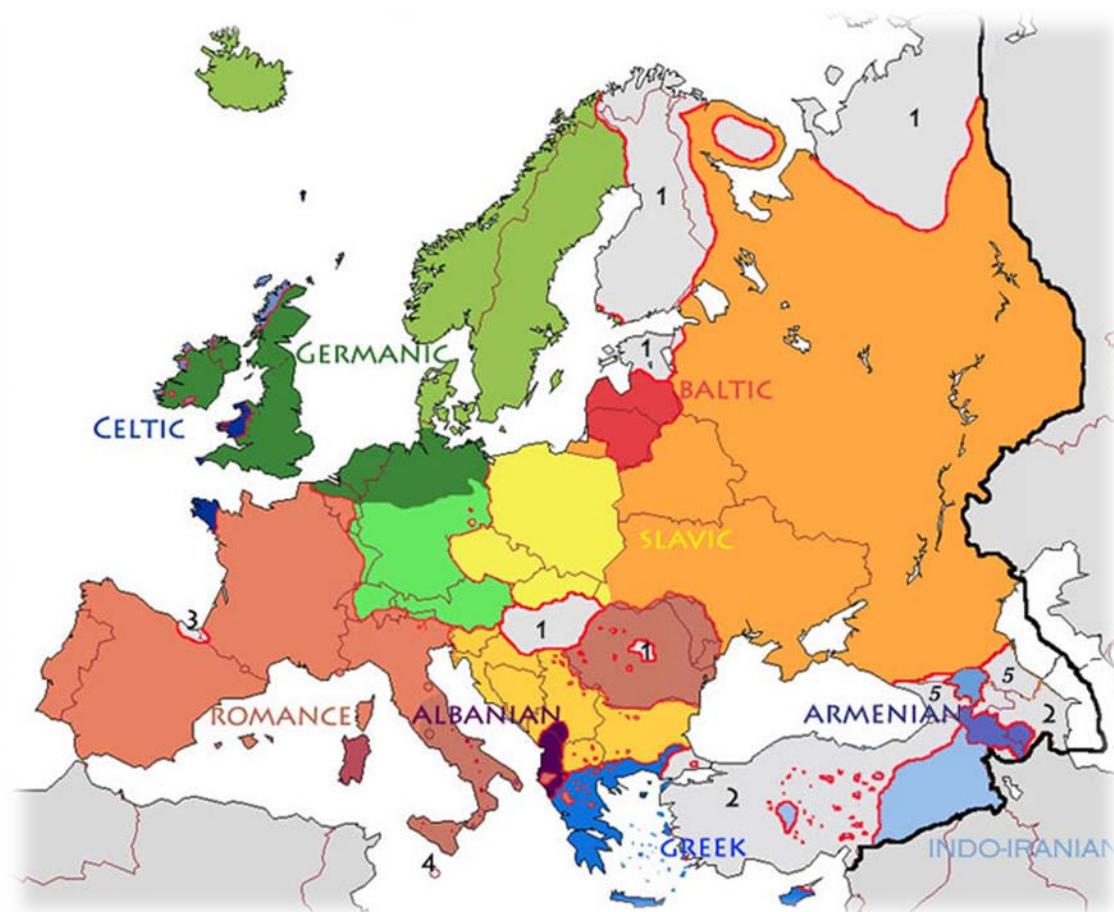


1.7. INDO-EUROPEAN DIALECTS



Languages of Europe. 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 PIE TO MODERN ENGLISH

The so-called *Schleicher's fable* is a poem composed in PIE, published by August Schleicher in 1868, originally named "*The Sheep and the Horses*". It is written here in the different reconstructible IE dialects for comparison.

The immediate parent dialect of each proto-language is enclosed in parentheses.

1. Introduction

A Common PIE version (ca. 3500 BC?): H_3owis $h_1ekw\ddot{o}s$ - k^{we} . • H_3owis , k^{wesjo} $w\grave{h}_1neh_2$ ne h_{est} , • $h_1ekwoms$ $spek\acute{e}t$, • h_1oinom $g^{w\grave{r}}h_3\acute{u}m$ $wog^{h}om$ $w\acute{e}g^{h}ontm_3$, • h_1oinom - k^{we} $megeh_2m$ b^h orom, • h_1oinom - k^{we} $d^h h_1g^h$ omonm₃ h_1oh_1ku $b^h\acute{e}rontm_3$. • H_3owis nu $h_1\acute{e}kwob^hjos$ $weuk^{w\acute{e}t}$: • “ $Kr\grave{d}$ $h_2\acute{e}g^h$ nutoi h_1moi ”, • $h_1ekwoms$ $h_2\acute{e}gontm_3$ $wih_1r\acute{o}m$ $w\acute{id}nt\acute{e}i$ ”. • $H_1ekw\ddot{o}s$ tu $weuk^{w\acute{o}nt}$: “ $Klud^h\acute{i}$, $h_3owi!$ ” • $kr\grave{d}$ $h_2\acute{e}g^h$ nutoi $\eta sm\acute{e}i$ $w\acute{id}ntb^hjos$: • h_2ner , $potis$, h_3owjom - r_3 $w\grave{h}_1neh_2m_3$ • $sweb^hei$ $g^{wh}ermom$ $westrom$ $k^{w\grave{r}}neuti$ ”. • H_3owjom - k^{we} $w\grave{h}hneh_2$ ne h_1esti . • Tod $k\acute{e}kluwos$ h_3owis h_2egrom $b^hug\acute{e}t$.

Common Anatolian (PAN), 2500 BC	Europe’s IE (IE IIIb), ca. 2500 BC
Howis ekwōs-k ^{we} .	Owis ekwōs-k ^{we} .
Howis, k ^{wesjo} w ^h neh ne est,	Owis, k ^{wesjo} w ^h nā ne est,
ekwons spekét,	ekwons spekét,
oikom g ^{wr} rúm wogom wégontm ₃ ,	oinom g ^{wr} rúm wog ^h om wég ^h ontm ₃ ,
oikom-k ^{we} megeh ₂ m borom,	oinom-k ^{we} megām b ^h orom,
oikom-k ^{we} dgomonm ₃ oku bérontm ₃ .	oinom-k ^{we} d ^h g ^h omonm ₃ ōkú b ^h érontm ₃ .
Howis nu ékwobos wūkwét:	Owis nu ékwob ^h os weuk ^{wét} :
“Krdi xégnutor moi,	“Krdi ág ^h nutoi moi,
ekwons xégonm ₃ wiróm wídn ^{te} ”.	ekwons ágontm ₃ wíróm wídn ^{te} i”.
Ekwōs tu weuk ^{wónt} : “Kludí, howi!	Ekwōs tu weuk ^{wónt} : “Klud ^h i, owi!
krdi hegnutor nsme wídn ^t bos:	krdi ág ^h nutoi nsmeí wídn ^t b ^h jos:
hner, potis, howjom-r ₃ w ^h nehm	ner, potis, owjom-r ₃ w ^h nām
swebę g ^w ermom wéstrom k ^{wr} nūdi”.	seb ^h ei g ^{wh} ormom westrom k ^{wr} neuti”.
Howjom-k ^{we} w ^h neh ne esti.	Owjom-k ^{we} w ^h nā ne esti.
Tod kékluwos howis hegrom bugét.	Tod kékluwos owis agrom b ^h ugét.

