

Modern Indo-European writing system

Carlos Quiles

Y P X Y T S q P M 7 0 田 4 7 J X | ⊗ E I 1 3 D 1 B A

Academia Prisca

2017

First Draft

Version 1 (October 2017)

© 2017 by Carlos Quiles

Academia Prisca

Full text: <<https://academiaprisca.org/>>

Correspondence: cquiles@academiaprisca.org



This work is licensed under the Creative Commons Attribution-ShareAlike 3.0 Unported License. To view a copy of this license, visit <<http://creativecommons.org/licenses/by-sa/3.0/>> or send a letter to Creative Commons, 171 Second Street, Suite 300, San Francisco, California, 94105, USA.

Table of Contents

Table of Contents	3
1. Introduction	5
2. Consonants	7
Other Indo-European phonemes (including laryngeals).....	7
Other common phonemes.....	8
Proto-Greek	8
Proto-Indo-Iranian	8
3. Vowels and vocalic allophones	9
4. Conventions.....	11
Accent.....	11
Vowel length: the apex and the macron	12
Resonants.....	13
Regressive assimilation of voicing and deaspiration.....	13
TD > DD.....	13
DT > TT (and DhT > TT).....	13
Laryngeals	13
Clitic words and compounds	14
Capital letters.....	14
5. Old Italic alphabet	15
Names	15
/g/, /g ^h /	15
/g ^w /, /g ^{wh} /.....	16
/k ^h /, /x/, /h/, and laryngeals	16

/d ^h / and /t ^h /	17
/b ^h /.....	17
/j/	17
Numerals.....	18
Typography.....	19
New Athena Unicode.....	19
Cardo	19
Aegean	19
Alphabetum	19
Segoe UI	19
Keyboard utilities	20
6. Latin alphabet	21
Typography.....	22
Telefon.....	22
Arial Unicode	22
Times New Roman	22
MS Reference, MS Reference Serif	22
7. Cyrillic alphabet	23
8. Greek alphabet.....	25
9. Perso-Arabic script	27
10. Brahmic scripts	29
11. Armenian alphabet.....	31

1. Introduction

This is the first writing convention for Modern Indo-European that includes the Old Italic script. This proposal is an experimental one, offering a recognizable different script for the reconstructed European language. It combines at the same time the disparate qualities of distinctness – of an archaic language compared to modern ones – and closeness – of a western European alphabet.

The Old Italic script originated from a western archaic Greek script, most likely all from the Cumaean alphabet, used in Ischia and Cumae (in the Bay of Naples) in the 8th century BC. It was brought to Italy by early Greek colonists using the Euboean alphabet (ca. 775-750 BC), possibly the first Greek alphabet, invented to record epic poetry¹.

These letters were adopted – albeit with local differences – to write Etruscan ca. 700 BC and Italic languages, e.g. Oscan ca. 7th c. (but the local variant was attested from ca. 5th c. BC), South Picene, and Archaic Latin ca. 7th c., from which the Old Latin alphabet was eventually derived. A common northern group of alphabets was formed by the Rhaetic, Venetic, Messapic, and Camunic alphabets, and included the alphabet of Lugano (used ca. 7th-5th c. BC). One alphabet of this latter group was most likely the origin of the Germanic runic alphabets.

Lepontic, written in the alphabet of Lugano, was probably a Celtic language related to Gaulish, but it could have been a Para-Celtic or Pre-Celtic Indo-European (i.e. Italo-Celtic) language. Venetic was a North-West Indo-European language, not clearly related more closely to either Italo-Celtic or Germanic. Messapic was an Indo-European language, maybe related to Illyrian, and therefore to Paleo-Balkan languages. The language written in the Camunic alphabet was possibly Indo-European, too.

Therefore, the Old Italic script was the first script used to write down North-West Indo-European languages, and it gave rise to the most commonly used alphabets to write Italo-Celtic (i.e. Celtic and Romance) and Germanic languages until modern times. Also, the Irish Ogham alphabet, dated from ca. 4th c. AD (but maybe as old as the 1st century BC) seems to have been born from the Latin alphabet, but possibly also from influences of the Elder Futhark, or some other variant of the Old Italic script.

¹ <http://scholar.lib.vt.edu/ejournals/ElAnt/V1N2/powell>

As a western archaic Greek script, the Euboean alphabet is “sister” of the eastern Ionic alphabet, which derived in the alphabet adopted by Athens in 403 BC under archon Eucleides, and later by most other parts of the Greek-speaking world during the 4th century BC. The Gothic and Cyrillic alphabets derived mainly from this classic Greek alphabet.

“Sister” systems were also the alphabets of Asia Minor, including regional variants like the Phrygian alphabet, and some divergent Anatolian scripts. Curiously, the Lydian script also developed the letter 8 to represent /f/, as did the Etruscan alphabet.

Archaic Greek scripts derived from the Phoenician alphabet, which in turn gave rise to the Aramaic script, and this after some iterations developed into the Arabic script ca. 400 AD (through the Syriac and then Nabatean script), and into the Nāgarī script ca. 1st-4th c. BC (through the Brahmi and then Gupta scripts).

The Old Italic script is therefore the ‘parent script’ of most western European scripts, whether ancient or modern, and seems the most natural choice for a common ancient European language. Just like the Cyrillic, Greek, Nāgarī, Arabic, Hebrew, or Chinese scripts all reveal at first sight the distinctive nature of the language they represent, the Old Italic script strikes visually as old European in nature. Because the most common letters can be adapted to the Late Proto-Indo-European phonetic system, the principle of letter-sound correspondence can be followed more precisely – albeit obviously incompletely – than with common modern alphabets.

