

1.7.2. SOUTHERN INDO-EUROPEAN DIALECTS

I. GREEK



Greek 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 (especially in the FYROM), Bulgaria, Albania and Turkey.

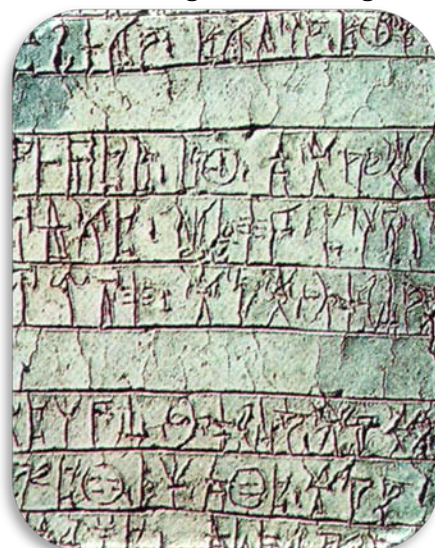
The major dialect groups of the Ancient Greek period can be assumed to have developed not later than 1120 BC, at the time of the Dorian invasions, and their first appearances as precise alphabetic writing began in the 8th century BC. The ancient Greeks themselves considered there to be three major divisions of the Greek people, into Dorians, Aeolians, and Ionians (including Athenians), each with their own defining and distinctive dialects. Allowing for their oversight of Arcadian, an obscure mountain dialect, and Cyprian, far from the center of Greek scholarship, this division of people and

1. Introduction

language is quite similar to the results of modern archaeological and linguistic investigation.

Greek has been spoken in the Balkan Peninsula since 2000 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.



Linear B has roughly 200 signs, divided into syllabic signs with phonetic values and logograms with semantic values.

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. 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 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*, '*Epikóheĩ*', 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 southern PIE dialect, was spoken in the late 3rd millennium BC, roughly at the same time as North-West Indo-European and Proto-Indo-Iranian, most probably in the Balkans. It was probably the ancestor of Phrygian too, and possibly that of Ancient

Macedonian, Dacian, Thracian, and arguably Armenian. The unity of Proto-Greek probably ended as Hellenic migrants, speaking the predecessor of the Mycenaean language, entered the Greek peninsula around the 21st century BC. They were then separated from the Dorian Greeks, who entered the peninsula roughly one millennium later, speaking a dialect that in some respects had remained more archaic.

NOTE. For Pelasgian and other Greek substrates as IE, some have cited different phonological developments in words like $\tau\upsilon\acute{\mu}\beta\omicron\varsigma$ (*tumbos* < PIE **d^hmb^hos*) or $\pi\upsilon\rho\gamma\omicron\varsigma$ (*purgos* < PIE **b^hrg^hos*).

Proto-Greek was affected by a late *Satemization* trend, evidenced by the (post-Mycenaean) change of labiovelars into dentals before *e* (e.g. *k^we* → *te* “and”).

The primary sound changes from (laryngeal) PIE to Proto-Greek include:

- Aspiration of PIE intervocalic **s* → PGk *h*.

NOTE. The loss of PIE prevocalic **s*- was not completed entirely, famously evidenced by *sus* “sow”, *dasus* “dense”; *sun* “with”, sometimes considered contaminated with PIE **kom* (cf. Latin *cum*) to Homeric / Old Attic *ksun*, is possibly a consequence of Gk. psi-substrate (See Villar).

- De-voicing of voiced aspirates: **b^h* → *p^h*, **d^h* → *t^h*, **g^h* → *k^h*, **g^{wh}* → *k^{wh}*.
- Dissimilation of aspirates (Grassmann’s law), possibly post-Mycenaean.
- PIE word-initial **j*- (not **Hj*-) is strengthened to PGk *dj*- (later Gk. ζ-).
- Vocalization of laryngeals between vowels and initially before consonants, i.e. **h₁* → *e*, **h₂* → *a*, **h₃* → *o*.

NOTE. The evolution of Proto-Greek should be considered with the background of an early Palaeo-Balkan *Sprachbund* that makes it difficult to delineate exact boundaries between individual languages. The characteristically Greek representation of word-initial laryngeals by prosthetic vowels is shared by the Armenian language, which also shares other phonological and morphological peculiarities of Greek, *vide infra*.

- The sequence CRHC (where C = consonant, R = resonant, H = laryngeal) becomes PIE *CRh₁C* → PGk CRēC; PIE *CRh₂C* → PGk CRāC; PIE *CRh₃C* → PGk CRōC.
- The sequence PIE *CRHV* (where V = vowel) becomes PGk CaRV.

NOTE. It has also been proposed by Sihler (2000) that *Vk^w* → *uk^w*; cf. PIE **nok^wts*, “night” → PGk *nuk^wts* → Gk. *nuks/nuxt*-; cf. also **k^wek^wlos*, “circle” → PGk *k^wuk^wlos* → Gk. *kuklos*; etc.

Later sound changes between Proto-Greek and the attested Mycenaean include:

1. Introduction

- Loss of final stop consonants; final $m \rightarrow n$.
- Syllabic $m \rightarrow am$, and $\eta \rightarrow an$, before resonants; otherwise both were nasalized $m/\eta \rightarrow \tilde{a} \rightarrow a$.
- loss of s in consonant clusters, with supplementary lengthening, e.g. $esmi \rightarrow \tilde{e}mi$.
- creation of secondary s from clusters, $ntja \rightarrow nsa$. Assibilation $ti \rightarrow si$ only in southern dialects.
- Mycenaean i-vocalism and replacement of double-consonance $-kw-$ for $-k^wk^w-$.

NOTE. On the problematic case of common Greek ἵππος (*hippos*), *horse*, derived from PIE and PGk *ekwos*, Meier-Brügger (2003): “the i-vocalism of which is best understood as an inheritance from the Mycenaean period. At that time, *e* in a particular phonetic situation must have been pronounced in a more closed manner, cf. *di-pa* i.e. *dipas* neuter ‘lidded container for drinking’ vs. the later δῆρας (since Homer): Risch (1981), O. Panagl (1989). That the i-form extended to the entire Greek region may be explained in that the word, very central during Mycenaean rule of the entire region (2nd millennium BC), spread and suppressed the *e*-form that had certainly been present at one time. On the *-pp-*: The original double-consonance $-k\underline{u}-$ was likely replaced by $-k^wk^w-$ in the pre-Mycenaean period, and again, in turn by *-pp-* after the disappearance of the labiovelars. Suggestions of an ancient $-k^wk^w-$ are already given by the Mycenaean form as *i-qa* (a possible $*i-ko-wo$ does not appear) and the noted double-consonance in alphabetic Greek. The aspiration of the word at the beginning remains a riddle”.

Other features common to the earliest Greek dialects include:

- The PIE dative, instrumental and locative cases were syncretized into a single dative.
- Dialectal nominative plural in *-oi*, *-ai* fully replaces Late PIE common $*-\bar{o}s$, $*-\bar{a}s$.
- The superlative on *-tatos* (<PIE $*-t\underline{m}-to-s$) 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.
- The first person middle verbal desinences *-mai*, *-mān* replace *-ai*, *-a*. The third singular *pherei* is an analogical innovation, replacing the expected PIE $*b^h\underline{e}reti$, i.e. Dor. $*phereti$, Ion. $*pheresi$.
- 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 also common to Greek dialects.

