split off from Dacian before 300 BC and that Dacian itself became extinct;

The argument for this early split (before 300 BC) is the following: inherited Albanian words (e.g. Alb. *motër* 'sister' < Late PIE *māter* 'mother') shows the transformation Late PIE *ā* > Alb. /o/, but all the Latin loans in Albanian having an /a:/ shows Lat. /a:/ > Alb. /a/. This indicates that the transformation P-Alb. /a:/ > P-Alb. /o/ happened and ended before the Roman arrival in the Balkans. On the other hand, Romanian substratum words shared with Albanian show a Romanian /a/ that correspond to an Albanian /o/ when both sounds source is an original common /a:/ (*mazāre/modhull* < **mādzula* 'pea'; *raṭā/rosë* < **rātja*: 'duck') indicating that when these words have had the same Common form in Pre-Romanian and Proto-Albanian the transformation P-Alb. /a:/ > P-Alb. /o/ had not started yet. The correlation between these two facts indicates that the split between Pre-Romanian (the Dacians that were later Romanized) and Proto-Albanian happened before the Roman arrival in the Balkans.

ILLYRIAN

The **Illyrian languages** are a group of Indo-European languages that were spoken in the western part of the Balkans in former times by ethnic groups identified as *Illyrians*: Delmatae, Pannoni, Illyrioi, Autariates, Taulanti. The Illyrian languages are generally, but not unanimously, reckoned as centum dialects.

Some sound-changes and other language features are deduced from what remains of the Illyrian languages, but because no writings in Illyrian are known, there is not sufficient evidence to clarify its place within the Indo-European language family aside from its probable centum nature. Because of the uncertainty, most sources provisionally place Illyrian on its own branch of Indo-European, though its relation to other languages, ancient and modern, continues to be studied and debated.

Today, the main source of authoritative information about the Illyrian language consists of a handful of Illyrian words cited in classical sources, and numerous examples of Illyrian anthroponyms, ethnonyms, toponyms and hydronyms.

A grouping of Illyrian with the Messapian language has been proposed for about a century, but remains an unproven hypothesis. The theory is based on classical sources, archaeology, as well as onomastic considerations. Messapian material culture bears a number of similarities to Illyrian material culture. Some Messapian anthroponyms have close Illyrian equivalents.

A relation to the Venetic language and Liburnian language, once spoken in northeastern Italy and Liburnia respectively, is also proposed.

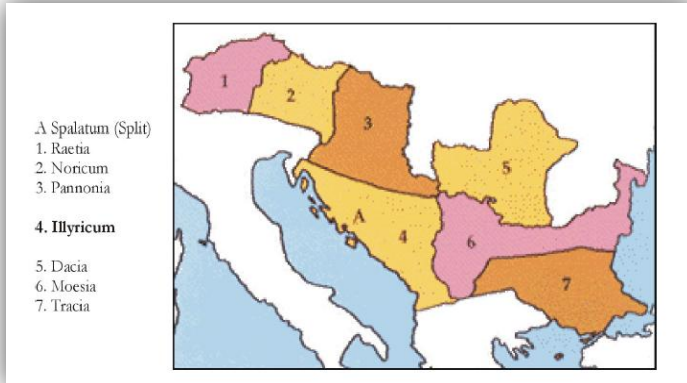


Figure 48. Territories where the different Paleo-Balkan languages were spoken.

language is widely but not unanimously accepted, although it is generally admitted that from what remains of the language, centum examples appear to greatly outnumber Satem examples. One of the few Satem examples in Illyrian appears to be *Osseriates*, probably from PIE **eghero-*, “lake”. Only a few Illyrian items have been linked to Albanian, and these remain tentative or inconclusive for the purpose of determining a close relation.

Only a few Illyrian words are cited in Classical sources by Roman or Greek writers, but these glosses, provided with translations, provide a core vocabulary. Only four identified with an ethnonym *Illyrii* or *Illurioi*; others must be identified by indirect means:

- brisa, “husk of grapes”; cf. Alb. *bërsi*.
- mantía, “bramble bush”; cf. Alb. (Tosk) *mën* “mulberry bush”, (Gheg) *mandë*.
- oseriates, “lakes”; akin to O.C.S. *ozero* (Sr.-Cr. *jezero*), Lith. *ẽžeras*, O.Pruss. *assaran*, Gk. *Akéroun* “river in the underworld”.
- rhinos, “fog, cloud”; cf. O.Alb. *ren*, mod. Alb. *re* “cloud”.
- sabaia, sabaïum, sabaius, “a type of beer”; akin to Eng *sap*, Lat. *sapere* “to taste”, Skr. *sabar* “sap, juice, nektar”, Av. *višāpa* “having poisonous juices”, Arm. *ham*, Greek *apalós* “tender, delicate”, O.C.S. *sveptŭ* “bee’s honey”.
- Lat. *sibina*, *sibyna*, *sybina*; Gk. *σιβυνη*, *σιβυνης*, *συβινη*, *ζιβυνη*: “a hunting spear”, “a spear”, “pike”; an Illyrian word according to Festius, citing Ennius; is compared to Gk. *συβηνη*, “flute case”, found in Aristophanes’ *Thesmophoriazusai*; the word appears in the context of a barbarian speaking. Akin to Persian *zôpîn*, Armenian *savîn* “spit”.
- tertigio, “merchant”; O.C.S. *trǫgŭ* (Sr.-Cr. *trg*), Lith. *tirgus* (Alb. *treg* “market” is a borrowing from archaic Slavic **trŭgŭ*)