Unlike other old European scripts, such as Linear B, Cuneiform scripts, or scripts of the Iberian Peninsula from Greek or Phoenician origin, the Old Italic script – like many scripts of Greek tradition – is easily deciphered by Europeans, and instantly recognizable as of western European heritage.

The modern use of the Old Italic script means using a late writing system, dated ca. 8th century BC, to write an older language, North-West Indo-European, probably spoken around the mid-3rd millennium. Such a use is not much different from the traditional use of the Latin alphabet to write ancient Latin texts; of the Classical Greek alphabet – and the more recent polytonic orthography – to write ancient Greek texts, like Homer’s epics; or of Nāgarī scripts to write ancient or classical Sanskrit texts.

2. Consonants

Phoneme	Italic	Greek	Latin	Cyrillic	Armenian	Perso-Arab	Nāgarī
[p]	Ɔ	Π π	P p	П п	Պ պ	پ-پ	प
[b]	Ɓ	Β β	B b	Б б	Բ բ	ب-ب	ब
[b ^h]	Φ	Βη βη	Bh bh	Бх бх	Բհ բհ	بهبهه	भ
[t]	Ƨ	Τ τ	T t	Т т	Տ տ	ت-ت	त
[t ^h]	↑	Θ θ	Th th	Тх тх	Թ թ	تتهه	थ
[d]	Ƨ	Δ δ	D d	Д д	Դ դ	د-د	द
[d ^h]	⊗	Δη δη	Dh dh	Дх дх	Դհ դհ	دهدهه	ध
[k]	Ƨ	Κ κ	K k	К к	Կ կ	ك-ك	क
[k ^h]	⊞	Χ χ	Kh kh	Кх кх	Կհ կհ	كهكهه	ख
[g]	Ƨ	Γ γ	G g	Г г	Գ գ	گ-گ	ग
[g ^h]	⊗	Γη γη	Gh gh	Гх гх	Գհ գհ	گهگهه	घ
[k ^w]	Ƨ	Ϟ ϟ	Q q	Қ қ	Խ խ	ق-ق	क्र
[g ^w]	Υ	Ϟ ϟ	C c	Гь гь	Ղ ղ	غ-غ	ग
[g ^{wh}]	δ	Ϟη ϟη	Ch ch	Гьх гьх	Ղհ ղհ	غهغهه	गह

[j]	ɹ	J j	J j	Й й	Յ յ	ي-ي	य
[w]	Ɔ	F Ɔ	W w	В в	Բ Բ	و-و	व
[r]	Ɔ	P ρ	R r	Р р	Ր Ր	ر-ر	र
[l]	Ɔ	Λ λ	L l	Л л	Լ լ	ل-ل	ल
[m]	Ɔ	M μ	M m	М м	Մ մ	م-م	म
[n]	Ɔ	N ν	N n	Н н	Ն ն	ن-ن	न
[s]	Ɔ	Σ σ ς	S s	С с	Ս ս	س-س	स

Other Indo-European phonemes (including laryngeals)

Phoneme	Italic	Greek	Latin	Cyrillic	Armenian	Perso-Arab	Nāgarī
[z]	Ɔ	Z ζ	Z z	З з	Ջ յ	ز-ز	ज़
[h]	Ɔ	H η	H h	Х х	Հ հ	ه-ه	ह
[x]	⊞	Χ χ	Kh kh	Кх кх	Կհ կհ	كهكهه	ख़
[ʔ]	‘	‘	‘	‘	‘	ء	॰
[ʕ ^w]	‘	‘	‘	‘	‘	ع-ع	

Other common phonemes

Phoneme	Italic	Greek	Latin	Cyrillic	Armenian	Perso-Arab	Nāgarī
[ɛ]	Ɔ	Μ μ	X x	Сч сч	(Վ վ)	بن-بن-بن	श
[v]	∇	Ϻ	V v	В в	Վ վ	ف-ف-ف	व

Proto-Greek

Phoneme	Italic	Greek	Latin	Cyrillic	Armenian	Perso-Arab	Nāgarī
[p ^h]	ƆΘ	Φ φ	Ph ph	Пх пх	Փ փ	په-په-په	फ
[k ^{wh}]	ƆΘ	Ϟη ϟη	Qh qh	Чх чх	Խհ խհ	ق-ق-ق	क्रह
[t̪s]	Τς	Τσ τσ	Ts ts	Ц ц	Ծ ծ	تس-تس-تس	तस
[d̪z]	ƆΙ	Δζ δζ	Dz dz	Дз дз	Զ ձ	دز-دز-دز	दष

Proto-Indo-Iranian

Phoneme	Italic	Greek	Latin	Cyrillic	Armenian	Perso-Arab	Nāgarī
[t̪ɛ]	ΤΜ	Τμ τμ	Κ κ	Ч ч	Զ չ	تژ-تژ-تژ	तश
[d̪z]	ƆΜ	Δμ δμ	Ġ ġ	Дщ дщ	Ֆ ց	دژ-دژ-دژ	दष
[d̪z ^h]	ƆΜΘ	Δμη δμη	Ġh ġh	Дщ дщ	Ֆհ ցհ	دژه-دژه-دژه	दषह
[t̪ʃ]	Τβ	Τβ τβ	Κ κ	Тш тш	Ը ճ	چ-چ-چ	च
[d̪ʒ]	Ɔβ	Δβ δβ	Ġ ġ	Дж дж	Ջ ղ	ج-ج-ج	ज
[d̪ʒ ^h]	ƆβΘ	Δβη δβη	Ġh ġh	Джх джх	Ջհ ղհ	جه-جه-جه	झ
[ʃ]	Ρ	Ρ ρ	Š š	Ш ш	Շ շ	ش-ش-ش	श