Proto-Aryan (IE IIIa), ca. 2500 BC	Proto-Greek (IE IIIa), ca. 2500 BC
Awis akwās-ka.	Owis ekwoi-k ^{we} .
Awis, kasja wīrnā na āst,	Ówis, k ^w eho w ^h nā ne ēst,
akwans spaqát,	ekwons spekét,
aikam grúm wag ^h am wág ^h antm ₃ ,	oiwom k ^{wh} rúm wok ^h om wek ^h ontm ₃ ,
aikam-ka mag ^h am b ^h aram,	oiwom-k ^{we} megām p ^h orom,
aikam-ka d ^h g ^h ámanm ₃ āku b ^h arantm ₃ .	oiwom-k ^{we} k ^h t ^h ómonm ₃ ōku p ^h érontm ₃ .
Awis nu ákwab ^h jas áwaukat:	Ówis nu ékwop ^h os éweuk ^{wet} :
“Krdi ág ^h nutai mai,	“Krdi ák ^h nutoi moi,
akwans aqantam wīrām wídn ^{tai} ”.	ekwons ágontm ₃ wíróm wídn ^{te} i”.
Ákwās tu áwawkant: “Krud ^h i avi!	Ékwoi tu éwewek ^{wont} : “Klut ^h i, owi!
krđ ág ^h nutai nsmaí wídn ^t b ^h jas:	krđ ág ^h nutoi nsmeí wídn ^t p ^h os:
nar, patis, awjam-r ₃ w ^h nām	anér, potis, owjom-r ₃ w ^h nām
swab ^h i g ^h armam wastram krnauti”.	sep ^h ei k ^{wh} ermom westrom k ^{wr} neuti”.
Awjam-ka wīrnā na asti.	Owjom-k ^{we} w ^h nā ne esti.
Tat kákruwas awis aqram áb ^h ugat.	Tot kékluwos owis agrom ép ^h uget.

A GRAMMAR OF MODERN INDO-EUROPEAN

Proto-Celtic (ca. 1000 BC)	Proto-Italic (ca. 1000 BC)
Owis ek ^w oi-k ^w e.	Owis ekwoi-k ^w e.
Owis, k ^w esjo wlānā ne est,	Owis, k ^w esjo wlānā ne est,
ek ^w ōs spekét,	ekwōs spekét,
oinom barúm woxom wéxontam,	oinom grāwúm woxom wéxontem,
oinom-k ^w e megam borom,	oinom-k ^w e megam φorom,
oinom-k ^w e dxoniom āku berontam.	oinom-k ^w e xomonem ōku φerontem.
Owis nu ék ^w obos weuk ^w ét:	Owis nu ékwoφos weuk ^w ét:
“Kridi áxnutor mai,	“Kordi axnutor mei,
ek ^w ōs ágontom wíróm wídanti”.	ekwōs ágontom wíróm wídatei”.
Ek ^w oi tu wewk ^w ónt: “Kludi, owi!	Ekwoi tu wewk ^w ónt: “Kluþi, owi!
kridi áxnutor ansméi wídanþjos:	kordi axnutor ensméi wídateφos:
ner, φotis, owjom-ar wlānām	ner, potis, owjom-or wlānām
sebi g ^w ormom westrom k ^w arneuti”.	seφei g ^h ormom westrom k ^w orneuti”.
Owjom-k ^w e wlānā ne esti.	Owjom-k ^w e wlānā ne esti.
Tod kéklowos owis agrom bugét.	Tud kékluwos owis agrom φugít.

Pre-Proto-Germanic (ca. 1000 BC)	Proto-Balto-Slavic (ca. 1000 BC)
Awiz exwaz-x ^w e.	Awis ekwōs-ke.
Awiz, h ^w es wulnō ne est,	Awis, kesja wilnā ne est,
ehwanz spexét,	ekwas spekét,
ainan karún wagan wéganðun,	ainan grun waġan wégantun,
ainan-x ^w e mekon baran,	ainan-ke meġan baran,
ainan-x ^w e gúmanan āxu bérandūn.	ainan-ke ġumanan ōķu bérantun
Awiz nu éxwamaz weux ^w éd:	Awis nu ekwamas wjaukét:
“Hurti ágnuðai mei,	“Ķirdi ágnutei mei,
exwanz ákanðun werán wítanðī”.	ekwans ágantun wirán wíduntei”.
Exwaz tu wewx ^w ant: “Hludi, awi!	Ekwōs tu wjaukunt: “Ķludi, awi!
hurti áknuðai unsmí wítunðmaz:	ķirdi ágnutei insméi wídūntmas:
ner, fapiz, awjan-aur wulnōn	ner, patis, awjam-ir wilnān
sibī warman wesbran h ^w urneupī”.	sebi gormom westran kirnjautī”.
Awjan-x ^w e wulnō ne isti.	Áwjam-ke wilnā ne esti.
Pat héxluwaz awiz akran bukép.	Ta ķéķluwas awis agram bugít.

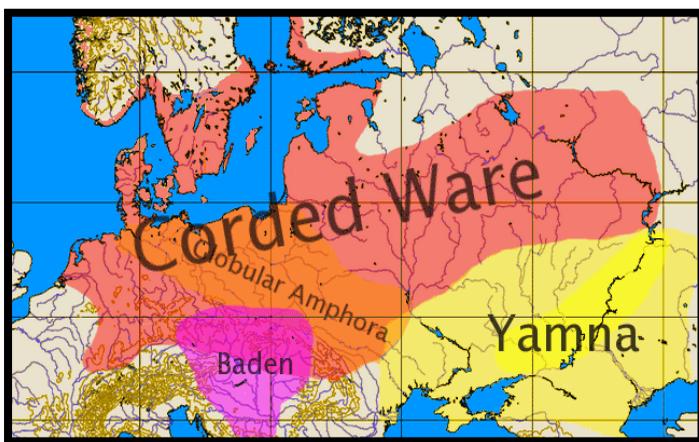
Translation: « *The Sheep and the Horses*. • A sheep that had no wool • saw horses, • one pulling a heavy wagon, • one carrying a big load, • and one carrying a man quickly. • The sheep said to the horses: • “My heart pains me, • seeing a man driving horses”. • The horses said: “Listen, sheep, • our 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. »

1.7.1. NORTHERN INDO-EUROPEAN DIALECTS

I. NORTH-WEST OR EUROPE'S INDO-EUROPEAN

The North-West Indo-European dialect *continuum*, also Europe's Indo-European, was spoken in the European Subcontinent in the centuries on either side of 2500 BC, evolving into the Pre-Celtic, Pre-Italic, Pre-Latin (probably within Pre-Italic), Pre-Germanic, Pre-Baltic, Pre-Slavic (or Pre-Balto-Slavic) IE dialects, among others. Its 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.

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 3200-2300 BC. The Globular Amphorae culture may be slightly earlier, but the relation between these cultures remains unclear.



Europe ca. 3200-2300 BC. The Germanic homeland is usually traced back to Jutland and southern Scandinavia; present-day Germany was the homeland for Celtic and Italic; the Eastern zone corresponds to Balto-Slavic. Beekes (1995).

NOTE. According to Adrados (1998), “[o]ne has to distinguish, in this huge geographical space, different locations. We have already talked about the situation of Germans to the West, and by their side, Celtic, Latin and Italic speakers; Balts and Slavs to the East, the former to the North of the later. See, among others, works by Bonfante (1983, 1984), about the old location of Baltic and Slavic-speaking communities. Isoglosses of different chronology let us partially reconstruct the language history. Note that the output obtained with Phonetics and Morphology match up essentially those of Porzig, who worked with Lexica”.

Kortlandt (1989), also considers that “[i]t is possible that the speakers of Italo-Celtic must be assigned to the Globular Amphora culture, and that Germanic grew out of a later component of the Corded Ware horizon (...) The Indo-Europeans who remained after the migrations became speakers of Balto-Slavic. If the speakers of the other *satem* languages can be assigned to the Yamnaya horizon and the western Indo-Europeans to the Corded Ware horizon, it is attractive to assign the ancestors of the Balts and the Slavs to the Middle Dnieper culture [an eastern extension of the Corded Ware culture, of northern Ukraine and Belarus]. If the origin of this culture “is to be sought in the Sredny Stog, Yamnaya and Late Tripolye cultures” and this phase is “followed by a middle period where the classic Corded Ware amphorae and beakers appear” (Mallory 1989: 248), the course of events corresponds nicely with the development of a *satem* language which was drawn into the western Indo-European sphere of influence”. Similarly, Adrados (1980) about the dialectal situation of Slavic (under a linguistic point of view): “To a layer of archaisms, shared or not with other languages (...) Slavic added different innovations, some common to Baltic. Some of them are shared with Germanic, as the oblique cases in -m and feminine participle; others with Indo-Iranian, so *satemization*, *Ruki* sound law (more present in Slavic than in Baltic) (...) Most probably, those common characteristics come from a recent time, from secondary contacts between IE III B [=Northern IE] (whose rearguard was formed by Balto-Slavs) and A [=Southern IE] (in a time when Greeks were not in contact anymore, they had already migrated to Greece)”.