A grouping of Illyrian with the Thracian and Dacian language in a “Thraco-Illyrian” group or branch, an idea popular in the first half of the 20th century, is now generally rejected due to a lack of sustaining evidence, and due to what may be evidence to the contrary.

A hypothesis that the modern Albanian language is a surviving Illyrian language remains very controversial among linguists. The identification of Illyrian as a centum

Some additional words have been extracted from toponyms, hydronyms, anthroponyms, etc.:

- *lougeon*, “a pool”; cf. Alb. *lag* “to wet, soak, bathe, wash” (< PA **lauga*), *lëgatë* “pool” (< PA. **leugatâ*), *lakshte* “dew” (< PA **laugista*); akin to Lith. *liūgas* “marsh”, O. Sla. *lužá* “pool”
- *teuta* < from the Illyrian personal name *Teuta* < PIE **teuta-*, “people”
- *Bosona*, “running water” (Possible origin of the name “Bosnia”, *Bosna* in Bosnian)

PAIONIAN

The **Paionian language** is the poorly attested language of the ancient Paionians, whose kingdom once stretched north of Macedon into Dardania and in earlier times into southwestern Thrace.

Several Paionian words are known from classical sources:

- *monapos*, *monaipos*, a wild bull.
- *tilôn*, a species of fish once found in Lake Prasias (Republic of Macedonia).
- *paprax*, a species of fish once found in Lake Prasias; masc. acc. pl. *paprakas*,

A number of anthroponyms (some known only from Paionian coinage) are attested, several toponyms (Bylazora, Astibos) and a few theonyms (*Dryalus*, *Dyalus*, the Paionian Dionysus), as well as:

- *Pontos*, affluent of the Strumica River, perhaps from **ponktos*, “wet” (cf. Ger. *feucht*, “wet”);
- *Stoboi* (nowadays Gradsko), name of a city, from **stob(h)* (cf. O.Pruss. *stabis* “rock”, O.C.S. *stoboru*, “pillar”, O.Eng. *stapol*, “post”, O.Gk. *stobos*, “scolding, bad language”);
- *Dóberos*, other Paionian city, from **dheubh-* “deep” (cf. Lith. *dubùs*, Eng. *deep*);
- *Agrianes*, name of a tribe, from **agro-* “field” (cf. Lat. *ager*, Gk. *agros*, Eng. *acre*).

Classical sources usually considered the Paionians distinct from Thracians or Illyrians, comprising their own ethnicity and language. Athenaeus seemingly connected the Paionian tongue to the Mysian language, itself barely attested. If correct, this could mean that Paionian was an Anatolian language.

On the other hand, the Paionians were sometimes regarded as descendants of Phrygians, which may put Paionian on the same linguistic branch as the Phrygian language.

Modern linguists are uncertain on the classification of Paionian, due to the extreme scarcity of materials we have on this language. However, it seems that Paionian was an independent IE dialect. It shows a/o distinctiveness and does not appear to have undergone Satemization. The Indo-European voiced aspirates **bh**, **dh**, etc., became plain voiced consonants, /b/, /d/, etc., just like in Illyrian, Thracian, Macedonian and Phrygian (but unlike Greek).

ANCIENT MACEDONIAN

The **Ancient Macedonian** language was the tongue of the Ancient Macedonians. It was spoken in Macedon during the 1st millennium BC. Marginalized from the 5th century BC, it was gradually replaced by the common Greek dialect of the Hellenistic Era. It was probably spoken predominantly in the inland regions away from the coast. It is as yet undetermined whether the language was a dialect of Greek, a sibling language to Greek, or an Indo-European language which is a close cousin to Greek and also related to Thracian and Phrygian languages.

Knowledge of the language is very limited because there are no surviving texts that are indisputably written in the language, though a body of authentic Macedonian words has been assembled from ancient sources, mainly from coin inscriptions, and from the 5th century lexicon of Hesychius of Alexandria, amounting to about 150 words and 200 proper names. Most of these are confidently identifiable as Greek, but some of them are not easily reconciled with standard Greek phonology. The 6,000 surviving Macedonian inscriptions are in the Greek Attic dialect.