II. ARMENIAN

Armenian is an Indo-European language spoken in the Armenian Republic, as well as in the region of Nagorno-Karabakh, and also used by ethnic Armenians in the Diaspora.

Armenian has been traditionally regarded as a close relative of Phrygian, apparently closely related to Greek, sharing major isoglosses with it. The *Graeco-Armenian hypothesis* proposed 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.

NOTE. That traditional linguistic theory, proposed by Pedersen (1924), establishes a close relationship between both original communities, Greek and Armenian, departing from a common subdialect of IE IIIa (Southern Dialect of Late PIE). That vision, accepted for a long time, was rejected by Clackson (1994) in *The linguistic relationship between Armenian and Greek*, which, supporting the *Graeco-Aryan* linguistic hypothesis, dismisses that the coincidences between Armenian and Greek represent more than those found in the comparison between any other IE language pair. Those findings are supported by Kortlandt in *Armeniaca* (2003), in which he proposes an old Central IE *continuum* Daco-Albanian / Graeco-Phrygian / Thracio-Armenian. Adrados (1998), considers an older Southern *continuum* Graeco-[Daco-]Thracio-Phrygian / Armenian / Indo-Iranian. Olteanu (2009) proposes a Graeco-Daco-Thracian language.

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



Distribution of ethnic Armenians in the 20th c.

subject of much speculation. It is clear that Armenian is an Indo-European language, but its development is opaque.

NOTE. Proto-Armenian sound-laws are varied and eccentric, such as IE **dw-* yielding Arm. *k-*, and in many cases still uncertain. In fact, that phonetic development is usually seen as **dw-* to *erk-*, based on PIE numeral **dwo-*, “two”, a reconstruction Kortlandt (*ibidem*) dismisses, exposing alternative etymologies for the usual examples.

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 Proto-Indo-European. In certain contexts, these aspirated stops are further reduced to *w*, *h* or zero in Armenian – so e.g. PIE **p^hots*, into Arm. *otn*, Gk. *pous*, “foot”; PIE **t^hreis*, Arm. *erek'*, Gk. *treis*, “three”.



Armenian manuscript, ca. 5th-6th c.

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.

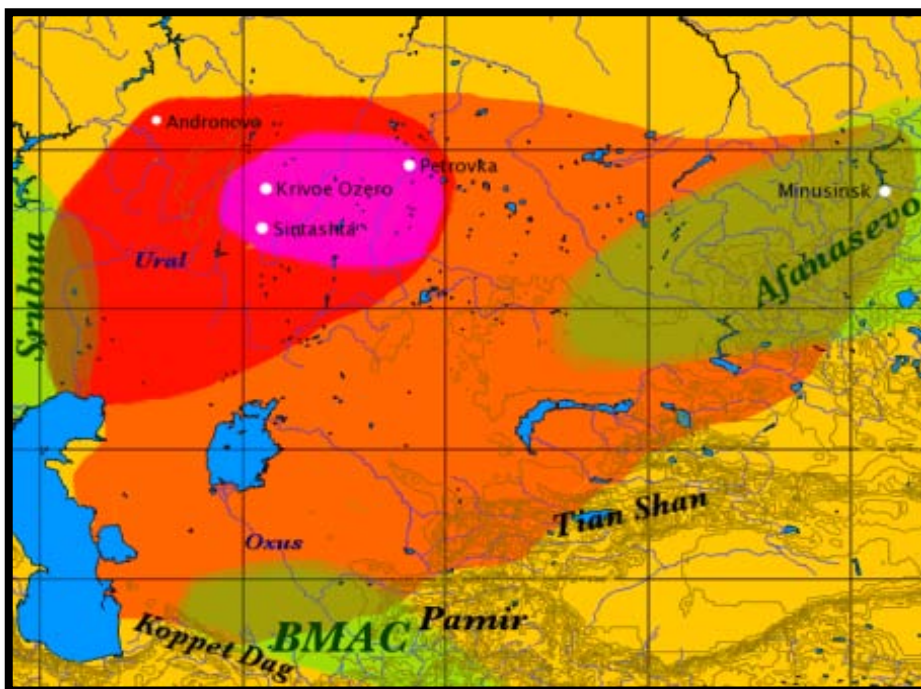
Grammatically, early forms of Armenian had much in common with classical Greek and Latin, but the modern language (like Modern Greek) has undergone many transformations. Interestingly enough, it shares with Italic dialects the secondary IE suffix **-tjōn*, extended from **-ti-*, cf. Arm թյուն (*t'youn*).

III. INDO-IRANIAN

The Indo-Iranian or Aryan language group constitutes the easternmost extant branch of the Indo-European family of languages. It consists of two main language groups, Indo-Aryan and Iranian, and probably Nuristani; Dardic is usually classified within the Indic subgroup.

The contemporary Indo-Iranian languages form therefore the second largest sub-branch of Indo-European (after North-West Indo-European), with more than one billion speakers in total, stretching from Europe (Romani) and the Caucasus (Ossetian) to East India (Bengali and Assamese). The largest in terms of native speakers are 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).

Map of the Sintashta-Petrovka culture (red), its expansion into the Andronovo culture during the 2nd millennium BC, showing the overlap with the BMAC in the south. The location of the earliest chariots is shown in purple.



Proto-Indo-Iranians are commonly identified with the bearers of the Andronovo culture and their homeland with an area of the Eurasian steppe that borders the Ural River on the west, the Tian Shan on the east – where the Indo-Iranians took over the area occupied by the earlier Afanasevo culture –, and Transoxiana and the Hindu Kush

1. Introduction

on the south. Historical linguists broadly estimate that a continuum of Indo-Iranian languages probably began to diverge by 2000 BC, preceding both the Vedic and Iranian cultures. A Two-wave model of Indo-Iranian expansion have been proposed (see Burrow 1973 and Parpola 1999), strongly associated with the chariot.

Aryans spread into the Caucasus, the Iranian plateau, and South Asia, as well as into Mesopotamia and Syria, introducing the horse and chariot culture to this part of the world. Sumerian texts from EDIIIb Ngirsu (2500-2350 BC) already mention the 'chariot' (*gigir*) and Ur III texts (2150-2000 BC) mention the horse (*anshe-zi-zi*). They left linguistic remains in a Hittite horse-training manual written by one "Kikkuli the Mitannian". Other evidence is found in references to the names of Mitanni rulers and the gods they swore by in treaties; these remains are found in the archives of the Mitanni's neighbors, and the time period for this is about 1500 BC.

The standard model for the entry of the Indo-European languages into South Asia is that the First Wave went over the Hindu Kush, either into the headwaters of the Indus and later the Ganges. The earliest stratum of Vedic Sanskrit, preserved only in the Rigveda, is assigned to roughly 1500 BC. From the Indus, the Indo-Aryan languages spread from ca. 1500 BC to ca. 500 BC, over the northern and central parts of the subcontinent, sparing the extreme south. The Indo-Aryans in these areas established several powerful kingdoms and principalities in the region, from eastern Afghanistan to the doorstep of Bengal.

The Second Wave is interpreted as the Iranian wave. The Iranians would take over all of Central Asia, Iran, and for a considerable period, dominate the European steppe (the modern Ukraine) and intrude north into Russia and west into central and eastern Europe well into historic times and as late as the Common Era. The first Iranians to reach the Black Sea may have been the Cimmerians in the 8th century BC, although their linguistic affiliation is uncertain. They were followed by the Scythians, who are considered a western branch of the Central Asian Sakas, and the Sarmatian tribes.