3. Vowels and vocalic allophones

Phoneme	Italic	Greek	Latin	Cyrillic	Armenian	Perso-Arab	Nāgarī
[a]	Ɑ	Α α	A a	А а	Ա ա	ا	अ
[e]	Ǝ	Ε ε	E e	Е е	Է է	ئ-ئ-ئ	ए
[o]	Ɔ	Ο ο	O o	О о	Օ օ	و-و-و	ओ
[a:]	Ɱ	Ἀ ᾱ	Ā ā	Ӓ ӓ	Ա̄ ա̄	اَ	आ
[e:]	Ǝ̄	Ἐ ἐ̄	Ē ē	Ӕ ӕ	Է̄ է̄	ئ̄-ئ̄-ئ̄	ऐ
[o:]	Ɔ̄	Ὀ ὄ	Ō ō	Ӗ ӗ	Օ̄ օ̄	و̄-و̄-و̄	औ

[i]	ɪ	Ι ι	I i	И и	Ի ի	ي-ي-ي	इ
[i:]	ɪ̄	Ἴ ῑ	Ī ī	Ӣ ӣ	Ի̄ ի̄	ي̄-ي̄-ي̄	ई
[u]	ʊ	Υ υ	U u	У у	Ո ո	و-و-و	उ
[u:]	ʊ̄	Ὼ ΐ	Ū ū	Ӧ ӧ	Ո̄ օ̄	و̄-و̄-و̄	ऊ

[r]	ɾ	Ρ ρ	R r	Р р	Ր ր	ر-ر-ر	ऋ
[l]	ɭ	Λ λ	L l	Л л	Լ լ	ل-ل-ل	ऌ
[m]	ɱ	Μ μ	M m	М м	Մ մ	م-م-م	अं
[n]	ɳ	Ν ν	N n	М м	Ն ն	ن-ن-ن	अः

4. Conventions

Accent

Indo-European is a tonal language with only one, acute accent per word, phonologically unpredictable, determined by morphological factors. It usually resides on the full grade vowel, more rarely on the zero-grade.

The accented syllable is higher in pitch than the surrounding syllables. The syllable after accent probably had falling pitch (cf. Skt. gen. sg. *vrkías* > *vrkyàs* (with falling tone), a contraction within Skt. history, the Rigveda scans with original syllabification.

Therefore, the natural way to represent the common acute accent is with the Greek acute accent (ὀξεῖα *sharp* or *high*) diacritic (´), and the less commonly represented falling pitch – especially interesting to facilitate performance of poetry – with the Greek grave (βαρεῖα *heavy* or *deep*) diacritic (`). These diacritics were introduced during the Hellenistic period in the 3rd century BC, but spread only from the 2nd century AD in papyri, to become standard in the Middle Ages.

The majuscule – writing entirely in capital letters – was the common formal writing style until the 8th century AD, when the minuscule polytonic began to replace it.

Just like Mandarin Chinese, PIE must have had both stress and pitch accent. Both were important, since some syllables must have had more prominence than others, and high pitch seems to have been more prominent (vowel length appears in most Anatolian words on PIE stressed syllables).

Stress is not generally marked in Modern Indo-European, although it could be represented with the syllable in bold. As a rule of thumb – as e.g. in Arabic, or in the Sezer stress pattern in Turkish –, syllable weight (the length of the syllable) marks the stress of words.

Whenever possible, then, syllables that include a long vowel or a diphthong (CVV) and those with more than one consonant (CVCC) are stressed. If in conflict, those with a combination of both (CVVCC) are probably the stressed ones².

² To get a general feeling of the pronunciation pattern of Modern Indo-European, with acute, grave, and stress accents, recordings of Stephen G. Daitz with restored Ancient Greek pronunciation are probably the best introduction. Even though Vedic Sanskrit must have had a similar pronunciation, there are currently no attempts as far as I know to record their original (i.e. restored) pronunciation, and most available recordings rely on regional traditions within India.

Since all non-clitic words of more than one syllable would be marked with one accent, as we have seen, a more elegant convention is not to write *all* accents *always*. The second to last syllable seems to be the most frequent accented syllable, so we can spare unnecessary diacritics if the accent is understood in that position, unless marked in another syllable. Therefore, acute accents are written over the vowel in accented syllables, except when it is on the second to last syllable (or *paenultima*), and in monosyllabic words, where it is not written.

Vowel length: the apex and the macron

Latin used the apex to mark long vowels.



Some have recently used the acute accent to represent the apex (for example in some Latin texts in the Wikipedia), in line with the modern Czech, Slovak, Irish, or Hungarian representation of vowel length.

However, the traditional representation of the apex is made with a different diacritic: a transverse, long, and curved mark. So for example the Vietnamese alphabet, which incorporated both acute and apex marks – with Alexandre de Rhodes describing them as different diacritics in his [Dictionarium Annamiticum Lusitanum et Latinum](#) (1651) –, where the apex marked final nasalization usually over letters ỗ, ữ.