The Pella curse tablet, a text written in a distinct Doric Greek idiom, found in Pella in 1986, dated to between mid to early 4th century BC, has been forwarded as an argument that the Ancient Macedonian language was a dialect of North-Western Greek. Before the discovery it was proposed that the Macedonian dialect was an early form of Greek, spoken alongside Doric proper at that time.

NOTE. Olivier Masson thinks that “in contrast with earlier views which made of it an Aeolic dialect (O.Hoffmann compared Thessalian) we must by now think of a link with North-West Greek (Locrian, Aetolian, Phocidian, Epirote). This view is supported by the recent discovery at Pella of a curse tablet which may well be the first ‘Macedonian’ text attested (...); the text includes an adverb “opoka” which is not Thessalian.” Also, James L. O’Neil states that the “curse tablet from Pella shows word forms which are clearly Doric, but a different form of Doric from any of the west Greek dialects of areas adjoining Macedon. Three other, very brief, fourth century inscriptions are also indubitably Doric. These show that a Doric dialect was spoken in Macedon, as we would expect from the West Greek forms of Greek names found in Macedon. And yet later Macedonian inscriptions are in Koine avoiding both Doric forms and the Macedonian voicing of consonants. The native Macedonian dialect had become unsuitable for written documents.”

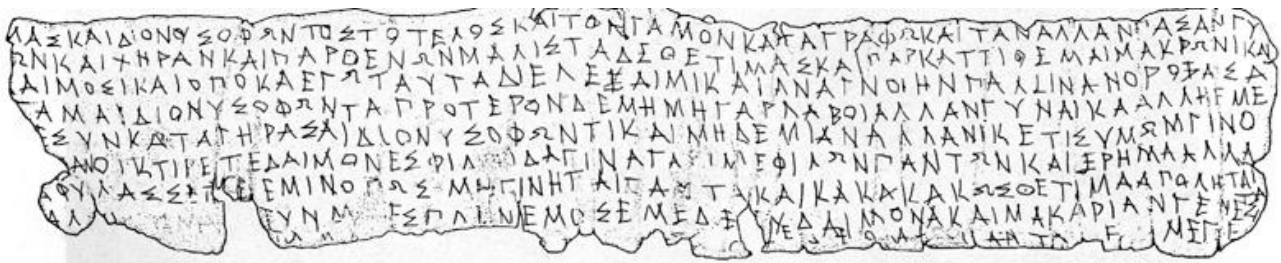


Figure 49. The Pella katadesmos, is a *katadesmos* (a curse, or magic spell) inscribed on a lead scroll, probably dating to between 380 and 350 BC. It was found in Pella in 1986.

From the few words that survive, a notable sound-law may be ascertained, that PIE voiced aspirates appear as voiced stops, written **β, γ, δ** in contrast to Greek dialects, which unvoiced them to **φ, χ, θ**.

- Mac. δανός *danós* ('death', from PIE **d^henh₂-* 'to leave'), compare Attic θάνατος *thánatos*.
- Mac. ἀβροῦτες *abrou̯tes* or ἀβροῦτες *abrou̯wes* as opposed to Attic ὀφρῦς *ophrús* for 'eyebrows'.
- Mac. Βερενίκη *Bereníkē* versus Attic Φερενίκη *Phereníkē*, 'bearing victory' *ἄδραια *adraia* ('bright weather'), compare Attic αἰθρία *aithría*, from PIE **h₂aid^h-*.
- βάσκιαι *báskioi* ('fascies'), from PIE **b^hasko*.
- According to Hdt. 7.73 (ca. 440 BC), the Macedonians claimed that the *Phryges* were called *Brygoi* before they migrated from Thrace to Anatolia ca. 1200 BC.
- μάγειρος *mágeiros* ('butcher') was a loan from Doric into Attic. Vittore Pisani has suggested an ultimately Macedonian origin, cognate to μάχαιρα *mákhaira* ('knife', <PIE **mag^h-*, 'to fight').

The same treatment is known from other Paleo-Balkan languages, e.g. Phrygian *bekos*, "bread", Illyrian *bagaron*, "warm", but Gk. φῶγω (*phōgō*), "roast", all from IE **b^heh₃g-*. Since these languages are all known via the Greek alphabet, which has no signs for voiced aspirates, it is unclear whether deaspiration had really taken place, or whether β, δ, γ were just picked as the closest matches to express voiced aspirates.

If γοτάν (*gotán*), "pig", is related to IE **g^wou* ('cattle'), this would indicate that the labiovelars were either intact, or merged with the velars, unlike the usual Gk. βοῦς (*boús*). Such deviations, however, are not unknown in Greek dialects; compare Doric Spartan γλεπ- (*glep-*) for common Greek βλεπ- (*blep-*), as well as Doric γλάχων (*gláchōn*) and Ionic γλήχων (*glēchōn*) for common Greek βλήχων (*blēchōn*).