The Medes, Parthians and Persians begin to appear on the Persian plateau from ca. 800 BC, and the Achaemenids replaced Elamite rule from 559 BC. Around the first millennium of the Common Era, the Iranian Pashtuns and Baloch began to settle on the eastern edge of the Iranian plateau, on the mountainous frontier of northwestern

Pakistan in what is now the North-West Frontier Province and Balochistan, displacing the earlier Indo-Aryans from the area.

The main changes separating Proto-Indo-Iranian from Late PIE include:

- Early *Satemization* trend:
 - Loss of PIE labiovelars into PII plain velars: $*k^w \rightarrow k$, $*g^w \rightarrow g$, $*g^wh \rightarrow g^h$.
 - Palatalization of PII velars in certain phonetic environments: $*k \rightarrow \check{k}$, $*g \rightarrow \check{g}$, $*g^h \rightarrow \check{g}^h$.
- Loss of laryngeals: $*HV \rightarrow a$, $*VH \rightarrow \bar{a}$. Interconsonantal $*H \rightarrow i$, cf. $*ph_2t\acute{e}r \rightarrow$ PII $pit\acute{a}r$.

NOTE. A common exception is the Brugmann's law. For those linguists who consider the laryngeal loss to have occurred already in Late PIE, Aryan vocalism is described as a collapse of PIE ablauting vowels into a single PII vowel; i.e. $*e, *o \rightarrow a$; $*\bar{e}, *\bar{o} \rightarrow \bar{a}$.

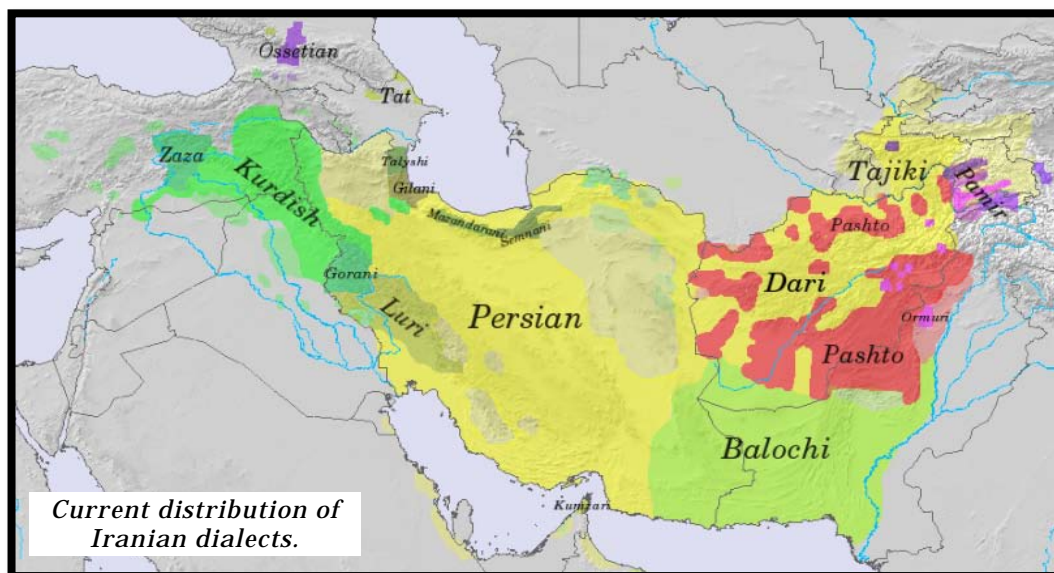
- Grassmann's law, Bartholomae's law, and the Ruki sound law were complete in PII.

NOTE. For a detailed description of those Indo-Iranian sound laws and the "*satemization*" process, see Appendix II. For Ruki sound law, v.s. Baltic in §1.7.1.

- Sonorants are generally stable in PII, but for the confusion $*l/*r$, which in the oldest Rigveda and in Avestan gives a general PIE $*l \rightarrow$ PII r , as well as $l \rightarrow r$.

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 PIE voiced aspirates.

A. IRANIAN



Current distribution of Iranian dialects.

1. Introduction

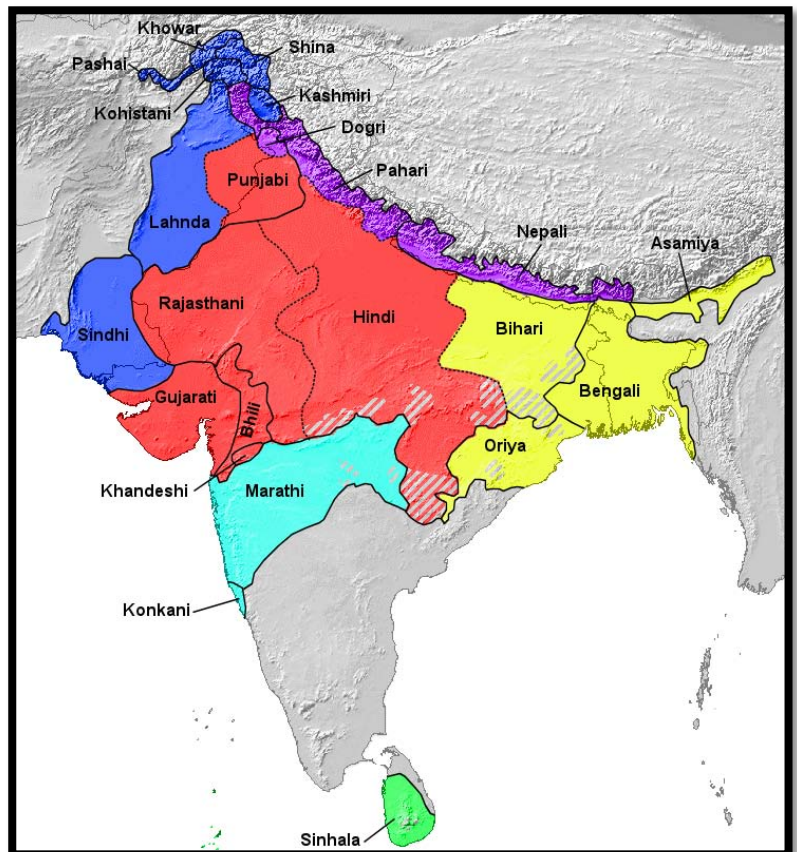
The Iranian languages are a branch of the Indo-Iranian subfamily, with an estimated 150-200 million native speakers today, the largest being Persian (ca. 60 million), Kurdish (ca. 25 million), Pashto (ca. 25 million) and Balochi (ca. 7 million).

Proto-Iranian dates to some time after the Proto-Indo-Iranian breakup, or the early second millennium BC, as the Old Iranian languages began to break off and evolve separately as the various Iranian tribes migrated and settled in vast areas of southeastern Europe, the Iranian plateau, and Central Asia. The oldest Iranian language known, Avestan, is mainly attested through the Avesta, a collection of sacred texts connected to the Zoroastrian religion.

Linguistically, the Old Iranian languages are divided into two major families, the Eastern and Western group, and several subclasses. The so-called Eastern group includes Scythian, even though the Scyths lived in a region extending further west than the Western group. The northwestern branch included Median, and Parthian, while the southwestern branch included Old Persian.