The apex is a rare diacritic (as infrequently used as Old Italic letters), and the macron has been used for a long time already to represent vocalic length in different scripts, including Roman, Greek, and Cyrillic alphabet. It seems therefore more interesting to keep using the macron in Modern Indo-European³ instead of the apex, if only to adapt to typographic constraints.

³ In Old Italic vowels, the symbol Combining Overline (0305), $\overline{\text{v}}$, offers a better output over letters than a Combining Macron (0304), $\text{v}^{\text{~}}$.

The acute accent offers a poor result with Old Italic letters, but it depends on the font type selected (see below the section on Typography). It seems to have a slightly higher position, which seems more in line with the current typographic custom, when using the New Athena Unicode font, if the Combining Acute Tone Mark (0341) – $\text{v}^{\text{~}}$, $\text{v}^{\text{~}}$ – is selected instead of the Combining Acute Accent (0301) – $\text{v}^{\text{~}}$, $\text{v}^{\text{~}}$. However, many software programs would not support this differentiation.

Resonants

The diacritic dot below (\cdot) should be used to clearly mark the vocalic realization of sonorants. It is usual in Indo-European linguistics to use the diacritic ring below (\circ), but it seems more appropriate to leave it for phonetic representations in Modern Indo-European.

Regressive assimilation of voicing and deaspiration

Regressive assimilation of voicing / aspiration, a typologically normal (i.e. natural) phenomenon, is not represented in writing. A phonemically-correct notation is preferred over a phonetically-correct one:

TD > DD

Examples include **nisdo** for /nizdo/, *nest*, from **sed-**, *sit*; or Greek /epibda/, from **epi-** and zero-grade **-pd-**, from **ped-**, *foot*.

DT > TT (and DhT > TT)

-gt-, **-ght-**, **-kt-** are pronounced /kt/; **-bt-**, **-bht-**, **-pt-** are pronounced /pt/. For example, **juktós** for /juktós/, *joint*, from **jeug-**, *join*; or **leghtos** for /lektos/, *bed*, from **legh-**, *lie*.

Laryngeals

The merged laryngeal /h/ of the Disintegrating Indo-European was probably still alive in NWIE, at least in *CRHC groups. Written with an **h**, its sound was most likely /h/ (that of the original *h₁). Its pronunciation in IEDs could have been:

- The sonorant plus /h/, without an auxiliary vowel.
- A voiceless vowel *h̥
- An auxiliary schwa (commonly a NWIE vowel /a/) plus /h/.

Therefore, examples of *CRHC groups might be pronounced differently, but a common, phonemic writing is preferred, i.e. that which represents the old situation. For example, *g η h₃tós, ‘known’, may be written **g η htós** or **gn η tós**, for an alternating Northern LIE *g η htós / *gn η tós (*gn^ahtós), cf. Toch. *gnatós, Ita.-Cel. *gnātós, PreGmc. *g η tós, PreBSl. *g η htós.

Common examples include:

- *p η h₁nós, ‘full’: **p η hnós**; Ita.-Cel. *plānós, PreGmc. *p η lnós, PreBSl. *p η hnós.
- *g η h₁tós, ‘born’: **g η htós**; Ita.-Cel. *gnātós, PreGmc. *g η tós, PreBSl. *g η htós.

- **grh₂nóm*, ‘corn’: **grhnóm**; Ita.-Cel. **grānóm*, PreGmc. **grnóm*, PreBSl. **grhnóm*.
- **prh₂tós*, ‘sold’: **prhtós**; Ita.-Cel. **prātós*, PreGmc. **prtós*, PreBSl. **prhtós*.
- **střh₃tós*, ‘strewn’: **strhtós**; Ita.-Cel. **strātós*, PreGmc. **strtós*, PreBSl. **strhtós*.
- **g^wřh₂ús*, ‘heavy’: **crhwús**; Ita.-Cel. **g^wrāwus*, PreGmc. **g^wřwus*, PreBSl. **g^wřhwus*.
- **krh₂tis*, ‘wickerwork’: **krhtis**.
- **péprhtai** < **péprh₃th₂ei*, ‘you got production’.
- **péprhdhi** < **péprh₂d^hi*, ‘keep selling!’.

Only for metric reasons would the insertion of a glottal stop be possible in Modern Indo-European in place of an older laryngeal, essentially between vowels, as found in certain old Indo-Iranian compositions: as, **wé’entos**, instead of **wēntos**, wind.

See our paper on the loss of laryngeals and vocalism in North-West Indo-European (at <https://oldeuropean.org/>) for more on these cases.

Clitic words and compounds

Some words must accompany the name it modifies. For example, particles -qe, and, -we, or, accompany the previous words. Some pronouns have enclitic forms that are placed after the name they define.

In all these cases, it is the most traditional solution to place them either joint with the noun (*populusque*, *avasca*), or separated. For Modern Indo-European, it seems thus more traditional to join particles with the accompanied word, but it can be separated with a dash to more clearly differentiate the particle from the word.

Compound words are written, as in Latin, Greek, Sanskrit, or modern German and most Slavic and Romance languages, without a dash; thus, **sindhueuropajóm**, Indo-European.