A number of examples suggest that voiced velar stops were devoiced, especially word-initially; as in κánaδοι (*kánadoi*, from PIE **genu-*), "jaws"; κόμβους (*kómbous*, from PIE **gombh-*), "molars"; within words, as in ἀρκόν (*arkón*) vs. Attic ἀργός (*argós*); the Macedonian toponym *Akesamenai*, from the Pierian name *Akesamenos* – if *Akesa-* is cognate to Greek *agassomai*, *agamai*, "to astonish"; cf. the Thracian name *Agassamenos*.

In Aristophanes' *The Birds*, the form κεβλήπυρις (*keblēpyris*), "red-cap bird", shows a voiced stop instead of a standard Greek unvoiced aspirate, i.e. Macedonian κεβ(α)λή (*kebalē*) vs. Greek κεφαλή (*kephalē*), "head".

1.7.4. ANATOLIAN LANGUAGES

The **Anatolian languages** are a group of extinct Indo-European languages, which were spoken in Asia Minor, the best attested of them being the Hittite language.

The Anatolian branch is generally considered the earliest to split off the Proto-Indo-European language, from a stage referred to either as Middle PIE (also IE II) or “Indo-Hittite”, typically a date in the mid-4th millennium BC is assumed. In a Kurgan framework, there are two possibilities of how early Anatolian speakers could have reached Anatolia: from the north via the Caucasus, and from the west, via the Balkans.

Attested dialects of the Anatolian branch are:

- **Hittite** (*nesili*), attested from ca. 1900 BC to 1100 BC, official language of the Hittite Empire.
- **Luwian** (*luwili*), close relative of Hittite spoken in adjoining regions, sometimes under Hittite control .
 - Cuneiform Luwian, glosses and short passages in Hittite texts written in Cuneiform script.
 - Hieroglyphic Luwian, written in Anatolian hieroglyphs on seals and in rock inscriptions.
- **Palaic**, spoken in north-central Anatolia, extinct around the 13th century BC, known only fragmentarily from quoted prayers in Hittite texts.
- **Lycian**, spoken in Lycia in the Iron Age, a descendant of Luwian, extinct in ca. the 1st century BC, fragmentary language.
- **Lydian**, spoken in Lydia, extinct in ca. the 1st century BC, fragmentary.
- **Carian**, spoken in Caria, fragmentarily attested from graffiti by Carian mercenaries in Egypt from ca. the 7th century BC, extinct ca. in the 3rd century BC.
- **Pisidian** and **Sidetic** (Pamphylian), fragmentary.
- **Milyan**, known from a single inscription.

There were likely other languages of the family that have left no written records, such as the languages of Mysia, Cappadocia and Paphlagonia.



Figure 50. Maximal extent of the Hittite Empire ca. 1300 BC is shown in **dark color**, the Egyptian sphere of influence in **light color**. The approximate extent of the Hittite Old Kingdom under Hantili I (ca. 1590 BC) in **darkest**.

Anatolia was heavily Hellenized following the conquests of Alexander the Great, and it is generally thought that by the 1st century BC the native languages of the area were extinct.

Hittite proper is known from cuneiform tablets and inscriptions erected by the Hittite kings. The script known as “Hieroglyphic Hittite” has now been shown to have been used for writing the closely related Luwian language, rather than Hittite proper. The later languages Lycian and Lydian are also attested in Hittite territory. Palaic, also spoken in Hittite territory, is attested only in ritual texts quoted in Hittite documents.

In the Hittite and Luwian languages there are many loan words, particularly religious vocabulary, from the non-Indo-European Hurrian and Hattic languages. Hattic was the language of the Hattians, the local inhabitants of the land of Hatti before they were absorbed or displaced by the Hittite invasions. Sacred and magical Hittite texts were often written in Hattic, Hurrian, and Akkadian, even after Hittite became the norm for other writings.

The Hittite language has traditionally been stratified into Old Hittite (OH), Middle Hittite (MH) and New or Neo-Hittite (NH), corresponding to the Old, Middle and New Kingdoms of the Hittite Empire, ca. 1750–1500 BC, 1500–1430 BC and 1430–1180 BC, respectively. These stages are differentiated partly on linguistic and partly on paleographic grounds.

Hittite was written in an adapted form of Old Assyrian cuneiform orthography. Owing to the predominantly syllabic nature of the script, it is difficult to ascertain the precise phonetic qualities of a portion of the Hittite sound inventory.

Hittite preserves some very archaic features lost in other Indo-European languages. For example, Hittite has retained two of three laryngeals, word-initial h_2 and h_3 . These sounds, whose existence had been hypothesized by Ferdinand de Saussure on the basis of vowel quality in other Indo-European languages in 1879, were not preserved as separate sounds in any attested Indo-European language until the discovery of Hittite. In Hittite, this phoneme is written as *h*.