On the archaeological quest for the *Urheimat*, Mallory & Adams (2006) make a complete summary of the different frameworks and models used. About the Retrospective Method, still favoured by many linguists, it is the “method where one examines those archaeological cultures that must have been associated with different Indo-European language groups and attempts to work backwards to the ‘proto-culture’. The unit of analysis here is the so-called ‘archaeological culture’, a classification device employed by archaeologists to deal with similar and geographically confined material culture and behaviour (...) Many of the language groups of Europe, i.e. Celtic, Germanic, Baltic, and Slavic, may possibly be traced back to the Corded Ware horizon of northern, central, and eastern Europe that flourished c. 3200-2300 BC. Some would say that the Iron Age cultures of Italy might also be derived from this cultural tradition. For this reason the Corded Ware culture is frequently discussed as a prime candidate for early Indo-European”.

Italic (with Latin), Celtic and Germanic are usually classified within a common West Indo-European *nucleus*. Balto-Slavic, on the other hand, is usually placed somewhere outside that West IE core, but always in close contact with it, as a North-West Indo-European dialect. Linguists have pointed out language contacts of Italic with Celtic,

1. Introduction

Celtic with Germanic, and Germanic with Balto-Slavic. Southern dialectal isoglosses affect Balto-Slavic and Tocharian, and only partially Germanic and Latin.

NOTE 1. Celtic too shares isoglosses with Southern dialects, according to Meier-Brügger (2003): “Celtic contacts with eastern Indo-Europe are ancient. Compare the case, among others, of relative pronouns, which in Celtic, contrarily to the Italic *kwo-/*kwi-, is represented by *Hio-, a characteristic that it shares with Greek, Phrygian, Indo-Iranian and Slavic”. Even though classifications of early proto-languages may vary depending on different criteria, they all have a known common origin, which is generally easier to reconstruct than their 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 (ca. 200 AD) – i.e. the features of the common language spoken by all Romance speakers, not the older, artificial, literary Classical Latin (ca. 100 BC) still less Old Latin (ca. 700 BC) – could be easily reconstructed, but the dialectal groups not. In fact, the actual groupings of the Romance languages are controversial, even knowing well enough Archaic, Classic and Vulgar Latin, and the history of Romance languages. Hence the difficulties in reconstructing and grouping individual North-West IE dialects, but the certainty in reconstructing a common North-West or Europe’s Indo-European language using raw linguistics, better explained if combined with archaeological data.

NOTE 2. On the inclusion of Pre-Latin IE within West Indo-Europe, against it there are some archaeological and linguistic theories (see Szemerényi, Colin Renfrew; v.s. for J.P. Mallory); Polomé (1983) & Schmidt (1984) say innovations common to Celtic and Germanic (later than those common to Celtic, Latin and Germanic), come from a time when Latin peoples had already migrated to the Italian peninsula. On the unity of Proto-Italic and Proto-Latin, Adrados (1998): “dubious is the old unity scheme, no doubt only partial, between Latin and Osco-Umbrian, which has been rejected by famous Italian linguists, relating every coincidence to recent contacts. I am not so sure about that, as the common innovations are big; cf. Beeler 1966, who doesn’t however dispel the doubts. Obviously, according to the decision taken, there are different historical consequences. If one thinks that both linguistic groups come from the North, through the Alps (cf. Tovar 1950), from the end of the 2nd millennium, a previous unity can be proposed. But authors like Devoto (1962) or Szemerényi (1962) made Latin peoples come from the East, through Apulia”. There has been a continued archaeological and (especially) linguistic support by mainstream IE studies to the derivation of Italic (and Latin) from a West Indo-European core, even after critics to the old Italo-Celtic concept (C. Watkins *Italo-Celtic Revisited*, 1963, K.H. Schmidt *Latein und Keltisch*, 1986); see Porzig (1954), Dressler (1971), Tovar (1970), Pisani (1974), Lehmann (1974), Bonfante (1983, 1984), Beekes (1995), Adrados, Bernabé, Mendoza (1998), etc.; on the archaeological question, see Ghirshman (1977), Thomas (1984), Gimbutas (1985), Harall (1995),...

Evolution of the reconstructed laryngeals of Proto-Indo-European in Europe's Indo-European include these vowel colourizations and compensatory lengthenings:

- PIE *H₁, the neutral laryngeal: *h₁a→**a**, *h₁e→**e**, *h₁o→**o**; *ah₁→**ā**, *eh₁→**ē**, *oh₁→**ō**.
- PIE *H₂, the a-colouring laryngeal: *h₂a→**a**, *h₂e→**a**, *h₂o→**a**; *ah₂→**ā**, *eh₂→**ā**.
- PIE *H₃, the o-colouring laryngeal: h₃e→**o**, h₃o→**o**; eh₃→**ō**, oh₃→**ō**.
- Often, but not always, interconsonantal H → **a**; as, *ph₂tér → **patér** (cf. PII pitár).
- PIH *rH→**r̥**, *lH→**l̥**, *n̥H→**n̥**, *m̥H→**m̥**; also, iH→**ī**, uH→**ū**.
- PIH *H before consonants → EIE Ø; cf. PIE *h₁dent-, EIE **dentis** (cf. PGk odōnts), “tooth”; PIE *h₂stér-, EIE **stér** (cf. PGk astér), etc.

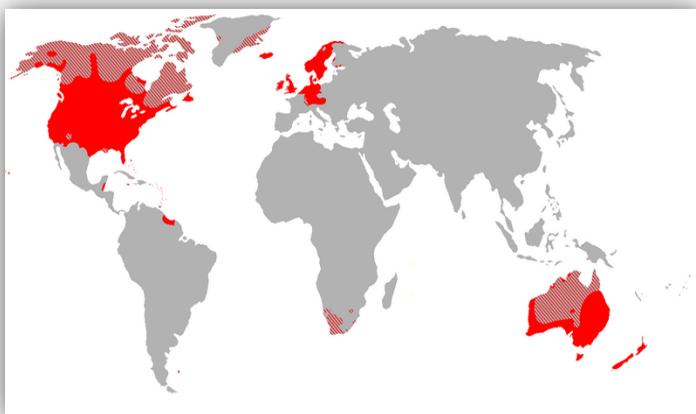
NOTE. The question is often made the other way round in IE studies, i.e. “according to these vowels reconstructed for North-West Indo-European, Proto-Greek and Proto-Indo-Iranian, which combination of *laryngeal+vowel* or *vowel+laryngeal* could make them all fit into a common mother-language?” For clarity purposes, Common PIE is taken in this book as example for the phonology of early dialects, but enough certainty in vocalism (for language revival purposes) is to be found only in EIE, PGk and PII; exact regularity or congruence of a common Proto-Indo-European phonology is neither necessary nor searched for, as there are many variations in the laryngeal theories proposed by scholars, who reconstruct from just one (Szemerényi) to eight (Puhvel) or nine (Adrados); a general reconstruction of three laryngeals is used here for its simplicity and wide acceptance today. For more on this see Appendix II.3, The Laryngeal Theory.

A. GERMANIC

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 Low Germanic dialects (like Dutch) and the Scandinavian languages.