Capital letters

All letters were capital letters in Latin and Classical Greek. Lowercase letters developed from cursive versions of uppercase letters, and its use in later Latin texts depended on the rules for medieval languages. Romance tradition dictates that only proper names, and not nouns (unlike German) or *local* adjectives (unlike English) need to be written in capitals, apart from the first word of the sentence. Therefore, **Rome**, but **roman**.

There was no ‘official’ rule for Latin or Classical Greek, though.

5. Old Italic alphabet

Names

Names and order of Old Italic letters imitate the traditional Latin *abecedarium*, with intercalation of aspirated and labialized phonemes for pragmatic purposes, in contrast with the Classical Greek *alphabet* order.

Phoneme	Italic	Name	Phoneme	Italic	Greek
[a]	ⱦ	ā	[k ^h]	ⱦ	khā
[b]	B	bē	[l]	L	ēl
[b ^h]	ⱦ	bhē	[m]	ⱦ	ēm
[g]	C	gē	[n]	ⱦ	ēn
[g ^h]	X	ghē	[o]	O	ō
[d]	D	dē	[p]	P	pē
[d ^h]	ⱦ	dhē	[k ^w]	Q	qū
[e]	E	ē	[r]	R	ēr
[w]	F	wā	[s]	S	ēs
[g ^w]	Y	cē	[t]	T	tē
[g ^{wh}]	8	chē	[t ^h]	↑	thē
[h]	H	hē	[u]	Y	ū
[i]	I	ī	[v]	V	vē
[j]	ⱦ	jōta	[x]	X	χā
[k]	K	kā	[z]	Z	zē(ta)

/g/, /g^h/

C is the common letter for /g/ in Old Italic scripts, and is also its most common value in archaic Greek and derived scripts, including Cyrillic and Gothic alphabets.

X represented /ks/ in the Latin script (from its original western Greek value) and /s/ in Etruscan. In Lugano it represented /g/, and the g-rune X in Elder Futhark evolved possibly from a similar northern alphabet, or maybe from an original /gs/ value. Because Proto-Germanic *g, after the effects of Grimm's law, comes (usually) from Indo-European *g^h, and the most common value in Greek scripts is of an aspirated sound (also found in Gothic and Cyrillic alphabets), it is proposed that X represents the aspirated voiced velar /g^h/ in Modern Indo-European.

/g^w/, /g^{wh}/

Υ, had a value of /k^h/ in western Greek, and was adopted as /k^h/, (or /kx/, /x/) for non-Indo-European languages of Italy (like Etruscan or Rhaetic), while in Indo-European scripts like Lepontic and Venetic variants⁴ it represented /g/. In Modern Indo-European, Υ is proposed to represent the common voiced labiovelar /g^w/.

8 was an Old Italic innovation representing /f/. It is found in Etruscan and Oscan, as well as in the Lydian script, probably as an independent parallel development. Its origin may have been an altered Β or Θ, or a new creation from Ϝ + Θ. Because of its likely original historical value as /w/+h/, and the similar derivation of the group *g^{wh} > *f in Latin and Venetic, it is probably the most appropriate letter to represent the aspirated labiovelar /g^{wh}/.

/k^h/, /x/, /h/, and laryngeals

Θ is the common Old Italic letter for /h/, related to Η in archaic Greek scripts, which represented almost uniformly /h/.

⌘, from the same letter as eastern Greek Ξ (which had a /ks/ value) was rarely used in Italic alphabets. For the proposed value of /k^h/ in Modern Indo-European, it can be interpreted as a ligature of ϙ + Θ.

⌘ as /k^h/ can also be used – as it was used in certain ancient Indo-European languages – with an alternative value /x/; cf. onomatopoetic laughing words reconstructed alternatively as *ha ha or *kha kha. A peculiar development in south Picene and in Camunic, ✱, probably representing /x/ in the former, could be used to represent more exactly this sound, but it has currently little support in common typefaces (it is only available in the Aegean typeface).

It is proposed that the common phonetic representation of the glottal stop /ʔ/ in European languages, the apostrophe ', also represents this sound in Modern Indo-European.

⁴ The similar z-rune Υ in Elder Futhark seems to be originally a variation of ʀ in certain northern alphabets.

/d^h/ and /t^h/

⊗ was adopted from *theta* in ancient Greek script, to represent /t^h/ in Etruscan and /t/ in Venetic and Lepontic. It is proposed to represent the common IE dental aspirated /d^h/.

For the less common aspirated dental sound, /t^h/, the northern group letter ↑ seems more appropriate. It represented Rhaetic /t'/ and Camunic /t^h/, and is found later in Germanic runes with a /t^h/ value.

/b^h/

Φ, as found in Etruscan, represented originally the voiceless bilabial aspirated /p^h/, equivalent to Greek Φ, but it was adopted as voiceless /b/ in Venetic. In Modern Indo-European, it is proposed to represent the common bilabial aspirated /b^h/.

/j/

Ancient languages of Italy used diverse methods represent the sound /j/. It was usually written with l, which could thus represent both /i/ and /j/. Lepontic and Venetic sometimes used ll or l.l to represent /j/ or /ij/. The alphabet of Lagole (for Venetic) shows a peculiar innovative use of Ɔ (originally /g/) to represent /j/, and this use is also later found in Elder Futhark ȝ, originally (probably) a Ɔ, runic ȝ, with an additional stroke. This tradition (which might have happened independently from the Venetic innovation) continued into Gothic *jer*, Ɔ j, and into Anglo-Saxon *Gēr* Ɔ (and more rarely *Ior* *).