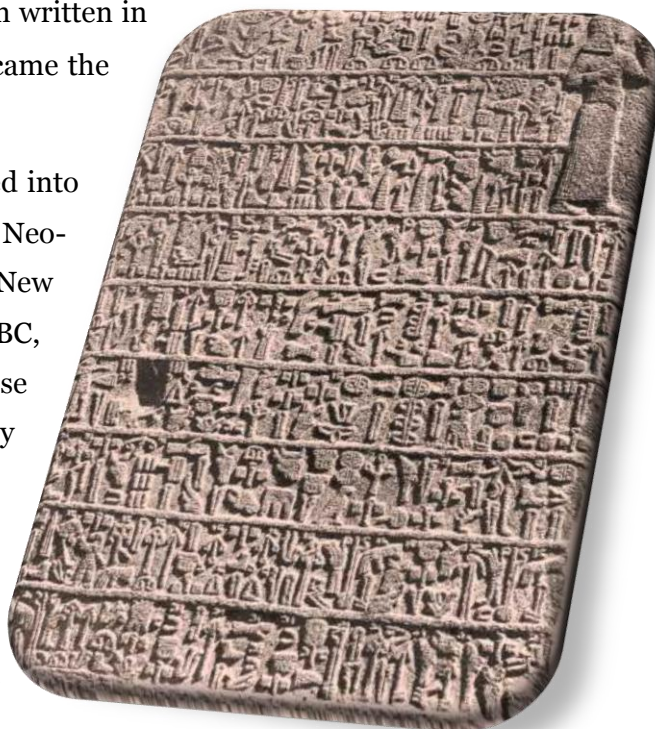


Figure 51. Hittite pictographic writing was directly derived from Old Assyrian cuneiform.

Hittite, as well as most other Anatolian languages, differs in this respect from any other Indo-European language, and the discovery of laryngeals in Hittite was a remarkable confirmation of Saussure's hypothesis.

The preservation of the laryngeals, and the lack of any evidence that Hittite shared grammatical features possessed by the other early Indo-European languages, has led some philologists to believe that the Anatolian languages split from the rest of Proto-Indo-European much earlier than the other divisions of the proto-language. In Indo-European linguistics, the term **Indo-Hittite** (also **Indo-Anatolian**) refers to the hypothesis that the Anatolian languages may have split off the Proto-Indo-European language considerably earlier than the separation of the remaining Indo-European languages. The majority of scholars continue to reconstruct a single Proto-Indo-European, but all believe that Anatolian was the first branch of Indo-European to leave the fold.

NOTE. The term is somewhat imprecise, as the prefix *Indo-* does not refer to the Indo-Aryan branch in particular, but is iconic for *Indo-European* (as in *Indo-Uralic*), and the *-Hittite* part refers to the Anatolian language family as a whole.

As the oldest attested Indo-European languages, Hittite is interesting largely because it *lacks* several grammatical features exhibited by other “old” Indo-European languages such as Sanskrit and Greek.

The Hittite nominal system consists of the following cases: Nominative, Vocative, Accusative, Genitive, Allative, Dative-Locative, Instrumental and Ablative. However, the recorded history attests to fewer cases in the plural than in the singular, and later stages of the language indicate a loss of certain cases in the singular as well. It has two grammatical genders, common and neuter, and two grammatical numbers, singular and plural.

Hittite verbs are inflected according to two general verbal classes, the *mi*-conjugation and the *hi*-conjugation. There are two voices (active and mediopassive), two moods (indicative and imperative), and two tenses (present and preterite). Additionally, the verbal system displays two infinitive forms, one verbal substantive, a supine, and a participle. Rose (2006) lists 132 *hi*-verbs and interprets the *hi/mi* oppositions as vestiges of a system of grammatical voice, i.e. “centripetal voice” vs. “centrifugal voice”.



Figure 52. Broken door jamb inscribed in raised Hittite hieroglyphs, c. 900 BC; in the British Museum.

1.8. ‘EUROPAIOM’ OR ‘SINDHUEUROPAIOM’

1.8.1. Modern Indo-European, for which we use the neutral name **Dínghūs** (also dialectally extended in **-ā**, Ita.-Cel., Ger. **dínghwā**), “*the language*”, is therefore a set of grammatical rules – including its writing system, noun declension, verbal conjugation and syntax –, designed to systematize the reconstructed Late Proto-Indo-European language, to adapt it to modern communication needs. As PIE was spoken by a prehistoric society, no genuine sample texts are available, and thus comparative linguistics – in spite of its 200 years’ history – is not in the position to reconstruct exactly their formal language (the one used by learned people), but only approximately how the spoken, vulgar language was like, i.e. the language that evolved into the different attested Indo-European dialects and languages.