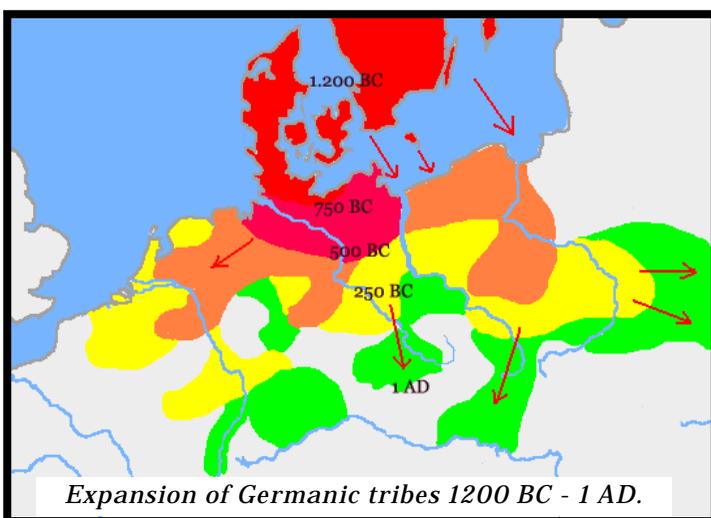


Spread of Germanic languages today.

1. Introduction

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 an earlier Pre-Proto-Germanic, a dialect of Europe's Indo-European branch dated ca. 1500-500 BC. The succession of archaeological horizons suggests that before their language differentiated into the individual Germanic branches the Proto-Germanic speakers lived in southern Scandinavia and along the coast from the Netherlands in the west to the Vistula in the east around 750 BC. 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.

NOTE. A few surviving inscriptions in a runic script from Scandinavia dated to ca. 200 are thought to represent a later stage of Proto-Norse; according to Bernard Comrie, it represents a *Late Common Germanic* which followed the "Proto-Germanic" stage. Several historical linguists have pointed towards the apparent material



and social continuity connecting the cultures of the Nordic Bronze Age (1800-500 BCE) and the Pre-Roman Iron Age (500 BCE - 1 CE) as having implications in regard to the stability and later development of the Germanic language group. Lehmann (1977) writes "Possibly the most important conclusion based on archeological evidence with relevance for linguistic purposes is the assumption of 'one huge cultural area' which was undisturbed for approximately a thousand years, roughly from 1500-500 BC. Such a conclusion in a stable culture permits inferences concerning linguistic stability, which are important for an interpretation of the Germanic linguistic data". Also, on setting the upper boundary of a comprehensive description of Proto-Germanic grammar, Lehmann (2005) wrote: "a grammar of Proto-Germanic must be a description of the language from approximately 2500 BC to the beginning of the common era".

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 Gothic alphabet developed by Bishop Ulfilas for his translation of the Bible in the 4th century. Later, Christian priests and monks who spoke and read Latin in addition to their native Germanic tongue began writing the Germanic languages with slightly modified Latin letters, but in Scandinavia, runic alphabets remained in common use throughout the Viking Age.

The so-called Grimm's law is a set of statements describing the inherited Europe's Indo-European stops as they developed in Pre-Proto-Germanic. 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:



Negau helmet. It reads (from right to left): harikastiteiva\\ip, "Harigast the priest".

- PIE voiceless stops change into PGmc. voiceless fricatives: **p**→*f*, **t**→*θ*, **k**→*x*, **k^w**→*x^w*.
- PIE voiced stops become PGmc. voiceless stops: **b**→*p*, **d**→*t*, **g**→*k*, **g^w**→*k^w*.
- PIE voiced aspirated stops lose their aspiration and change into plain voiced stops: **b^h**→*b*, **d^h**→*d*, **g^h**→*g*, **g^{wh}**→*g^w*, *g*, *w*.

Verner's Law addresses a category of exceptions, stating that unvoiced fricatives are voiced when preceded by an unaccented syllable: PGmc. *s*→*z*, *f*→*v*, *θ*→*ð*; as, EIE **b^hratēr** → PGmc. *brōþēr*, "brother", but EIE **mātér** → PGmc. *mōðēr* "mother".

NOTE 1. W. P. Lehmann (1961) considered that Jacob Grimm's "First Germanic Sound Shift", or Grimm's Law and Verner's Law, which pertained mainly to consonants and were considered for a good many decades to have generated Proto-Germanic, were Pre-Proto-Germanic, and that the "upper boundary" was the fixing of the accent, or stress, on the root syllable of a word, typically the first. Proto-Indo-European had featured a moveable pitch accent comprising "an alternation of high and low tones" as well as stress of position determined by a set of rules based on the lengths of the word's syllables.

The fixation of the stress led to sound changes in unstressed syllables. For Lehmann, the "lower boundary" was the dropping of final -a or -e in unstressed syllables; for example, PIE **woid-á** >, Goth. *wait*, "knows" (the > and < signs in linguistics indicate a genetic descent). Antonsen (1965) agreed with Lehmann about the upper boundary but later found runic evidence that the -a was not

1. Introduction

dropped: Gmc. *ékwakraz ... wraita*, “*I wakraz ... wrote (this)*”. He says: “We must therefore search for a new lower boundary for Proto-Germanic”.

NOTE 2. Sometimes the shift produced allophones (consonants that were pronounced differently) depending on the context of the original. With regard to original PIE **k** and **k^w**, Trask (2000) says that the resulting PGmc. *x* and *x^w* were reduced to *h* and *h^w* in word-initial position. Consonants were lengthened or prolonged under some circumstances, appearing in some daughter languages as geminated graphemes. Kraehenmann (2003) states that Proto-Germanic already had long consonants, but they contrasted with short ones only word-medially. Moreover, they were not very frequent and occurred only intervocally almost exclusively after short vowels. The phonemes *b*, *d*, *g* and *g^w*, says Ringe (2006) were stops in some environments and fricatives in others.

Effects of the aforementioned sound laws include the following examples:

- **p**→*f*: EIE **pod̥s** “*foot*”, PGmc. *fōts*; cf. Goth. *fōtus*, O.N. *fōtr*, O.E. *fōt*, O.H.G. *fuoz*.
- **t**→*þ, ð*: EIE **tritjós** “*third*”, PGmc. *þriðjaz*; cf. Goth. *þridja*, O.N. *þriðe*, OE. *þridda*, O.H.G. *dritto*.
- **k**→*x, h*: EIE **kwon** “*dog*”, PGmc. *xunðaz*; cf. Goth. *hunds*, O.N. *hundr*, O.E. *hund*, O.H.G. *hunt*.
- **k^w**→*x^w, h^w*: EIE **k^wos** “*what, who*”, Gmc. *h^woz*; cf. Goth. *hwas*, O.N. *hverr*, O.S. *hwe*, O.E. *hwā*, O.Fris. *hwa*, O.H.G. *hwër*.
- **b**→*p*: EIE **werbō** “*throw*”, Gmc. *werpō*; cf. Goth. *wairpan*, O.S. *werpan*, O.N. *verpa*, O.E. *weorpan*, M.L.G., Du. *werpen*, Ger. *werfen*.
- **d**→*t*: EIE **dekm̥** “*ten*”, Gmc. *tehun*; cf. Goth. *taihun*, O.S. *tehan*, O.N. *tiu*, O.Fris. *tian*, O.Du. *ten*, O.H.G. *zehan*.
- **g**→*k*: EIE **gelu** “*ice*”, Gmc. *kaldaz*; cf. Goth. *kalds*, O.N. *kaldr*, O.E. *cald*, O.H.G. *kalt*.
- **g^w**→*kw*: EIE **g^wīwós** “*alive*”, Gmc. *k^wi(k)waz*; cf. Goth. *k^wius*, O.N. *kvikr*, O.E. *cwic*, O.H.G. *quec*.
- **b^h**→*b*: EIE **b^hrātēr** “*brother*”, Gmc. *brōþēr*; cf. Goth. *bróþar*, O.N. *brōþir*, O.E. *brōþor*, O.H.G. *bruoder*.
- **d^h**→*d*: EIE **d^hworis** “*door*”, Gmc. *duriz*; cf. Goth. *daúr*, O.N. *dyrr*, O.E. *duru*, O.H.G. *turi*.