In the Oscan alphabet long vowels ū and ī are represented with V and ʒ, respectively. An innovative *heta*, F, a similar symbol developed in Greek colonies of southern Italy to represent /h/ (when H was eventually used to represent a vowel), was later incorporated as the rough breathing diacritic in Greek.

The use of ʒ seems appropriate to distinguish /j/ from /i/ in Modern Indo-European, given its tradition in Greek-related scripts, and its similarity with Latin script j.

Numerals

Numerals can be written with Arabic numerals, or with Old Italic numerals. Old Italic numerals – from which Roman numerals derived – are hypothesized to have derived from notches on tally sticks, as is common in European cultures that use tally marks.

Old Italic	I	II	III	IIII	Λ	ΛI	ΛII	ΛIII	ΛIIII	×	↑	8	∇	⊗
Roman	I	II	III	IV	V	VI	VII	VIII	IX	X	L	C	D	M
Arabic	1	2	3	4	5	6	7	8	9	10	50	100	500	1000

Typography

New Athena Unicode

A sans serif typeface, it has a bold, italic, and bold-italic fonts, and downloadable files for its use in websites. Probably the most frequently updated from the list, it offers support for multiple ancient languages, so that you can use it seamlessly in texts that include different scripts. [Download](#).

ΕΥΡΩΠΑΪΟΝ: ΣΤΕΦΕ ΜΑΡΙΑ ΥΡΑΤ-Α ΝΕΜΑ ΤΥ

Cardo

It looks quite well with macrons, but accents are not rendered correctly. It would be a good choice otherwise. [Download](#).

ΕΥΡΩΠΑΪΟΝ: ΣΤΕΦΕ ΜΑΡΙΑ ΥΡΑΤ-Α ΝΕΜΑ ΤΥ

Aegean

It has a bold version. Download the [fonts Aegean](#).

ΕΥΡΩΠΑΪΟΝ: ΣΤΕΦΕ ΜΑΡΙΑ ΥΡΑΤ-Α ΝΕΜΑ ΤΥ

Alphabetum

Commercial, right to left typeface. You can try the free Alpha-test font [downloading it here](#):

ΕΥΡΩΠΑΪΟΝ: ΣΤΕΦΕ ΜΑΡΙΑ ΥΡΑΤ-Α ΝΕΜΑ ΤΥ

Segoe UI

Included by default in Windows. **Segoe UI Symbol** worked also well with many scripts, but Microsoft has discontinued its support in Windows 10, and it is not included anymore. You can still [download](#) and install it (version 5.90 for Windows 8 or 8.1) instead of the font by default in Windows 10 or in Windows 7 and previous versions.

ΕΥΡΩΠΑΪΟΝ: ΣΤΕΦΕ ΜΑΡΙΑ ΥΡΑΤ-Α ΝΕΜΑ ΤΥ

Keyboard utilities

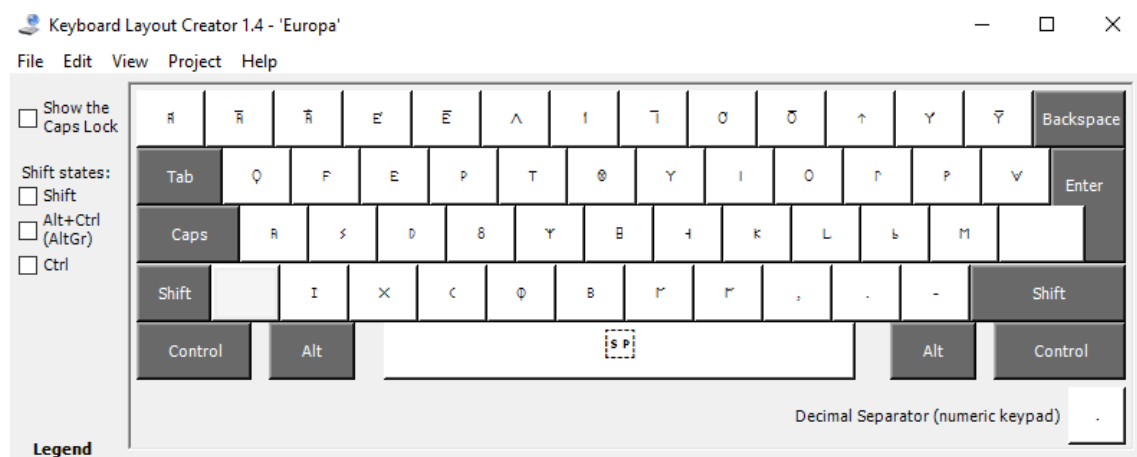
Our efforts to create a utility to write with Old Italic characters have been unsuccessful.

Our experience with Keyman Developer was disappointing at best, due to the need of additional user software to use it.

The free software Microsoft Keyboard Layout Creator (KLC)⁵, still in version 1.4, is an old tool, whose output we were not able to use with Windows 10. We have not tried it in other versions.

However, you can download our prepared **Old Italic keyboard layout source file** from the website of Academia Prisca, to use it strictly within Microsoft KLC.

Once the software is open, select *File > Load Source File*, and open our source file, called INEeu.klc.



Microsoft Keyboard Layout for Modern Indo-European in Old Italic characters

Then select *Project > Test Keyboard Layout...*, and you can write the text in the test windows. With your text written in Old Italic characters, you can copy & paste it anywhere.

