L2/12-386

Doc Type: Working Group Document
Title: Revised Proposal to Encode Additional Old Italic Characters
Source: UC Berkeley Script Encoding Initiative (Universal Scripts Project)
Author: Christopher C. Little (chrislit@berkeley.edu)
Status: Liaison Contribution
Action: For consideration by JTC1/SC2/WG2 and UTC
Replaces: N4046 (L2/11-146R)
Date: 2012-11-06

0. Introduction

The existing Old Italic character repertoire includes 31 letters and 4 numerals. The Unicode Standard, following the recommendations in the proposal L2/00-140, states that Old Italic is to be used for the encoding of Etruscan, Faliscan, Oscan, Umbrian, North Picene, and South Picene. It also specifically states that Old Italic characters are inappropriate for encoding the languages of ancient Italy north of Etruria (Venetic, Raetic, Lepontic, and Gallic). It is true that the inscriptions of languages north of Etruria exhibit a number of common features, but those features are often exhibited by the other scripts of Italy. Only one of these northern languages, Raetic, requires the addition of any additional characters in order to be fully supported by the Old Italic block. Accordingly, following the addition of this one character, the Unicode Standard should be amended to recommend the encoding of Venetic, Raetic, Lepontic, and Gallic using Old Italic characters. In addition, one additional character is necessary to encode South Picene inscriptions.

This proposal is divided into five parts: The first part (§1) identifies the two unencoded characters (Raetic 𐌑 and South Picene 𐌟) and demonstrates their use in inscriptions. The second part (§2) examines the use of each Old Italic character, as it appears in Etruscan, Faliscan, Oscan, Umbrian, South Picene, Venetic, Raetic, Lepontic, Gallic, and archaic Latin, to demonstrate the unifiability of the northern Italic languages' scripts with Old Italic. The third part (§3) considers arguments against the unification of Venetic, Raetic, Lepontic, and Gallic scripts with Unicode's existing Old Italic script. The fourth part (§4) demonstrates the viability of this unification via sample encodings of inscriptions from many of the northern Italic languages. And the fifth part (§5) suggests revised text for section 14.2 of the standard.

This proposal is intentionally silent on the Camunic writing system, which may or may not be unifiable with the Unicode Old Italic script. Encoders should not be specifically dissuaded from employing the Old Italic script for Camunic, as they are currently discouraged from using it to encode Venetic, Raetic, Gallic, and Lepontic. However, the user community does not currently exhibit consensus on the Camunic character repertoire and it may require additional characters in Old Italic if it can be unified therewith.

0.1. Changes

The previous revision of this document (N4046 =L2/11-146R) referenced and claimed to cover inscriptions written in Ligurian. There are no known extant inscriptions in this language, and identification of Ligurian glyph forms has been corrected to refer to the Raetic glyphs that they reflect. This change has no effect on the proposed character repertoire.

Standard transliterations have been added to all example transcriptions.

Sections 1.5 (Proposed code chart annotations), 3 (Against the disunification of northern and southern Old Italic), and 5 (Recommendation for revised text of section 14.2 of the standard) are entirely new to this revision.

The spelling *Rhetic* has been changed to *Raetic* throughout. Although *Rhetic*, *Raetic*, and *Rhaetic* are all accepted English spellings, *Raetic* appears to be most common.

1. New Characters and Annotations

1.1. Justification

1.1.1. Raetic 𐌸

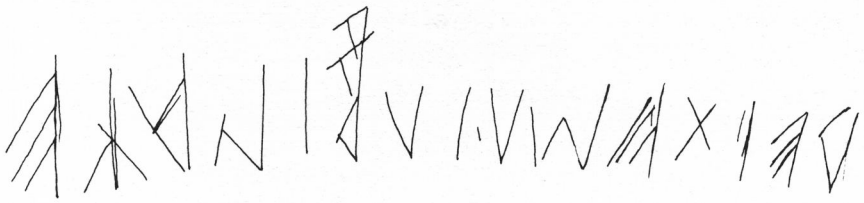
𐌸	U+1032F	OLD ITALIC LETTER TTE

Raetic exhibits a triangle symbol in inscriptions from Magrè. The shape variably appears as 𐌸 or 𐌹, but most frequently as 𐌸. The symbol is interpreted to be a dental phoneme, transliterated as *t'* by Bonfante (1996) and as *th* or *þ* by Jensen (1969). Diringer (1968) acknowledges the existence of the letter, but offers no transcription. And Schumacher (1992), writing on the inscriptions of Raetia, presents the inscriptions in transliteration but gives no transliteration of the 𐌸 glyph, rendering it instead with a drawing of the sign itself.

Two inscriptions in which it appears are MA-8/PID 227 and MA-10/PID-229, illustrated below:

73. Corno cervino. Conservato nel Museo di Este. Lungo cm. 10,5. Presenta modanature alle due estremità. L'iscrizione ha ductus sinistrorso; alcune lettere sono capovolte. Si ha un punto all'interno di *uiu* (Tav. XLIII, 3).

G. PELLEGRINI, in *NSc,* 1918, p. 183; *PID* 227.



A. *reitem uiu t'inaχe*
B. *VII*(?)

Fig. 1-1: (Morandi 1982:199)

MAGRE' 8 (PID 227) MNA I.G. 58808 MA-8

Hirschhornsprosse, an den beiden Enden gerillt; am schmäleren Ende ein Loch. L. 10,5 cm.