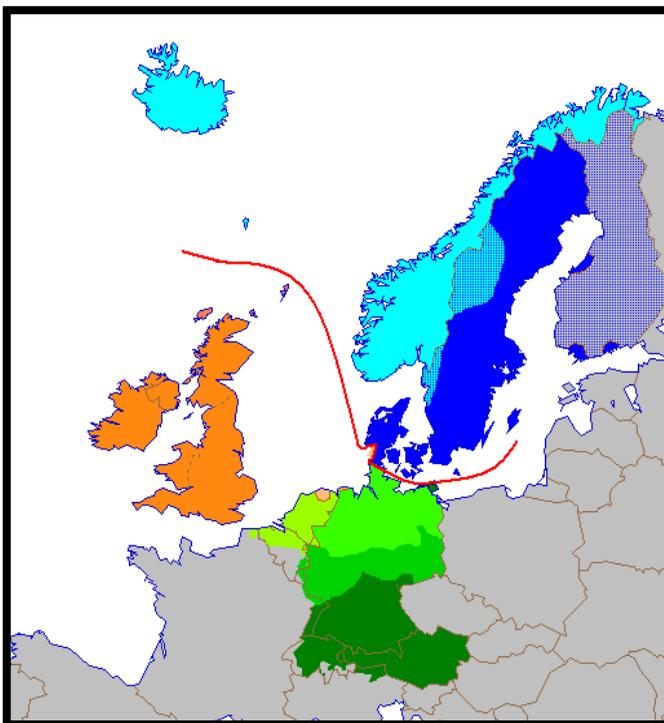
- **g^h→g**: EIE **g^hansis** “goose”, Gmc. *gansiz*; cf. Goth *gansus*, O.N. *gās*, O.E. *gōs*, O.H.G. *gans*.
- **g^{wh}→gw/g/w**: EIE **g^{wh}ormos** “warm”, Gmc. *warmaz*; cf. O.N. *varmr*, O.E. *wearm*, O.H.G. *warm*. EIE **g^{wh}ondos** “fight”, Gmc. *gandaz*; cf. Goth. *gunþs*, O.N. *gandr*, O.E. *gūþ*, O.H.G. *gund*.

A known exception is that the voiceless stops did not become fricatives if they were preceded by PIE **s**., i.e. **sp**, **st**, **sk**, **sk^w**. 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*.

EIE vowels: **a, o→a**; EIE **ā, ō→ō**. PGmc. had then short *i*, *u*, *e*, *a*, and long *ī*, *ū*, *ē*, *ō*, *ǣ*?

NOTE 1. Similar mergers happened in the Slavic languages, but in the opposite direction. At the time of the merge, the vowels probably were [ɔ] and [ɔ:] before their timbres differentiated into maybe [α] and [ɔ:].

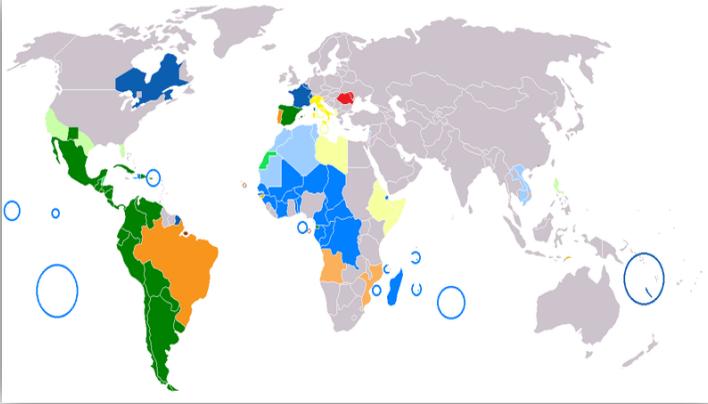
NOTE 2. PGmc. *ǣ* and *ē* are also transcribed as *ē¹* and *ē²*; *ē²* is uncertain as a phoneme, and only reconstructed from a small number of words; it is posited by the comparative method because whereas all probable instances of inherited EIE *ē* (PGmc. **ē¹*) are distributed in Gothic as *ē* and the other Germanic languages as *ā*, all the Germanic languages agree on some occasions of *ē* (e.g. PGmc. *hē²r* → Goth., O.E., O.N. *hēr*, “here”). Krahe treats *ē²* (secondary *ē*) as identical with *ī*. It probably continues EIE **ei** or **ēi**, and it may have been in the process of transition from a diphthong to a long simple vowel in the Proto-Germanic period. Gothic makes no orthographic and therefore presumably no phonetic distinction between *ē¹* and *ē²*. The existence of two Proto-Germanic [e:] -like phonemes is supported by the existence of two e-like Elder Futhark runes, *Ehwaz* and *Eihwaz*.



Germanic dialects in Europe. The line divides Western from Northern dialects.

1. Introduction

B. LATIN

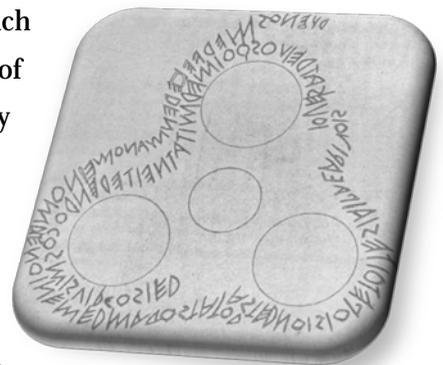


Regions where Romance languages are spoken, either as mother tongue or as second language.

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 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 oversea empires established by Spain, Portugal and France after the 15th century then spread Romance to the other continents — to such an extent that about two thirds of all Romance speakers are now outside Europe.

Latin is usually classified, along with Faliscan, as an Italic dialect. The Italic speakers were not native to



The Duenos (O.Lat. duenus, Lat. buenus) Inscription in Old Latin, ca. 6th century BC.

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. As Rome extended its political dominion over Italy, Latin became dominant over the other Italic languages, which ceased to be spoken perhaps sometime in the 1st century AD.

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, spoken in the area around *Falerii Veteres*, north of the city of Rome.
 - Latin, spoken in west-central Italy. The Roman conquests eventually spread it throughout the Roman Empire and beyond.



Iron Age Italy, ca 800 BC. In central Italy, Italic languages. In southern and north-western Italy, other Indo-European languages. Venetic, Sicilian and Sikel were possibly IE.

The ancient Venetic language, as revealed by its inscriptions (including complete sentences), was also closely related to the Italic languages and is sometimes even

1. Introduction

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 IE language.

Phonetic changes from EIE to Latin include: **b^h**→*f/b*, **d^h**→*f/b*, **g^h**→*h/f*, **g^w**→*w/g*, **k^w**→*kw/k*, **p**→*p/kw*.

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, Lat. *uolo, uelle, uolui*, and other such forms from PIE **wel-**, *will*, were represented by words derived from **g^{her-}**, *desire*, cf. Osc. *herest*, “*he wants, desires*” as opposed to Lat. *uult* (id.). Lat. *locus*, “*place*” was absent and represented by Osc. *slaagid*.

In phonology, Oscan also shows a different evolution, as EIE **k^w**→ Osc. *p* instead of Lat. *kw* (cf. Osc. *pis*, Lat. *quis*); EIE **g^w**→ Osc. *b* instead of Latin *w*; EIE medial **b^h**, **d^h**→ Osc. *f*, in contrast to Lat. *b* or *d* (cf. Osc. *mefiai*, Lat. *mediae*); etc.