B. INDO-ARYAN

The Indo-Aryan or Indic languages are a branch of the Indo-Iranian subfamily with a total number of native speakers of more than 900 million. The largest languages in terms of native speakers are Hindustani (about 540 million), Bengali (about 200 million), Punjabi (about 100 million), Marathi



(about 90 million), Gujarati (about 45 million), Nepali (about 40 million), Oriya (about 30 million), Sindhi (about 20 million) and Assamese (about 14 million).

The earliest evidence of the group is from Vedic Sanskrit, the language used in the ancient preserved texts of the Indian subcontinent, the foundational canon of Hinduism known as the Vedas. The Indo-Aryan superstrate in Mitanni is of similar age as the Rigveda, but the only evidence is a number of loanwords.

In the 4th c. BC, the Sanskrit language was codified and standardised by the grammarian Panini, called “Classical Sanskrit” by convention. Outside the learned sphere of Sanskrit, vernacular dialects (Prakrits) continued to evolve and, in medieval times, diversified into various Middle Indic dialects.

C. NURISTANI

The recent view is to classify Nuristani as an independent branch of the Indo-Iranian language family, instead of the the Indic or Iranian group. In any event, it would seem they arrived in their present homeland at a very early date, and never entered the western Punjab of Pakistan.

1.7.3. OTHER INDO-EUROPEAN DIALECTS OF EUROPE

I. ALBANIAN

Albanian is 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 Albanian language has no living close relatives among the



Albanian dialects Gheg, Tosk. Communities of Arbëreshë- and Arvanitika-speakers

1. Introduction

modern IE languages. There is no consensus over its origin and dialectal classification, although some scholars derive it from Illyrian, and others claim that it derives from Thracian.

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.

NOTE. 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 14th c. AD, 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 Durrës in 1462 in the Gheg dialect, and some New Testament verses from that period.

II. PALEO-BALKAN LANGUAGES

A. PHRYGIAN

The Phrygian language was the IE language spoken by the Phrygians, a people that settled in Asia Minor during the Bronze Age. It survived probably into the 6th 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 migrated 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.



Phrygian region and expanded Kingdom.

1. Introduction

Phrygian is attested by two corpora, one, Palaeo-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 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.

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 migrated 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.

Phrygian seems to exhibit an augment, like Greek and Armenian, as in Phryg. *eberet*, probably corresponding to PIE **é-b^her-e-t* (cf. Gk. *epheret*).



Phrygian inscription in Midas City.

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.

Phrygian words with possible PIE origin and Graeco-Armenian cognates include:

- Phryg. *bekos*, “bread”, from PIE **b^heh₃g-*; cf. Gk. *p^hōgō*, “to roast”.
- Phryg. *bedu*, “water”, from PIE **wed-*; cf. Arm. *get*, “river”.
- Phryg. *anar*, “husband”, “man”, PIE **h₂ner-*, “man”; cf. Gk. *aner-*, “man, husband”.
- Phryg. *belte*, “swamp”, from PIE root **b^hel-*, “to gleam”; cf. Gk. *baltos*, “swamp”.
- Phryg. *brater*, “brother”, from PIE **b^hreh₂ter-*; cf. Gk. *p^hrāter-*.
- Phryg. *ad-daket*, “does, causes”, from PIE stem **d^hē-k-*; cf. Gk. *et^hēka*.
- Phryg. *germe*, “warm”, from PIE **g^wher-mo-*; cf. Gk. *t^hermos*.
- Phryg. *gdan*, “earth”, from PIE **d^hg^hom-*; cf. Gk. *k^ht^hōn*.

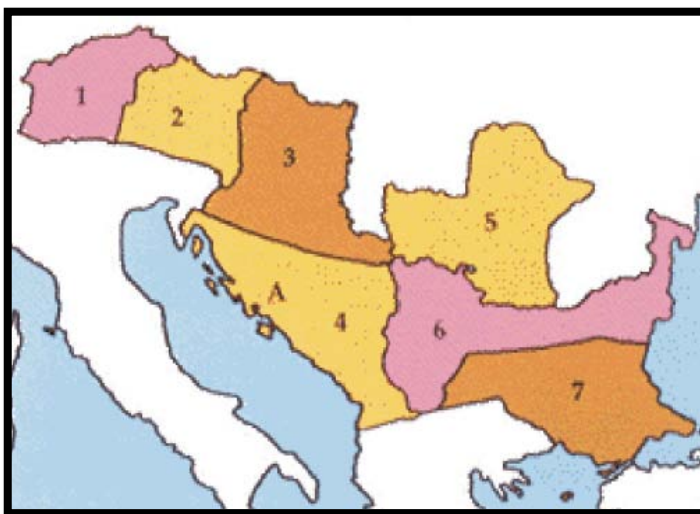
NOTE. For more information on similarities between Greek and Phrygian, see Neumann *Phrygisch und Griechisch* (1988).

B. 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 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. Some sound-changes and other



Roman provinces in the Balkans, 2nd century AD: A. Spalatum (Split); 1. Raetia; 2. Noricum; 3. Pannonia; 4. Illyricum; 5. Dacia; 6. Moesia; 7. Tracia.

1. Introduction

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.

NOTE. 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. A grouping of Illyrian with the Thracian and Dacian language in a “Thracico-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. Also, the hypothesis that the modern Albanian language is a surviving Illyrian language remains very controversial among linguists.

B. THRACIAN

Excluding Dacian, whose status as a Thracian language is disputed, Thracian was spoken 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). 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 AD.

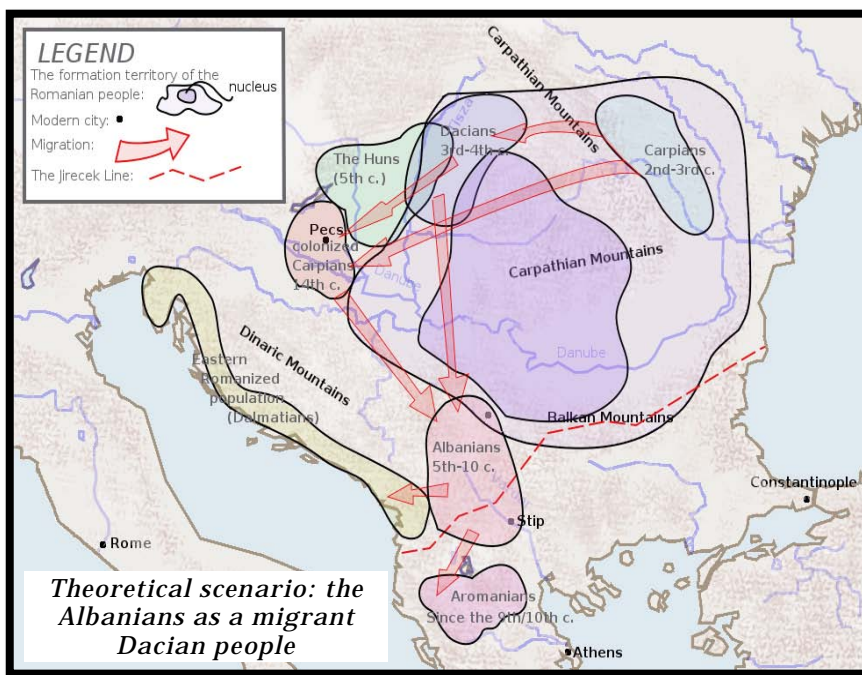
NOTE. As an extinct language with only a few short inscriptions attributed to it (*v.i.*), there is little known about the Thracian language, but a number of features are agreed upon. A number of probable Thracian words are found in inscriptions – most of them written with Greek script – on buildings, coins, and other artifacts. Some Greek lexical elements may derive from Thracian, such as *balios*, “*dappled*” (< PIE **b^hel-*, “*to shine*”, Pokorny also cites Illyrian as possible source), *bounos*, “*hill, mound*”, etc.