NOTE. Reconstructed languages like Modern Hebrew, Modern Cornish, Modern Coptic or Modern Indo-European may be revived in their communities without being as *easy*, as *logical*, as *neutral* or as *philosophical* as the million artificial languages that exist today, and whose main aim is to be supposedly ‘*better*’, or ‘*easier*’, or ‘*more neutral*’ than other artificial or natural languages they want to substitute. Whatever the sociological, psychological, political or practical reasons behind the success of such ‘*difficult*’ and ‘*non-neutral*’ languages instead of ‘universal’ ones, what is certain is that if somebody learns Hebrew, Cornish, Coptic or Indo-European (or Latin, German, Swahili, Chinese, etc.) whatever the changes in the morphology, syntax or vocabulary that could follow (because of, say, ‘*better*’ or ‘*purser*’ or ‘*easier*’ language systems recommended by their language regulators), the language learnt will still be the same, and the effort made won’t be lost in any possible case.

1.8.2. We deemed it worth it to use the Proto-Indo-European reconstruction for the revival of a complete modern language system, because of the obvious need of a common language within the EU, to substitute the current deficient linguistic policy. This language system, called *European* or *European language* (**Eurōpáiom**), is mainly based on the features of the European or northwestern dialects, whose speakers – as we have already seen – remained in loose contact for some centuries after the first PIE migrations, and have influenced each other in the last millenia within the European subcontinent.

NOTE. As Indo-Europeanist López-Menchero puts it, “there are three Indo-European languages which must be clearly distinguished: 1) The Proto-Indo-European language, spoken by a prehistoric people, the so-called Proto-Indo-Europeans, some millennia ago; 2) The reconstructed Proto-Indo-European language, which is that being reconstructed by IE scholars using the linguistic, archaeological and historical data available, and which is imperfect by nature, based on more or less certain hypothesis and schools; and 3) The Modern Indo-European language system(s) which, being based on the later, and trying to come near to the former, is neither one nor the other, but a modern language systematized and used in the modern word”. We should add that, unlike artificial languages, Indo-European may not be substituted by different languages, although – unlike already systematized languages like Classic Latin or English – it could be changed by other dialectal, older or newer versions of it, as e.g. ‘*Graeco-Aryan*’, i.e. a version mainly based on the Southern Dialect, or ‘*Indo-Hittite*’, a version using laryngeals, not separating feminines from the animates, and so on.

NOTE 2. A Modern PIE is probably the best option as an International Auxiliary Language too, because a) French, German, Spanish, and other natural and artificial languages proposed to substitute English dominance, are only supported by their small cultural or social communities, while the communities of IE speakers make up the majority of the world's population, being thus the most 'democratic' choice for a language spoken within international organizations and between the different existing nations; and b) only a major change in the political arena could make a language different than English succeed as a spoken IAL; if the European Union makes Modern Indo-European its national language, it would be worth it for the rest of the world to learn it as second language and use it as the international language instead of English.

1.8.5. Words to complete the MIE vocabulary (in case that no common PIE form is found) are to be taken from present-day IE languages. Loan words – from Greek and Latin, like *philosophy*, *hypothesis*, *aqueduct*, etc. –, as well as modern Indo-European borrowings – from English, like *software*, from French, like *ambassador*, from Spanish, like *armadillo*, from German, like *Kindergarten*, from Italian, like *casino*, from Russian, like *icon*, from Hindi, like *pajamas*, etc. –, should be used in a *pure IE form* when possible. They are all Indo-European dialectal words, whose original meaning is easily understood if translated; as, e.g. Greek loan *photo* could appear in Modern Indo-European either as **phótos** ['p'o-tos] or ['fo-tos], a loan word, or as **bháwtos** ['b'hau-tos], a loan translation of Gk. "bright", IE **bháuesos**, from genitive **bhauesós**, from PIE verb **bhā**, to *shine*, which gives in Greek *phosphorus* and *phot*. The second, translated word, should be preferred.² See §2.9.4, point 4.

1.8.6. A comparison with Modern Hebrew seems adequate, as it is one successful precedent of an old, reconstructed language becoming the living language of a whole nation.