You can also use the website Europajom.org with its MediaWiki WYSIWYG editor including many Unicode characters to write texts, or correct what you have written using Microsoft KLC.

Any help in this field is indeed welcome.

⁵ Now downloadable at <<https://www.microsoft.com/en-us/download/details.aspx?id=22339>>

6. Latin alphabet

The Latin alphabet used to write Indo-European is similar to the English, which is in turn borrowed from the Late Latin *abecedarium*. Because of the role of this alphabet as model for other ones, simplicity and availability of the characters is preferred over tradition and exactitude.

The Latin alphabet was borrowed in very early times either directly from the Cumaeen alphabet or from the Etruscan one, and did not at first contain the letter G. The letters Y and Z were introduced still later, about 50 BC. The Latin character C originally meant /g/, a value always retained in the abbreviations *C.* (for *Gaius*) and *Cn.* (for *Gnaeus*). That was probably due to Etruscan influence, which copied it from Greek Γ, *gamma*, just as later Cyrillic Γ, *Ge*. In early Latin script C came also to be used for /k/, and K disappeared except before in a few words, as *Kal.* (*Kalendae*), *Karthago*. Thus there was no distinction in writing between the sounds /g/ and /k/. This defect was later remedied by forming (from C, the original g-letter) a new character G. In Modern Indo-European, unambiguous **K** stands for /k/, and **G** for /g/, so **C** is left without value, being used (taking its oldest value /g/) to represent the labiovelar /g^w/.

V originally denoted the vowel sound /u/ (Eng. ‘oo’), and F stood for the sound of consonant /w/ (from Gk. Ϝ, called *digamma*). When F acquired the value of our /f/, V came to be used for consonant /w/ as well as for the vowel /u/. The Latin semivowel /w/ developed into Romance /v/; therefore V no longer adequately represented /u/ or /w/, and the Latin alphabet had to develop alternative letters. The Germanic /w/ phoneme was therefore written as VV (a doubled V or U) by the seventh or eighth century by the earliest writers of Old English and Old High German. During the late Middle Ages, two forms of V developed, which were both used for its ancestor /u/ and modern /v/. The pointed form V was written at the beginning of a word, while a rounded form U was used in the middle or end, regardless of sound. The more recent letters **U** and Germanic **W** probably represent the sounds /u/ and /w/ respectively more unambiguously than Latin V.

The letter **I** stood for the vowel /i/, and was also used in Latin (as in Modern Greek) for its consonant sound /j/. **J** was originally developed as a swash character to end some Roman numerals in place of I; both I and J represented /i/, /i:/, and /j/. In MIE, **J** represents the semivowel /j/, an old Latin value current in most Germanic and Slavic languages. **Y** is used to represent the vowel /y/ in foreign words. That /j/ value is retained in English J

only in foreign words, as *Hallelujah* or *Jehovah*. Because Romance languages developed new sounds (from former /j/ and /g/) that came to be represented as I and J, English J (from French J), as well as Spanish, Portuguese or Italian J have sound values quite different from /j/. The romanisation of the sound /j/ from different writing systems (like Devanagari) as Y – which originally represented in Latin script the Greek vowel /y/ – is due to its modern value in English and French, and has spread a common representation of /j/ as Y in Indo-European studies, while J is used to represent other sounds.

A different use of the Latin alphabet to represent PIE, following the Classical Latin tradition, is available at <<http://verger1.narod.ru/lang1.htm>>.

Typography

Telefon

Telefon <<https://monokrom.no/fonts/telefon>> is a beautiful font with specific support for Proto-Indo-European and Modern Indo-European characters, reworked by [Frode Bo Helland](#) in collaboration with Fernando López-Menchero.

ūpādhyāy kūse
 ūpādhyāy kūse

Arial Unicode

Times New Roman

MS Reference, MS Reference Serif

7. Cyrillic alphabet

The Cyrillic script is used following its modern trends, taking on account that Russian is the model for most modern keyboards and available typography.

Non-Russian characters have been avoided, and we have followed the principle of one letter for each sound: While Ъ is commonly used to represent /j/, Cyrillic scripts usually lack a character to represent consonantal /w/, given that usually /v/ (written В) replaces it. While У is generally used in Cyrillic for foreign words, a ‘one character, one sound’ policy requires the use of a character complementary to Ъ, which is logically found in В – a sound lacking in Indo-European.

In Slavistic transcription *jer* Ъ and front *jer* В were used to denote Proto-Slavic extra-short sounds /ǔ/ and /ǐ/ respectively (e.g. *slověnskъ* adj. ‘slavonic’). Today they are used with other values in the different languages that still use them, but the need for traditional ‘labial’ /^w/ and ‘palatal’ /^j/ signs available in most Cyrillic keyboards made them the most logical selection to mark a change of value in the characters representing stops.

8. Greek alphabet

The Modern Greek alphabet lacks letters to represent PIE phonetics properly. Therefore, the Ancient Greek letters and values assigned to them are used instead.

The consonant cluster /k^h/ was in Ancient Greece written as X (Chi) in eastern Greek, and Ξ (Xi) in western Greek dialects. In the end, **X** was standardised as /k^h/ (/x/ in modern Greek), while Ξ represented /ks/. In the Greek alphabet used for IE, **X** represents /k^h/, while Ξ represents /k^{wh}/, necessary for the representation of a Proto-Greek voiceless aspirate. As in Ancient Greek, Φ stands for /p^h/, and Θ for /t^h/.