C. DACIAN

The Dacian language was an Indo-European language spoken by the ancient people of Dacia. It is often considered to have been either 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 of Slavic languages.

According to the hypothesis of Hasdeu (1901), a branch of Dacian continued as the Albanian language. A refined version of that hypothesis considers Albanian to be a Daco-Moesian Dialect that split off before 300 BC, and that Dacian became extinct.

NOTE. The arguments for this early split before 300 BC include:

- Inherited Albanian words (e.g. PIE **mātēr* → Alb. *motër*) shows the evolution PIE **ā* → Alb. *o*, but all the Latin loans in Albanian having an *ā* (<PIE **ā*) shows Lat. *ā* → Alb. *a*. Therefore, the transformation happened and ended before the Roman arrival in the Balkans.
- Romanian substratum words shared with Albanian show a Rom. *a* that corresponds to Alb. *o* when the source for both sounds is an original common *ā* (cf. *mazāre/modhull*<**mādzula*,

1. Introduction

“pea”; *raṭã/rosë* < **rātja*: “duck”); therefore, when these words had the same common form in Pre-Romanian and Proto-Albanian the transformation PIE **ā* → Alb. *o* had not started yet.

The correlation between these two facts could indicate that the split between Pre-Romanian (the Dacians later Romanized) and Proto-Albanian happened before the Roman arrival in the Balkans.

E. 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.

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.

NOTE. 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 distinction and does not appear to have undergone Satemization. The Indo-European voiced aspirates became plain voiced consonants, i.e. **b^h* → *b*, **d^h* → *d*, **g^h* → *g*, **g^{wh}* → *g^w*; as in Illyrian, Thracian, Macedonian and Phrygian (but unlike Greek).

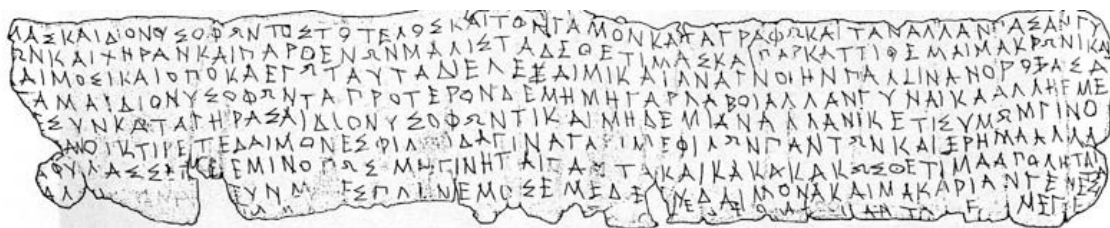
F. ANCIENT MACEDONIAN

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.



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

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.”

From the few words that survive, a notable sound-law may be ascertained, that PIE voiced aspirates **d^h*, **b^h*, **g^h*, appear as δ (=d^h), β (=b^h), γ (=g^h), in contrast to Greek dialects, which unvoiced them to θ (=t^h), φ (=p^h), χ (=k^h).

NOTE. Since these languages are all known via the Greek alphabet, which has no signs for voiced aspirates, it is unclear whether de-aspiration had really taken place, or whether the supposed voiced stops β, δ, γ were just picked as the closest matches to express voiced aspirates PIE **b^h*, **d^h*,

1. Introduction

*g^h. As to Macedonian β, δ, γ = Greek φ, θ, χ, Claude Brixhe[(1996) suggests that it may have been a later development: The letters may already have designated not voiced stops, i.e. [b, d, g], but voiced fricatives, i.e. [β, δ, γ], due to a voicing of the voiceless fricatives [φ, θ, χ] (= Classical Attic [p^h, t^h, k^h]). Brian Joseph (2001) sums up that “The slender evidence is open to different interpretations, so that no definitive answer is really possible”, but cautions that “most likely, Ancient Macedonian was not simply an Ancient Greek dialect on a par with Attic or Aeolic”. In this sense, some authors also call it a “deviant Greek dialect”.

- PIE *d^henh₂-, “to leave”, → A.Mac. δανός (*danós*), “death”; cf. Attic θάνατος (*t^hánatos*).
PIE *h₂aid^h- → A.Mac. *ἄδραια (*adraia*), ‘bright weather’, Attic αἰθρία (*ait^hria*).
- PIE *b^hasko- → A.Mac. βάσκοιοι (*βάskioi*), “fasces”. Compare also for A.Mac. ἀβροῦτες (*abroūtes*) or ἀβροῦφες (*abroūwes*), Attic ὄφρυς (*op^hrūs*), “eyebrows”; for Mac. Βερενίκη (*Berenikē*), Attic Φερενίκη (*P^herenikē*), “bearing victory”.
 - According to Herodotus (ca. 440 BC), the Macedonians claimed that the *Phryges* were called *Brygoi* (<PIE *b^hrugo-) before they migrated from Thrace to Anatolia ca. 1200 BC.
 - 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 (*kep^halē*), “head”.
- If A.Mac. γοτάν (*gotán*), “pig”, is related to PIE *g^wou-, “cow”, this would indicate that the labiovelars were either intact (hence *g^wotán), or merged with the velars, unlike the usual Gk. βοῦς (*boûs*).

NOTE. Such deviations, however, are not unknown within Greek dialects; compare Dor. γλεπ- (*glep-*) for common Gk. βλεπ- (*blep-*), as well as Dor. γλάχων (*glāchōn*) and Ion. γλήχων (*glēchōn*) for Gk. βλήχων (*blēchōn*).

- Examples suggest that voiced velar stops were devoiced, especially word-initially: PIE *genu- → A.Mac. κánaδοι (*kánadoi*), “jaws”; PIE *gomb^h- → A.Mac. κόμβους (*kómbous*), “molars”.
 - Compared to Greek words, there is A.Mac. 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. also the Thracian name *Agassamenos*.

1.7.4. ANATOLIAN LANGUAGES

The Anatolian languages are a group of extinct Indo-European languages, which were spoken in Anatolia for millennia, 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 or “Proto-Indo-Hittite” (PIH), typically a date ca. 4500-3500 BC is assumed. Within 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.



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

NOTE. The term *Indo-Hittite* 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.

Attested dialects of the Anatolian branch are:

- Hittite (*nesili*), attested from ca. 1800 BC to 1100 BC, official language of the Hittite Empire.
- Luwian (*luwili*), close relative of Hittite spoken in Arzawa, to the southwest of the core Hittite area.
- 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, most likely a descendant of Luwian, extinct in ca. the 1st century BC. A fragmentary language, it is also a likely candidate for the language spoken by Trojans.

1. Introduction

- 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 Anatolian branch that have left no written records, such as the languages of Mysia, Cappadocia and Paphlagonia.

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 and written in an adapted form of Old Assyrian



Hittite pictographic writing

cuneiform orthography. Owing to the predominantly syllabic nature of the script, it is difficult to ascertain the precise phonetic qualities of some Hittite sounds.

NOTE. 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.