NOTE. A specimen of Faliscan appears written round the edge of a picture on a patera: “*foied vino pipafo, cra carefo*”, which in Old Latin would have been “*hodie vinom bibabo, cras carebo*”, translated as “*today I will drink wine; tomorrow I won't have any*” (R. S. Conway, *Italic Dialects*). Among other distinctive features, it shows the retention of medial *f* which in Latin became *b*, and evolution of EIE **g^h**→*f* (*fo-*, contrast Lat. *ho-*).

Hence the reconstructed changes of North-West Indo-European into Proto-Italic:

- Voiced labiovelars unround or lenite: **g^w**→*g/w*, **g^{wh}**→*g^h*.



Forum inscription in Latin, written boustrophedon.

- Voiced aspirates become first unvoiced, then fricativize: **b^h** → **p^h** → **ϕ** → **f**; **d^h** → **t^h** → **θ**; **g^h** → **k^h** → **x**.

NOTE. About PIE intervocalic **g^h** → Ita. *x*, linguists (see Joseph & Wallace 1991) generally propose that it evolves as Faliscan *g* or *k*, while in Latin it becomes glottal *h*, without a change of manner of articulation. Picard (1993) rejects that proposal citing abstract phonetic principles, which Chela-Flores (1999) argues citing examples of Spanish phonology.

- EIE **s** → Ita. *θ* before *r* (cf. Ita. *kereθrom*, Lat. *cerebrum*); unchanged elsewhere.

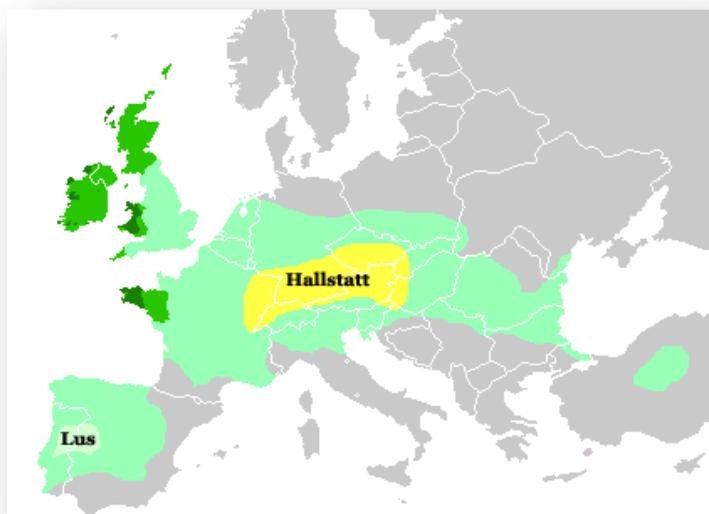
Up to 8 cases are found; apart from the 6 cases of Classic Latin (i.e. N-V-A-G-D-Ab), there was a Locative (cf. Lat. *proximae viciniae, domī, carthaginī*; Osc. *aasai*, Lat. “*in ārā*” etc.) and an Instrumental (cf. Columna Rostrata Lat. *pugnandod, marid, naualid*, etc; Osc. *cadeis amnud*, Lat. “*inimicitiae causae*”; Osc. *preiuatud*, Lat. “*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*.

C. CELTIC

The Celtic languages are the languages descended from Proto-Celtic, or “Common Celtic”, an Indo-European proto-language.

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



Diachronic distribution of Celtic peoples: maximal expansion (ca. 200 BC) and modern “Celtic nations” and Celtic-speaking territories.

and the North Sea, up the Rhine and down the Danube to the Black Sea and the Upper

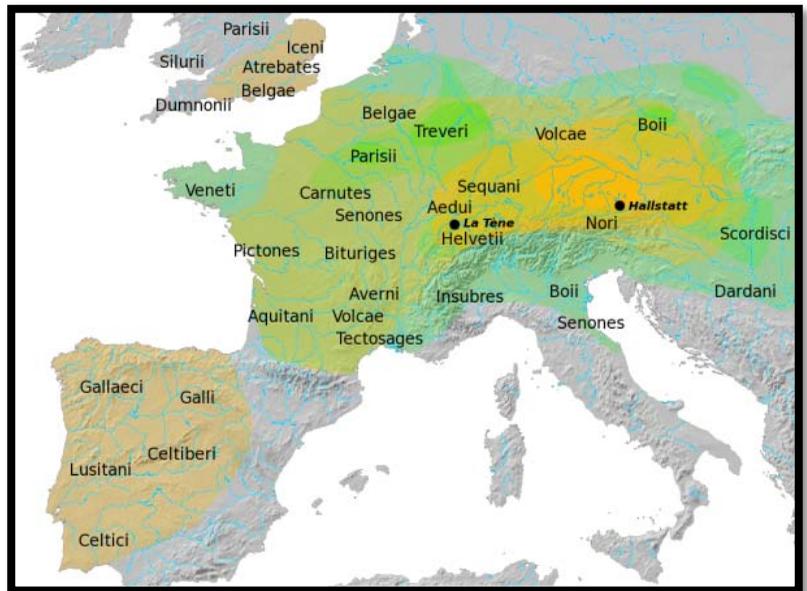
1. Introduction

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.

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.

NOTE. 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 EIE k^w , which became $*p$ in the P-Celtic languages but $*k$ in Goidelic. An



Hallstatt core territory (ca. 800 BC) and its influence (ca. 500 BC); and La Tène culture (ca. 450) and its influence (ca. 50 BC). Some major Celtic tribes have been labeled.

example is the Cel. verbal root k^wrin - “to buy”, which became Welsh $pryn$ -, but O.Ir. $cren$ -.

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 $[k^w]$ to $[p]$ occurred independently or regionally. 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 $[\beta]$, a nasalized voiced bilabial fricative (an extremely rare 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 include:

- Consonants: **p** → ϕ → *h* → \emptyset in initial and intervocalic positions. Cel. ϕs → *xs*, ϕt → *xt*

NOTE. EIE **p** was lost in Proto-Celtic, apparently going through the stages ϕ (perhaps in Lus. *porcos*, *v.i.*) and *h* (perhaps attested by the toponym *Hercynia* if this is of Celtic origin) before being lost completely word-initially and between vowels. EIE **sp-** became Old Irish *s* and Brythonic *f*; while Schrijver (1995) argues there was an intermediate stage *s ϕ -* (in which ϕ remained an independent phoneme until after Proto-Insular Celtic had diverged into Goidelic and Brythonic), McCone (1996) finds it more economical to believe that **sp-** remained unchanged in PC, that is, the change **p** to ϕ did not happen when **s** preceded.

- Aspirated: **d^h** → *d*, **b^h** → *b*, **g^h** → *x*, **g^{wh}** → *g^w*; but **g^w** → *b*.
- Vowels: \bar{o} → \bar{a} , \bar{u} (in final syllable); \bar{e} → \bar{i} ; EIE **u-w** → Cel. *o-w*.
- Diphthongs: **$\bar{a}i$** → *ai*, **$\bar{e}i$** → *ei*, **$\bar{o}i$** → *oi*; **$\bar{a}u$** → *au*, **$\bar{e}u, \bar{o}u$** → *ou*.
- Sonorants: **l̥** → *la, li* (before stops); **r̥** → *ar, ri* (before stops); **m̥** → *am*; **n̥** → *an*.

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).

NOTE. 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 PIE gentivie **-osjo** 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 \bar{i} 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.

Other arguments include that both Celtic and Italic have collapsed the PIE Aorist and Perfect into a single past tense, and the \bar{a} -subjunctive, because both Italic and Celtic have a subjunctive descended from an earlier optative in \bar{a} -. Such an optative is not known from other languages, but the suffix occurs in Balto-Slavic and Tocharian past tense formations, and possibly in Hitt. *-ahh-*.

1. Introduction

D. 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.

There is much debate whether Pre-Proto-Slavic branched off directly from Europe's Indo-European in 2000 BC, or whether it passed through a common *Proto-Balto-Slavic* stage which had necessarily split apart before 1000 BC in its two main sub-branches.