The Greek alphabet lacks a proper representation for long vowels, so they are all marked (as in the other alphabets) with diacritics. **H** is used to represent the sound /h/, as it was originally used in most Ancient Greek dialects; it is also used to mark (voiced) aspirated phonemes. Ē represents /e:/ and Ō stands for /o:/ in the Greek alphabet for IE. For more on the problem of historical Eta and its representation in the Modern Greek alphabet, see <http://www.opoudjis.net/unicode/unicode_aitch.html>.

While not a practical solution (in relation to the available Modern Greek keyboards), we keep a traditional Ancient Greek script, assuming that it will enjoy the transliteration of texts mainly written in Latin or Cyrillic letters; so e.g. Archaic koppa **Q** stood for /k/ before back vowels (e.g. *Qóρνθοϝ*, *Korinthos*), hence its IE value /k^w/. Archaic digamma **F** represented /w/, a sound lost already in Classical Greek. Additions to the IE alphabet are new letter koppa **Ϟ** for /g^w/, based on the alternative Unicode shapes of the archaic koppa, and the ‘more traditional’ inverted iota **Ϡ** for /j/, preferred over Latin yot – although the lack of capital letter for inverted iota makes the use of (at least) a capital **J** necessary to distinguish /j/ from /i/. See <<http://www.opoudjis.net/unicode/yot.html>>.

9. Perso-Arabic script

The Perso-Arabic script has been adapted to the needs of a fully differentiated PIE alphabet, following Persian, Urdu and Kurdish examples.

The Perso-Arabic script is a writing system that is originally based on the Arabic alphabet. Originally used exclusively for the Arabic language, the Arabic script was modified to match the Persian language, adding four letters: پ [p], ت [tʃ], ژ [ʒ], and گ [g]. Many languages which use the Perso-Arabic script add other letters. Besides the Persian alphabet itself, the Perso-Arabic script has been applied to the Urdu or Kurdish Soranî alphabet.

Unlike the standard Arabic alphabet, which is an abjad (each symbol represents a consonant, the vowels being more or less defective), the MIE perso-arabic script is a true alphabet, in which vowels are mandatory, making the script easy to read.

Among the most difficult decisions is the use of letters to represent vowels – as in modern alphabets like Kurdish or Berber – instead of diacritics – as in the traditional Arabic or Urdu scripts. Following tradition, hamza (originally a glottal stop) should probably be placed on the short vowels and resonants, instead of the long ones (especially above ‘alif), but automatic equivalence with the other alphabets make the opposite selection more practical.

Because waw و and yodh ي could represent short and long vowels /u/ and /i/, and consonantal /w/ and /j/, a conventional selection of current variants has been made: Arabic letter Ve, sometimes used to represent the sound /v/ when transliterating foreign words in Arabic, and also used in writing languages with that sound (like Kurdish) is an obvious selection for consonantal /w/ because of its availability. The three-dotted yodh becomes then a consequent selection for consonantal yodh. Hamza distinguishes then the long vowel from the short ones, which is represented with the original symbols.

10. Brahmic scripts

The Brahmic or Indic scripts are a family of abugida (alphabetic-syllabary) writing systems, historically used within their communities – from Pakistan to Indochina – to represent Sanskrit, whose phonology is similar to the parent PIE language. Devanāgarī has come to be the most commonly used Brahmic script to represent Sanskrit, hence our proposal of its character values for the rest of them.

The characters and accents are generally used following their traditional phonetic value. Exceptions are the lack of vocalic characters to properly represent /ṁ/ and /ṇ/. Hence anusvara ँ, which represents /ṁ/, is used to represent /ṁ/. Also, visarga ॐ, which stands for /ḥ/ (allophonic with word-final /r/ and /s/) is proposed for /ṇ/.

Automatic transliteration between many Brahmic scripts is usually possible, and highly available within scripts used in India.

That happens e.g. with the InScript keyboard: because all Brahmic scripts share the same order, any person who knows InScript typing in one script can type in any other Indic script using dictation even without knowledge of that script.

11. Armenian alphabet

Armenian characters, similarly to Greek, need to be adapted to a language with a different series of short and long vowels and aspirated phonemes.

Because of that, a tentative selection is made, which needs not be final – as with any other script. Because Armenian lacks a proper character for /u/, and because it has not different characters to represent long vowels other than /e:/ or /o:/, the more practical choice is to imitate the other alphabets to allow for equivalence. The characters that represent short vowels also represent different sounds; as, Է for /e/ and word initially /je/, and Ո for /o/ and word initially /vo/, so a less ambiguous choice would be Է̄ for /e/ and Օ for /o/. Hence the letter Ո historically used to write /o/ and /u/ (in digraphs) stands for /u/.

The conventional selection of one-character representation of aspirated voiceless consonants follows Armenian tradition and equivalence with Greek, a closely related language, as we have already seen; i.e. Proto-Greek is probably the nearest branch to the one Pre-Armenian actually belonged to, and it is therefore practical to retain equivalence between both scripts.

Armenian diacritics (like the abbreviation mark proposed for long vowels) are defined as ‘modifier letters’, not as ‘combining diacritical marks’ in Unicode, so they do not combine as true superscript. Some fonts do combine them, as Everson Mono Ս̄ զ̄ զ̄̄ Է̄ Է̄̄ Օ̄ Օ̄̄ Ի̄ Ի̄̄ Ո̄ Ո̄̄.