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

Luwian was spoken by population groups in Arzawa, to the west or southwest of the core Hittite area. In the oldest texts, e.g. the Hittite Code, the Luwian-speaking areas including Arzawa and Kizzuwatna were called Luwia. From this homeland, Luwian speakers gradually spread through Anatolia and became a contributing factor to the

downfall, after circa 1180 BC, of the Hittite Empire, where it was already widely spoken. Luwian was also the language spoken in the Neo-Hittite states of Syria, such as Milid and Carchemish, as well as in the central Anatolian kingdom of Tabal that flourished around 900 BC. Luwian has been preserved in two forms, named after the writing systems used: Cuneiform Luwian and Hieroglyphic Luwian.



Luwian use according to inscriptions found

For the most part, the immediate ancestor of the known Anatolian languages, Common Anatolian (the Late Proto-Anatolian dialect spoken ca. 2500) has been reconstructed on the basis of Hittite. However, the usage of Hittite cuneiform writing system limits the enterprise of understanding and reconstructing Anatolian phonology, partly due to the deficiency of the adopted Akkadian cuneiform syllabary to represent Hittite sounds, and partly due to the Hittite scribal practices.

NOTE 1. This especially pertains to what appears to be confusion of voiceless and voiced dental stops, where signs -dV- and -tV- are employed interchangeably different attestations of the same word. Furthermore, in the syllables of the structure VC only the signs with voiceless stops are generally used. Distribution of spellings with single and geminated consonants in the oldest extant monuments indicates that the reflexes of PIE voiceless stops were spelled as double consonants and the reflexes of Proto-Indo-European voiced stops as single consonants. This regularity is the most consistent in in the case of dental stops in older texts; later monuments often show irregular variation of this rule.

NOTE 2. For a defence of Etruscan as an IE language, classified within the Anatolian branch, see Adrados (2005) at <<http://emerita.revistas.csic.es/index.php/emerita/article/viewArticle/52>>.

Known changes from Middle PIE into Common Anatolian include:

- Voiced aspirates merged with voiced stops: *d^h→d, *b^h→b, *g^h→g.

1. Introduction

- Voiceless stops become voiced after accented long-vowel or diphthong: PIH **wēk-* → CA *wēg-* (cf. Hitt. *wēk-*, “ask for”); PIH **d^heh₁ti*, “putting” → CA *dādi* (cf. Luw. *taac-* “votive offering”).
- Conditioned allophone PIH **tj-* → CA *tsj-*, as Hittite still shows.
- PIH **h₁* is lost in CA, but for **eh₁→ā*, appearing as Hitt., Pal. *ē*, Luw., Lyc., Lyd. *ā*; word-initial **h₂→x*, non-initial **h₂→h*; **h₃→h*.

NOTE 1. Melchert proposes that CA *x* (voiceless fricative) is “lenited” to *h* (voiced fricative) under the same conditions as voiceless stops. Also, word-initial **h₃* is assumed by some scholars to have been lost already in CA.

NOTE 2. There is an important assimilation of laryngeals within CA: a sequence -VRHV- becomes -VRRV-; cf. PIH **sperh₁V-* → Hitt. *isparr-*, “kick flat”; PIH **sun-h₃-V-* → Hitt. *sunna-*, “fill”, Pal. *sunnuttil-*, “outpouring”; etc.

- PIH sonorants are generally stable in CA. Only word-initial **r̥* has been eliminated. Word-initial **je-* shows a trend to become CA *e-*, but the trend is not complete in CA, as Hittite shows.
- Diphthong evolved as PIH **ei* → CA long *ē*; PIH **eu* → CA *ū*. PIE **oi*, **ai*, **ou*, **au*, appear also in CA.

NOTE. Common Anatolian preserves PIE vowel system basically intact. Some cite the merger of PIH **o* and (controversial) **a* as a Common Anatolian innovation, but according to Melchert that merger was secondary shared innovation in Hittite, Palaic and Luwian, but not in Lycian. Also, the lengthening of accented short vowels in open syllables cannot be of Common Anatolian, and neither can lengthening in accented closed syllables.

- The CA nominal system shows an archaic productive declension in **-i*, **-u*. There are only two grammatical genders, animate and inanimate.
- Hittite verbs are inflected according to two general verbal classes, the *mi-* and the *hi-* conjugation.

NOTE. 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”. Additionally, the Hittite verbal system displays two voices (active and mediopassive), two moods (indicative and imperative), and two tenses (present and preterite), two infinitive forms, one verbal substantive, a supine, and a participle.

1.8. MODERN INDO-EUROPEAN

1.8.1. Modern Indo-European (MIE) is therefore a set of grammatical rules – including its writing system, noun declension, verbal conjugation and syntax –, designed to systematize the reconstructed Proto-Indo-European dialect North-West Indo-European – described (v.s.) as the last *IE dialect continuum* (spoken in Europe for some centuries within the time frame 3000-2000 BC) – to adapt it to modern communication needs.

Because such PIE dialects were spoken by prehistoric societies, no genuine sample texts are available, and thus comparative linguistics – in spite of its 200 years' history – is not (and will not be) in the position to reconstruct exactly their formal languages (the one used by learned people at the time), but only approximately how the spoken, vulgar languages were like, i.e. the proto-languages that later evolved into the different attested Indo-European dialects and languages.

NOTE. Reconstructed languages like Modern Hebrew, Modern Cornish, Modern Coptic, Modern Prussian 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*' natural languages instead of 'universal' ones, what is certain is that if somebody learns Hebrew, Cornish, Coptic, Prussian or Indo-European (or Latin, Gothic, Greek, Sanskrit, 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. That cannot be said of personal inventions.

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 for a common language within the EU, to substitute the current deficient linguistic policy. This language system, called *European* or *European language* (**eurōpājóm**), is mainly based on the features of the European or Northwestern IE dialects, whose speakers – as we have already seen – remained in close contact for some centuries after the first Late PIE migrations, and have influenced each other in the last millennia within Europe.

NOTE. As Indo-Europeanist F. López-Menchero (2008) puts it, "there are 'three (Late) Proto-Indo-European languages' which might be distinguished today:

1. Introduction

1) The actual Proto-Indo-European language and its early dialects, spoken by prehistoric peoples of Eurasia in the Neolithic and Bronze Age, some millennia ago;

2) the reconstructed Late Proto-Indo-European language system, which has been studied by IE scholars using the linguistic, archaeological and historical data available, and which is (and will remain) imperfect by nature, based on more or less certain hypotheses and schools of thought; and

3) the modern Indo-European language systems (European, Hellenic, Aryan) which, being based on the later, and trying to come near to the former, are neither one nor the other, but modern languages systematized to be used in the modern world”.

NOTE 2. In that sense, some critics have considered the so-called “*Indo-European language revival*” to be different from (and thus not comparable to) other language revivals, like – as they put it – Hebrew or Cornish, because of the ‘obvious differences that will exist between that ancient North-West Indo-European language and the Modern Indo-European or European language’. It is important to note that, even though there is a general belief that Modern Hebrew and Ancient Hebrew are the same languages, among Israeli scholars there have been continued calls for the “Modern Hebrew” language to be called “Israeli Hebrew” or just (preferably) “Israeli”, due to the strong divergences that exist – and further develop with its use – between the modern language spoken in Israel and its theoretical basis, the Ancient Hebrew of the Tanakh, its contents (and language variations) having being compiled probably between 450-200 BC, i.e when the language was being substituted by Aramaic. On that interesting question, Prof. Ghil’ad Zuckermann considers that “Israelis are brainwashed to believe they speak the same language as the prophet Isaiah, a purely Semitic language, but this is false. It’s time we acknowledge that Israeli is very different from the Hebrew of the past”. He points out to the abiding influence of modern Indo-European dialects – especially Yiddish, Russian and Polish –, in vocabulary, syntax and phonetics, as imported by Israel’s founders. The same could certainly be said of Cornish and other language revivals, and even of some death languages with a continued use, like the Modern Latin language used by the Catholic Church, which is not comparable to the Classical Latin used by Cicero, not to talk about the real, Vulgar Latin used by the different peoples who lived in the Roman Empire.