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 secession from a common Proto-Balto-Slavic is estimated on archaeological and glottochronological criteria to have occurred between 1500-1000 BC. Common Slavic is usually reconstructible to around 600 AD.



Historical distribution of the Slavic languages. The larger shaded area is the Prague-Penkov-Kolochin complex of cultures of the 6th to 7th 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, dated ca. 500 AD.

By the 7th century, Common Slavic 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.

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.



Page from Codex Zographensis (10th 11th c. AD) in Old Church Slavonic.

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. *Vel'komoravské učilište*) and were used for government and religious documents and books. In 885, the use of the O.C.S. 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

1. Introduction

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

Vowel changes from North-West Indo-European to Proto-Slavic:

➤ EIE **ī**, **ei** → Sla. *i₁*; EIE **i** → **i* → Sla. *b*; EIE **u** → **u* → Sla. *τ*; EIE **ū** → Sla. *y*.

➤ EIE **e** → Sla. *e*; EIE **ē** → Sla. *ě₁*;

➤ EIE **en**, **em** → Sla. *ę*; EIE **an**, **on**; **am**, **om** → **an*; **am* → Sla. *φ*.

➤ EIE **a**, **o** → **a* → Sla. *O*; EIE **ā**, **ō** → **ā* → Sla. *a*; EIE **ai**, **oi** → **ai* → Sla. *ě₂*.
reduced **ai* (**āi*/**ui*) → Sla. *i₂*; EIE **au**, **ou** → **au* → Sla. *u*.

NOTE 1. Apart from this simplified equivalences, other evolutions appear (see Kortlandt's *From Proto-Indo-European to Slavic* at <<http://www.kortlandt.nl/publications/art066e.pdf>>):

○ The vowels *i₂*, *ě₂* developed later than *i₁*, *ě₁*. In Late Proto-Slavic there were no differences in pronunciation between *i₁* and *i₂* as well as between *ě₁* and *ě₂*. They had caused, however, different changes of preceding velars, see below.

○ Late Proto-Slavic yers *ь*, *ъ* < earlier *i*, *u* developed also from reduced EIE **e**, **o** respectively. The reduction was probably a morphologic process rather than phonetic.

○ We can observe similar reduction of **ā** into **ū* (and finally *y*) in some endings, especially in closed syllables.

○ The development of the Sla. *i₂* was also a morphologic phenomenon, originating only in some endings.

○ Another source of the Proto-Slavic *y* is **ō* in Germanic loanwords – the borrowings took place when Proto-Slavic no longer had *ō* in native words, as EIE **ō** had already changed into **ā*.

○ EIE **a** (from PIE **ǵ**) disappeared without traces when in a non-initial syllable.

○ EIE **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, EIE 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 **ē*, **ā*, **ō* with further development like above.

NOTE 2. Other vocalic changes from Proto-Slavic include **jo*, **jь*, **jy* changed into **je*, **jь*, **ji*; **o*, **ь*, **y* also changed into **e*, **ь*, **i* after **c*, **з*, **s*' which developed as the result of the 3rd palatalization; **e*, **ě* changed into **o*, **a* after **č*, **ж*, **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 found:

- EIE **p** → Sla. *p*; EIE **b**, **b^h** → Sla. *b*.
- EIE **t** → Sla. *t*; EIE **d**, **d^h** → Sla. *d*.
- EIE **k**, **k^w** → Sla. *K* (palatalized ***kⁱ** → Sla. *s*); EIE **g**, **g^h**, **g^w**, **g^{wh}** → Sla. *g* (palatalized (***gⁱ**, ***gⁱ^h** → Sla. *z*)
- EIE **s** → Sla. *s*; before a voiced consonant EIE [**z**] → Sla. *z*; before a vowel when after **r**, **u**, **k**, **i**, probably also after **l** → Sla. *x*.
- EIE word-final **m** → Sla. *n* (<BSl. **n*).
- EIE **m̥** → Sla. *im*, *um*; EIE **m̥** → Sla. *in*, *un*; EIE **l̥** → Sla. *il*, *ul*; EIE **r̥** → Sla. *ir*, *ur*.
- EIE **w** → Sla. *v* (<BSl. **w*); EIE **j** → Sla. *j*.

In some words the Proto-Slavic *x* developed from other PIE phonemes, like **kH**, **ks**, **sk**.

E. BALTIC

The Baltic languages are a group of related languages belonging to the IE language family, spoken in areas extending east and southeast of the Baltic Sea in Northern Europe.

The language group is often 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.



1. Introduction

While related, Lithuanian, Latvian, and particularly Old Prussian differ substantially from each other and are not mutually intelligible.

The oldest Baltic linguistic record is the Elbinger lexicon of the beginning of the 14th century AD. It contains 802 Old Prussian equivalents of Old Middle German words. The oldest Baltic text is Old Prussian as well; it comes from the middle of the 14th century AD and includes only eleven words. The first Old Lithuanian and Old Latvian texts come from the 16th century and appear already in book form, and were translations of a catechism and the Lord's Prayer.

Baltic and Slavic share so many similarities that 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, dated ca. 1500-500 BC, depending on the different guesstimates.

NOTE 1. For those guesstimates, "Classical glottochronology" conducted by Czech Slavist M. Čejka in 1974 dates the Balto-Slavic split to -910±340 BCE, Sergei Starostin in 1994 dates it to 1210 BCE, and "recalibrated glottochronology" conducted by Novotná & Blažek dates it to 1400-1340 BCE. This agrees well with Trziniec-Komarov culture, localized from Silesia to Central Ukraine and dated to the period 1500–1200 BCE.

NOTE 2. 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*". 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.

A new theory was proposed in the 1960s by V. Ivanov and V. Toporov: that the Balto-Slavic proto-language split from the start into West Baltic, East Baltic and Proto-Slavic. In their framework, Proto-Slavic is a peripheral and innovative Balto-Slavic dialect which suddenly expanded, due to a conjunction of historical circumstances. Onomastic evidence shows that Baltic languages were once spoken in much wider territory than the one they cover today, and were later replaced by Slavic.

NOTE. The most important of these common Balto-Slavic isoglosses are:

- Winter's law: lengthening of a short vowel before a voiced plosive, usually in a closed syllable.
- Identical reflexes of PIE syllabic sonorants, usually developing *i* and *u* before them. Kuryłowicz thought that *uR reflexes arose after PIE velars, and also notable is also older opinion of J.Endzelīns and R. Trautmann according to whom *uR reflexes are the result of zero-grade of morphemes that had EIE o → PBSl. *a in normal-grade. Matasović (2008) proposes following internal rules after EIE syllabic R → BSl. *əR: 1) *ə→*i in a final syllable; 2) *ə→*u after velars and before nasals; 3) *ə→*i otherwise.
- Hirt's law: retraction of PIE accent to the preceding syllable closed by a laryngeal.
- Rise of the Balto-Slavic acute before PIE laryngeals in a closed syllable.
- Replacement of PIE genitive singular of thematic nouns with ablative.
- Formation of past tense in *-ē (cf. Lith. pret. *dāvė*, "he gave", O.C.S. imperfect *bě*, "he was")
- Generalization of the IE neuter **to-** stem to the nominative singular of masculine and feminine demonstratives instead of IE **so-** pronoun, **so**, **sā**, **tod** → BSl. *tos*, *tā*, *tod*.
- Formation of *definite adjectives* with a construction of adjective and relative pronoun; cf. Lith. *geràsis*, "the good", vs. *gėras*, "good"; O.C.S. *dobrŭjb*, "the good", vs. *dobrŭ*, "good".