HEBREW REVIVAL	INDO-EUROPEAN REVIVAL
<i>ca. 3000 BC: Proto-Aramaic, Proto-Ugaritic, and other Canaanite languages spoken.</i>	<i>ca. 3000 BC: Middle Proto-Indo-European dialects, Pre-IE III and Pre-Proto-Anatolia, spoken. ca. 2,500 BC: Late PIE spoken.</i>
<i>ca. 1000 BC: The first written evidence of distinctive Hebrew, the Gezer calendar.</i>	<i>ca. 1600 BC: first written evidence, Hittite and Luwian tablets (Anatolian). ca. 1500 BC: Linear B tablets in Mycenaean Greek.</i>
<i>Orally transmitted Tanakh, composed between 1000 and 500 BC.</i>	<i>Orally transmitted Rigveda, in Vedic Sanskrit, (similar to older Indo-Iranian), composed in parts, from 1500 to 500 BC. Orally transmitted Zoroastrian works in Avestan (Iranian dialect), from 1000 to 700 BC. Homeric works dated from ca. 700 BC. Italic inscriptions, 700-500 BC.</i>
<i>Destruction of Jerusalem by the Babylonians under Nebuchadnezzar II, in 586 BC. The Hebrew language is then replaced by Aramaic in Israel under the Persian Empire. Destruction of Jerusalem and Expulsion of Jews by the Romans in 70 AD.</i>	<i>Italics, Celts, Germanics, Baltics and Slavics are organized mainly in tribes and clans. Expansion of the great Old Civilizations, such as the Persians, the Greeks and the Romans. Behistun Inscription, Celtic inscriptions ca 500 BC; Negau Helmet in Germanic, ca. 200 BC.</i>

<i>70-1950 AD. Jews in the Diaspora develop different dialects with strong Hebrew influence, with basis mainly on Indo-European (Yiddish, Judeo-Spanish, Judeo-Italian, etc.), as well as Semitic languages (Judeo-Aramaic, Judeo-Arab, etc.)</i>	<i>Expansion of the renowned Antique, Mediaeval and Modern IE civilizations, such as the Byzantines, the Franks, the Persians, the Spanish and Portuguese, the Polish and Lithuanians, the French, the Austro-Hungarians and Germans and the English among others.</i>
<i>1880 AD. Eliezer Ben-Yehuda begins the construction of a modern Hebrew language for Israel based on Old Hebrew.</i>	<i>1820 AD. Bopp begins the reconstruction of the common ancestor of the Indo-European languages, the Proto-Indo-European language.</i>
<i>19th century. Jews speaking different Indo-European and Semitic languages settle in Israel. They use different <i>linguae francae</i> to communicate, such as Turkish, Arab, French or English.</i>	<i>1949-1992. European countries form an International European Community, the EEC. 1992-2007: A Supranational entity, the European Union, substitutes the EEC. There are 23+3 official languages</i>
<i>1922 AD. Hebrew is named official language of Palestine, along with English and Arabic. From that moment on, modern Hebrew becomes more and more the official national language of the Israelis. The settlers' native languages are still spoken within their communities and families.</i>	<i>Present. New steps are made to develop a national entity, a confederation- or federation-like state. The EU Constitution and the linguistic policy are two of the most important issues to be solved before that common goal can be achieved. More than 97% of the EU populations has an Indo-European language as mother tongue.</i>

NOTE. Even though it is clear that our proposal is different from the Hebrew language revival, we think that: a) Where Jews had only some formal writings, with limited vocabulary, of a language already dead five centuries before they were expelled from Israel, Indo-European has hundreds of living dialects and other very old dead dialects attested. Thus, even if we had tablets of PIE written in some dialectal predominant formal IE language (say, from pre-Proto-Indo-Iranian), the current PIE reconstruction would probably still be used as the main source for PIE revival today. b) The common culture and religion was possibly the basis for the Hebrew language revival in Israel. Proto-Indo-European, whilst the mother tongue of some prehistoric tribe with a common culture and religion, spread into different peoples, with different cultures and religions. There was never a concept of “Indo-European community” after the migrations. But today Indo-European is the language spoken by the majority of the population – in the world and especially within Europe –, and it is therefore possible to use it as a *natural* and *culturally* (also “*religiously*”) neutral language, what may be a significant advantage of IE.

1.7.7. The noun **Eurōpaíos** comes from adjective **eurōpaiós**, from special genitive **europai** of Old Greek Εὐρώπη (*Eurṓpē*), Εὐρώπα (*Eurṓpā*), both forms alternating already in the oldest Greek, and both coming from the same PIE feminine ending **ā** (see § 4.9.3). The Greek ending **-ai-o-** (see § 4.7.8 for more on this special genitive in **-ai**) turns into Latin **-ae-u-**, and so *Europaeus*. The forms **Eurṓpā** and **Eurōpaiós** are, then, the ‘correct’ ones in MIE, as they are the original Classic forms – other dialectal variants, as **Eurōpās**, **Eurōpaís**, **Eurōpaikós**, **Eurōpaiskós**, etc. could be also used.