1.8.3. Late Proto-Indo-European features that are common to early PIE dialects (mainly North-West IE, Proto-Greek and Proto-Indo-Iranian), like nominal and verbal inflection, morphology and syntax, make it possible for PIE to be proposed as *Dachsprache* for an *Indo-European International Auxiliary Language* project. Obviously, French, German, Spanish, Hindustani, Chinese, and other natural and artificial languages proposed to substitute English dominance, are only supported by their cultural or social communities, whereas IE native 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.

NOTE 1. Because Modern Indo-European (a revived North-West IE proto-language) has other sister dialects that were spoken by coeval prehistoric communities, languages like Modern Hellenic (a revived Proto-Greek) and Modern Aryan (a revived Proto-Indo-Iranian) can also be used in the regions where they are currently spoken in the form of their surviving dialects, as those proto-languages were not much more different from North-West IE than Swedish from Danish, or Spanish from Portuguese. They might also serve as *linguae francae* for closely related languages or neighbouring regions, i.e. Aryan for Asia, Hellenic for and Armenian-speaking territories.

Anatolianism (Turkish *Anadoluculuk*) asserts that Turks descend from the indigenous population of ancient Anatolia, based on historical and genetic views. Supported by Turkish intellectuals in the 20th century, it became essential to the process of nation-building in Turkey, but was substituted by the Pan-Turkic nationalism Mustafa Kemal Atatürk had discouraged before his death. If accepted again, Turks could embrace their historical culture by adopting Modern Anatolian (a revived Common Anatolian, "*cousin dialect*" of EIE, PGk. and PII) as a modern second language for Turkey, which shares close historical and cultural ties with Europe and Asia.

NOTE 2. The terms *Ausbausprache-Abstandsprache-Dachsprache* were coined by Heinz Kloss (1967), and they are designed to capture the idea that there are two separate and largely independent sets of criteria and arguments for calling a variety an independent "language" rather than a "dialect": the one based on its social functions, and the other based on its objective structural properties. A variety is called an *ausbau language* if it is used autonomously with respect to other related languages. This typically means that it has its own standardized form independent of neighbouring standard languages, like (in this hypothetical future) Modern Indo-European in Europe and the Americas, Modern Aryan in Asia. This often involves being taught in schools, and being used as a written language in a wide variety of functions, possibly including that of an official national language. In contrast, varieties that are *abstand languages* are those that are only spoken and typically only used in private contexts.

Dachsprache means a language form that serves as standard language for different dialects, even though these dialects may be so different that mutual intelligibility is not possible on the basilectal level between all dialects, particularly those separated by significant geographical distance. So e.g. the *Rumantsch Grischun* developed by Heinrich Schmid (1982) as such a *Dachsprache* for a number of quite different Romansh language forms spoken in parts of Switzerland; or the *Euskara Batua*, "Standard Basque", and the Southern Quechua literary standard, both developed as standard languages for dialect continua that had historically been thought of as discrete languages

1. Introduction

with many dialects and no “official” dialect. Standard German and standard Italian to some extent function in the same way. Perhaps the most widely used *Dachsprache* is Modern Standard Arabic, which links together the speakers of many different, often mutually unintelligible Arabic dialects. Hence a *Standard Indo-European*, which might take rules from Late Proto-Indo-European reconstruction and the Modern Indo-European rules presented here, would be the wide *Dachsprache* necessary to encompass (i.e. to serve as linguistic umbrella for) the modern revival of early PIE dialects.

NOTE 3. Our proposal is different from the Hebrew language revival, but we think that:

a) The reconstruction of a common Late PIE (laryngeal?) phonology, nominal or verbal inflection system results at best mainly in abstract *formulae* or vague approximations – following the dissertation of Mallory & Adams (v.s. § 1.1.8) –; they are very useful for a Standard Indo-European *Dachsprache*, but the reconstruction unfortunately does not have enough certainty to be used for a common, modern revived language. Reconstructions of early PIE dialects, on the other hand, result in approximations with strong statistical confidence, offering a practical system for common West European, Greek and Indo-Iranian phonetics and inflection system, but they lack enough data on their oldest morphology, syntax and vocabulary, which were obscured by later innovations. Therefore, reconstructions of Late PIE and early PIE dialects complement each other.

b) Where Zionism had only some formal writings, with limited vocabulary, of an ancient language already dead centuries before their latest sacred texts were compiled (ca. 200 BC), and their people expelled from Israel (in 70 AD), Pro-Europeanism and Indo-Europeanism have PIE and its early dialects (EIE, PGk and PII) with a continued history of use in Eurasia and hundreds of living dialects, and other very old dead dialects attested, so that their modern revival can be considered ‘less artificial’. Thus, even if Europeans had tablets dating from 2000 BC in some dialectal predominant formal EIE language (say, from Pre-Germanic or Pre-Celtic IE), the current North-West Indo-European reconstruction should probably still be used as the main source for Indo-European language revival in the European Union. Just taking a look at Mycenaean inscriptions and its difficult phonetic decipherment is enough to realize how little EIE reconstruction would change if writings were found.

c) The common culture and religion was probably the basis for the Hebrew language revival in Israel. Proto-Indo-European, whilst the mother tongue of some prehistoric tribe with an own culture and religion, spread into different peoples, with different cultures and religions. There was never a concept of “*Indo-European community*” after the migrations. However, Indo-European languages are spoken today by the majority of the population – in the world and especially within Europe –, and its early dialects spread into two main communities, EIE and PGk in Europe, PII in

South Asia. It is therefore possible to speak them as *natural*, *cultural* and *national* common languages, what may be a significant advantage of IE as IAL over any other natural language.

Also, blessed Anne Catherine Emmerich (1774-1824), a Roman Catholic Augustinian nun, stigmatic, mystic, visionary and ecstatic, had revelations about the Confusion of Tongues (Genesis 11:1-9): “*Upon Heber who, as we have said, took no part in the work, God cast His eyes; and amid the general disorder and corruption, He set him and his posterity apart as a holy nation. God gave him also a new and holy language possessed by no other nation, that thereby his race should be cut off from communication with all others. This language was the pure Hebrew, or Chaldaic [=NW Semitic]. The first tongue, the mother tongue,*



Gustave Doré's Confusion of Tongues

*spoken by Adam, Sem, and Noe, was different, and it is now extant only in isolated dialects. Its first pure offshoots are the Zend, the sacred tongue of India, and the language of the Bactrians [=early Indo-Iranian dialects]. In those languages, words may be found exactly similar to the Low German of my native place [=any modern EIE regional dialect]. The book that I see in modern Ctesiphon [=Greek colony, site of church councils of the Church of the East], on the Tigris, is written in that language [=a PGk dialect]”. Her visions receive particular veneration from Traditional Catholics, and this one refers clearly to the three “isolated dialects” (PII, EIE and PGk) derived directly from the *confounded* divine tongue – which some have since identified as the common Proto-Indo-European language – spoken until ca. 3000-2500 BC according to Biblical chronology and archaeological findings of the great temple towers (*ziggurats*) of ancient Sumer. For the Catholic Catechism, the Genesis “uses figurative language, but affirms a primeval event”, see <http://www.catholic.com/library/Adam_Eve_and_Evolution.asp>.*

1.8.4. Modern Indo-European words to complete the lexicon of North-West Indo-European, in case that no common PIE form is found, are to be loan-translated from present-day North-West IE languages. Common loan words from sister dialects can also be loan-translated or taken as (proto-language) loan words.