Common Balto-Slavic innovations include several other prominent, but non-exclusive isoglosses, such as the satemization, Ruki, change of PIE **o** → BSl. **a* (shared with Germanic, Indo-Iranian and Anatolian) and the loss of labialization in PIE labiovelars (shared with Indo-Iranian, Armenian and Tocharian). Among Balto-Slavic archaisms notable is the retention of traces of an older PIE accent. 'Ruki' is the term for a sound law which is followed especially in BSl. and Aryan 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**, **g^h**). 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 IE by Holger Pedersen.

Baltic and Slavic show a remarkable amount of correspondence in vocabulary too; there are at least 100 words exclusive to BSl., either being a common innovation or sharing the same semantic development from a PIE root; as, BSl. **l̥ēpā*, "*tilia*" → Lith. *lėpa*, O.Prus. *līpa*, Ltv. *liēpa*; Sla. **lipa*; BSl. **rankā*, "*hand*" → Lith. *rankà*, O.Prus. *rānkan*, Ltv. *rūoka*; Sla. **rōkà* (cf. O.C.S. *rōka*). BSl. **galwǎ*, "*head*" → Lith. *galvà*, O.Prus. *galwo*, Ltv. *galva*; Sla. **golvà* (cf. O.C.S. *glava*).

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. It became extinct after the Roman Empire conquered the region and assimilated the inhabitants.

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.

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

- PIE short *o→a, as in the last syllable of the genitive *kalatoras*.
- PIE final *m→n, as in *aran*.
- PIE *nj→nn, as in the Messapian praenomen *Dazohannes* vs. the Illyrian praenomen *Dazonius*; the Messapian genitive *Dazohonnihi* vs. Illyrian genitive *Dasonii*, etc.
- PIE *tj→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.
- PIE *sj→ss, as in Messapian *Vallasso* for *Vallasio*, a derivative from the shorter name *Valla*.
- The loss of final *-d, as in *tepise*, and probably of final *-t, as in *-des*, perhaps meaning “set”, from PIE *d^he-, “set, put”.
- The change of voiced aspirates in Proto-Indo-European to plain voiced consonants: PIE *d^h→d, as in Messapian *anda* (< PIE *en-d^ha- < PIE *en-, “in”, compare Gk. *entha*); and PIE *b^h→b, as in Messapian *beran* (< PIE *b^her-, “to bear”).
- PIE *au→ā before (at least some) consonants: *Bāsta*, from *Bausta*.
- The form *penkaheh* – which Torp very probably identifies with the Oscan stem *pompaio* – a derivative of the Proto-Indo-European numeral *penk^we, “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

Venetian 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. It was a *Centum* dialect.

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. The inscriptions use a variety of the Northern Italic alphabet, similar to the Old Italic alphabet. It became extinct around the 1st century when the local inhabitants were assimilated into the Roman sphere.

NOTE. 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.

Interesting parallels with Germanic have also been noted, especially in pronominal forms:

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

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

Venetian 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. *vhrateri* (< PIE **b^hreh₂tereī*), “to the brother”.

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

1. Introduction

There are also indications of the developments of PIE initial *g^w→w-, PIE *k^w→kv and PIE initial *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...** (e.g. **penkwe** → *k^wenk^we, “five”, **perkwo** → *k^werk^wu, “oak”), 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 IE languages, primarily Celtic (Gaulish) and 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 IE language which was spoken by the ancient Liburnians in the region of Liburnia (south of the Istrian peninsula) in classical times. It is usually classified as a *Centum* language. 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.

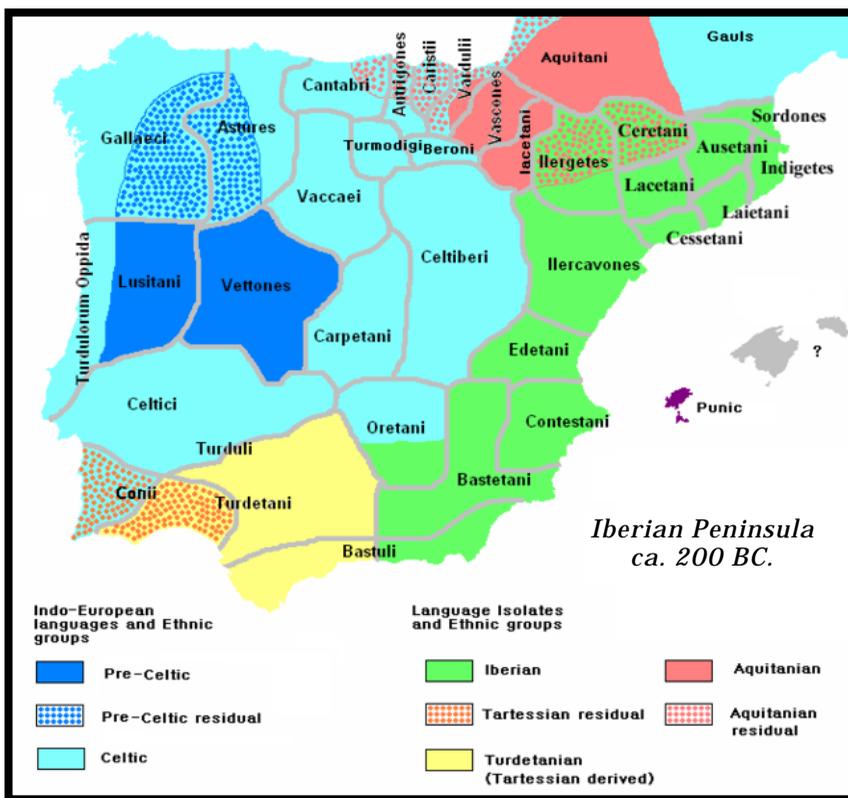
NOTE. No writings in Liburnian are known, though. The grouping of Liburnian with Venetic is based on the Liburnian onomastics. In particular, Liburnian anthroponyms show strong Venetic affinities, with many common or similar names and a number of common roots, such as *Vols-*, *Volt-*, and *Host-* (<PIE *g^hos-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.

Liburnia was conquered by the Romans in 35 BC, and its language was eventually replaced by Latin, undergoing language death probably very early in the Common Era.

LUSITANIAN

Lusitanian or Lusatian (so named after the *Lusitani* or Lusitanians) was a Paleohispanic IE 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 river in the western area of the Iberian Peninsula, where they were established already before the 6th c. BC.

Their language is usually considered a Pre-Celtic (possibly Italo-Celtic) IE dialect, and it is sometimes associated with the language of the Vettones and with the linguistic substratum of the Gallaeci and Astures, based on archaeological findings and descriptions of ancient historians.

NOTE. The affiliation of the Lusitanian language within a Pre-Celtic (or Italo-Celtic) IE group is still debated. There are those who endorse that it is a Celtic language, a theory 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 PIE 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, it does not necessarily

1. Introduction

preclude the possibility of Lusitanian being Celtic, because of the supposed evolution of PIE initial $*p \rightarrow *ϕ \rightarrow *h \rightarrow \text{Cel. } \emptyset$, so it might have been an early Proto-Celtic (or Italo-Celtic) dialect that split off before the loss of $p-$, or when $p-$ had become $*ϕ-$ (before shifting to $h-$ and then being lost); the letter p of the Latin alphabet could have been used to represent either sound.

F. Villar and R. Pedrero relate 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 Marrucinian *Iovia*, Lus. *Iovea(i)*, etc. and other lexical items, as Umb. *gomia*, Lus. *comaia*, with some other grammatical elements.

II. 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*”).

NOTE. These are sometimes identified with the Yuezhi and the Kushans, while the term *Tokharistan* usually refers to 1st millennium Bactria. A Turkic text refers to the Turfanian language (Tocharian A) as *twqry*. F. W. K. Müller has associated this with the name of the Bactrian *Tokharoi*. In Tocharian, the language is referred to as *arish-kāna* 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.

Note. 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.

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 were not intercomprehensible. The common Proto-Tocharian language must have preceded the attested languages by several centuries, probably dating to the 1st millennium BC.