NOTE 1. For Homer, *Eurṓpē* was a mythological queen of Crete – abducted by Zeus in bull form when still a Phoenician princess –, and not a geographical designation. Later *Europa* stood for mainland Greece, and by 500 B.C. its meaning had been extended to lands to the north. The name *Europe* is possibly derived from the Greek words *eurús* (*eurús*, “broad”, from IE **h₁urhu-*) and *ops* (*ops*, “face”, from IE **h₃ekw-*), thus maybe

reconstructable as MIE **Ūróqā** – *broad* having been an epithet of Earth in PIE religion. Others suggest it is based on a Semitic word cognate with Akkadian *erebu*, “sunset” (cf. Arabic *maghreb*, Hebrew *ma'ariv*), as from the Middle Eastern vantage point, the sun does set over Europe. Likewise, *Asia* is sometimes thought to have derived from a Semitic word such as the Akkadian *asu*, meaning “sunrise”, and is the land to the east from a Middle Eastern perspective, thus maybe MIE **Eróbā**. In Greek mythology Ἔρεβος (*Erebos*, “deep blackness/darkness or shadow”) was the son of Chaos, the personification of darkness and shadow, which filled in all the corners and crannies of the world. The word is probably from IE **h₁regʷos* (cf. O.N. *rækkr*, Goth. *riqis*, Skr. *rajani*, Toch. *orkām*), although possibly also a loan from Semitic, cf. Hebrew *erebh* and Akkadian *erebu*, etc.

NOTE 2. ‘Europe’ is a common evolution of Latin a-endings in French; as in ‘Amerique’ for *America*, ‘Belgique’ for *Belgica*, ‘Italie’ for *Italia*, etc. Eng. *Europe* is thus a French loan word, as may be seen from the other continents’ names: *Asia* (not **Asy*), *Africa* (not **Afrik*), *Australia* (not **Australy*), and *America* (not **Amerik*).

NOTE 3. Only Modern Greek maintains the form Ευρώπη (*Európi*) for the subcontinent, but still with adjective ευρωπαϊκό (*europaikó*), with the same old irregular a-declension and IE ethnic ending **-iko-**. In Latin there were two forms: *Europa*, *Europaeus*, and lesser used *Europe*, *Europensis*. The later is usually seen in scientific terms.

NOTE 4. For adj. “European”, compare derivatives from O.Gk. *eurōpai-ós* (< IE **eurōp-ai-ós**), also in Lat. *europaé-us* -> M.Lat. *europé-us*, in turn giving It., Spa. *uropeo*, Pt., Cat. *uropeu*; from Late Latin base *europé-* (< IE **eurōp-ái-**) are extended **europe-is*, as Du. *europes*; from extended **europe-anos* are Rom. *europene*, or Fr. *européen* (into Eng. *european*); extended **europe-iskos* gives common Germanic and Slavic forms (cf. Ger. *Europäisch*, Fris. *europesk*, Sca. *uropeisk*, Pl. *uropejski*, common Sla. *evropsk-*, etc.); other extended forms are Ir. *Eorpai-gh*, Lith. *europo-s*, Ltv. *eiropa-s*, etc. For European as a noun, compare, from **europé-anos*, Du., Fris. *europaan*, from **europé-eros*, Ger. *Europäer*, from ethnic **-ikos*, cf. Sla. *evropejk-*, Mod.Gk. *europai-kó*, etc.

The regular genitive of the word **Eurōpā** in Modern Indo-European is **Eurōpās**, following the first declension. The name of the European language system is **Eurōpáiom**, inanimate, because in the oldest IE dialects attested, those which had an independent name for languages used the neuter, cf. Gk. n.pl. ελληνικά (*ellēniká*), Skr. n.sg. संस्कृतम् (*saṃskṛtam*), also in Tacitus Lat. *uōcābulum latīnum*.

In other languages, however, the language name is an adjective which defines the noun “language”, and therefore its gender follows the general rule of concordance; cf. Lat. f. *latīna lingua*, or the Slavic examples³; hence MIE **eurōpaiá dñghūs** or **eurōpaiá dñghwā**, *European language*.

1.7.8. **Sindhueurōpáiom** (n.) means *Indo-European (language)*. The term comes from Greek Ἰνδός (*^hIndos*), *Indus river*, from Old Persian *Hinduš* - listed as a conquered territory by Darius I in the Persepolis terrace inscription.

NOTE. The Persian term (with an aspirated initial [s]) is cognate to *Sindhu*, the Sanskrit name of the Indus river, but also meaning *river* generically in Indo-Aryan (cf. O.Ind. *Saptasindhu*, “[region of the] seven rivers”). The Persians, using the word *Hindu* for *Sindhu*, referred to the people who lived near the Sindhu River as *Hindus*, and their religion later became known as *Hinduism*. The words for their language and region, *Hindī* or *Hindustanī* and *Hindustan*, come from the words *Hindu* and *Hindustan*, “India” or “Indian region” (referring to the Indian subcontinent as a whole, see **stā**) and the adjectival suffix *-ī*, meaning therefore originally “Indian”.