NOTE 1. Even though the vocabulary reconstructible for early PIE languages is indeed wider than the common Proto-Indo-European lexicon, a remark of Mallory & Adams (2006) regarding

1. Introduction

reconstructible Common PIE words is interesting: “Yet we know that our reconstructed lexicon falls far short of the full language, e.g. we can reconstruct ‘eye’ and ‘eyebrow’ but not ‘eyelash’. We can most easily gain an impression of what may be missing when we consider modern ethnobotanical studies. In Proto-Indo-European we can offer about thirty-two plant names and an additional twenty-six tree names. In contrast, Brent Berlin examined the languages of ten traditional farming societies and found that the average number of botanical taxa reported in each language was 520. If we were to treat such comparisons at face value this would suggest that we are recovering only about 11 per cent of the probable botanical lexicon known to the Proto-Indo-Europeans. Or compare, for example, the fact that we can reconstruct only a few terms relating to the horse in Proto-Indo-European; in English this semantic field includes horse, pony, nag, steed, prancer, dobbin, charger, courser, colt, foal, fielly, gelding, hack, jade, crock, plug, and many more terms, including the many specific terms describing the colour of the horse, e.g. bay, chestnut, sorrel, pinto. There is no reason to suspect that PIE did not behave similarly”.

NOTE 2. For examples of loan translations from modern EIE languages, cf. from Latin *aquaeduct* (Lat. *aquaeductus* → MIE **aqāsduktos**) or *universe* (Lat. *uniuersus* < **oin(i)-uors-o-* < **oino-wrt-to-* → MIE **oinówr̥stos** ‘turned into one’); from English, like *software* (from Gmc. *samþu-*, *warō* → MIE **somtúworā**); from French, like *ambassador* (from Cel. *amb(i)actos* → MIE **ambhiagtos** ‘public servant’); or *chamber* (from O.Lat. *camera*, from PGk. *kamarā*, ‘vault’ → MIE **kamarā**); from Russian, like *bolshevik* (MIE **belijówikos**); etc.

Modern loan words from sister or cousin IE dialects can be either loan-translated or directly taken as loan words, depending on the nature of the individual words:

- Loan words should be taken directly in MIE from forms which are found only in one proto-language or restricted to southern dialects; as e.g. Gk. *photo*, which should be taken directly as loan word **pháwotos**, from PGk *phawots*, gen. *phawotós*, as Gk. φῶς (<φάφος), φωτός, in compound **phawotogr̥phjā**, *photography*, derived from IE root **bhā-**, *shine*, which could be loan-translated as MIE ***bháwotos**, from ***bhawotogr̥bhjā**, but without having a meaning for extended **bha-wes-**, still less for **bha-wot-**, in North-West Indo-European or even Proto-Indo-European, as it is only found in Ancient Greek dialects. Compare also MIE **skholá**, from Lat. *schola*, taken from Gk. σχολή (<PGk. *skholá*), *spare time, leisure, tranquility*, borrowed from Greek with the meaning “*school*”, which was in O.Gk. σχολεῖον (*scholeíon*), translated as PGk. *skholehjom* (<PIE *-*esjo-m*), from IE root **segh-**, which could also be loan-translated as MIE ***sgholá** or even more purely (and artificially) ***sgholesjom**, none of them being Proto-Indo-European or common Indo-European terms. Examples from Indo-Iranian include **wasākáranas**, *bazaar*, from O.Ira. *vahacarana*, “*sale-traffic*”, *bazaar*, which could also be translated as proper MIE ***wesāqólenos**, from PIE roots **wes-**

and **qel-**; or **katúrangam**, *chess*, from Skr. *caturangam* (which entered Europe from Pers. *shatranj*) a *bahuvrihi* compound, meaning “having four limbs or parts”, which in epic poetry often means “army”, possibly shortened from **katurangabalam**, Skr. *caturangabalam*, lit. “four-member force”, “an army comprising of four parts”, could be loan-translated as MIE ***qatúrangom** and ***qaturangobelom**, from roots **qetwṛ-**, **ang-** and **bel-**.

- Loan words and loan translations might also coexist in specialized terms; as, from PIE **h₁rudʰrós*, *red*, PGk *eruthrós*, in common loan **eruthrókutos**, *erythrocyte*, proper MIE **rudhrós**, in **rudhrá** (**ésenos**) **kētjā**, *red (blood) cell*; cf. also MIE **mūs**, **musós**, *mouse*, *muscle*, PGk *mūs*, *muhós*, in loan **muhokutos**, *myocyte*, for **muskosjo kētjā**, *muscle cell*.

1.8.5. The adjective **eurōpājós**, m. *European*, comes from the Greek noun **Eurōpā**.

NOTE. Gk. *Eurōpā* is from unknown origin, even though it was linked with Homer’s epithet for Zeus *euruopā*, from **hurú-oqeh₂* “far-seeing, broad”, or **h₁urú-woqeh₂* “far-sounding” (Heath, 2005). Latinate adj. *europaeus*, which was borrowed by most European languages, comes from Gk. adj. *eurōpaíos*, in turn from PGk *eurōpai-jós* < PIE **eurōpeh₂-jós* → MIE **eurōpā-jós**. For the evolution PIH **-eh₂jo-* → PGk **-aijo-*, cf. adjective formation in Gk. *agorá-agoraíos*, Ruigh (1967).

The name of the European language system is **eurōpājóm**, inanimate, because of the oldest IE dialects, those which had an independent name for languages used the neuter.

NOTE. Compare Gk. n.pl. Ἑλληνικά (*hellēniká*), Skr. n.sg. संस्कृतम् (*saṃskṛtam*), O.H.G. *diutisc*, O.Prus. *prūsiskan*, etc.; cf. also in Tacitus Lat. *uōcābulum latīnum*.

In most IE languages, the language is also referred to as “language” defined by an adjective, whose gender follows the general rule of concordance; as in MIE f. **eurōpājā dnghwā**, *European language*.

NOTE. Cf. Lat. *latīna lingua*, Gk. ελληνική γλώσσα, O.H.G. *diutiska sprāhha* (Ger. *Deutsche Sprache*), O.Prus. *prūsiskai bilā*, O.C.S. словѣньскыи ѡзыкъ (*slověňskyyi jezýkjū*), etc.

1.8.6. Because the term *Indo-European* is common today to refer to the reconstructed language, we decided to use that traditional name to describe the Proto-European language, as a way to familiarize the reader with the European language system as a natural, dead language, and to distinguish it clearly from other language inventions.

NOTE. However, when speaking in European, **sindhueurōpājóm**, *Indo-European*, **p̄rmosindhueurōpājóm**, *Proto-Indo-European*, **Eurōpās sindhueurōpājóm**, *Europe’s Indo-European*, should refer to the theoretical linguistic concepts, to the ancient reconstructed dialects, while **eurōpājóm**, *European*, should be preferred for the modern language, just like *Israeli* is probably the most suited name to refer to Modern Hebrew.