V: ⟵ reit^{z}emu.iu$\{$inaχe

R: Nichtschriftliche Zeichen

Fig. 1-2: (Schumacher 1992:163)

An example transcription of MA-8/PID 227, supplemented with a PUA , is: PEITEMY.IY IMAYE <reitemu.iut'inaχe>

G. PELLEGRINI, in *NSc*, 1918, p. 187; *PID* 229; A. MANCINI, in *REI* di *StEtr*, XLIII, 1975, pp. 254-255.

A. *rit'ie kerrinake*
B. *z*(?)

Fig. 1-3: (Morandi 1982:200)

MAGRE' 10 (PID 229, IR 9) MNA I.G. 58809 MA-10

Hirschhornsprosse, am schmäleren Ende ein Loch. L. 10 cm.

V: ⟶ $\frac{r^{z}}{p}$ i?ieker$\frac{r^{z}}{p}$inake

R: Nichtschriftliche Zeichen

Fig. 1-4: (Schumacher 1992:163)

An example transcription of MA-10/PID 229, supplemented with a PUA , is: 𐌓𐌉𐌉𐌄𐌊𐌄𐌓𐌓𐌉𐌍𐌀𐌊𐌄 <rit'iekerrinake>.

1.1.2. South Picene 𐌟

𐌟	U+1031F	OLD ITALIC LETTER ESS

In one South Picene inscription, TE-5, a stele from Penna Sant'Andrea, an unencoded character appears twice. It is believed to be derived from the letter *ka* (𐌊) mirrored across its *y*-axis. The phonemic value is believed to be some variety of sibilant, transliterated variously as *ś* (Marinetti 1985) and *σ* (Rix 2002). Relative to its first instance, the sign itself appears rotated 90° in its second instance, but the orientation of the first instance is typically cited in sign lists as the exemplar.

Fig. 17. – Museo Nazionale di Chieti. Stele da Penna S. Andrea (TE. 5). Ricostruzione della stele dai tre frammenti combacianti.

Fig. 1-5: TE-5 assembled (Marinetti 1985:Fig. 17)

Fig. 20. – Stele da Penna S. Andrea (TE. 5). Frammento inferiore.

Fig. 1-6: TE-5 lower section (Marinetti 1985:Fig. 20)

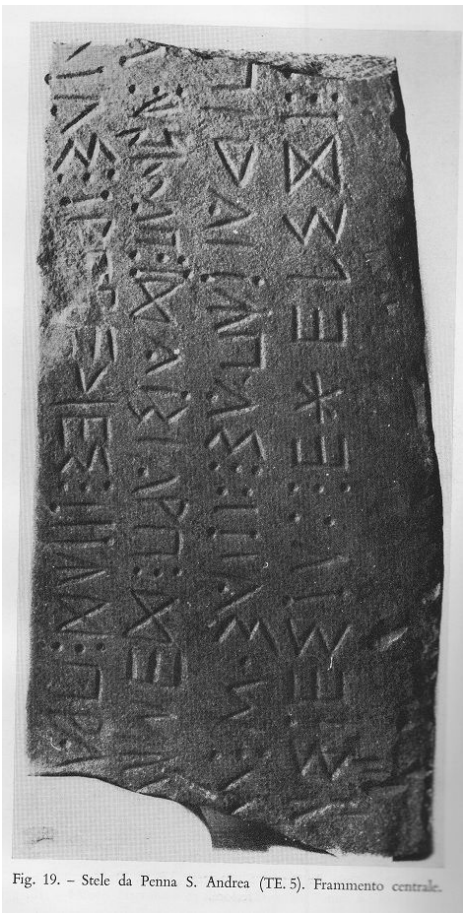


Fig. 19. – Stele da Penna S. Andrea (TE. 5). Frammento centrale.

Fig. 1-7: TE-5 middle section (Marinetti 1985:Fig. 19)

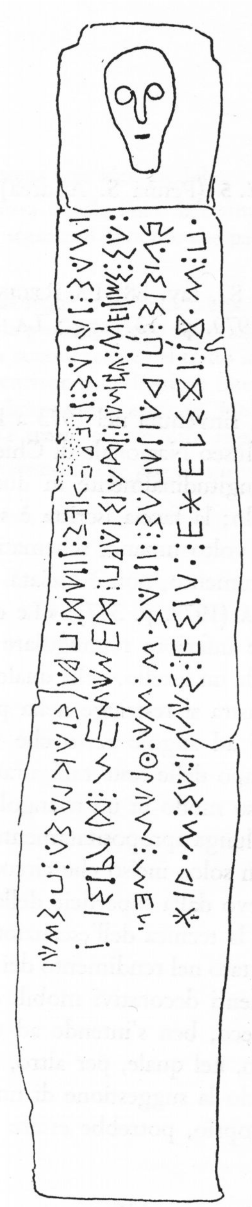


Fig. 1-8: (Marinetti 1985:216)

śidom : safinús : estuf : eśelsít : tíom :po/
vaisis : pidaitúpas : fitiasom : múfqlúm : me
⟦n/t⟧fistrúí : nemúneí : praistaít : panivú :
meitims : saf/inas : tútas : tre$\frac{b}{r}$egies : titúí :
praistaklasa : posmúi

Fig. 1-9: (Marinetti 1985:217)

Sp TE 5 stela *Penna S. Andrea*	**σidom: safinús: estuf: eσelsít: tíom: povaisis:** **pidaitúpas: fitiasom: múfqlúm: me{n}fistrúí:** **nemúneí: praistaít: panivú: meitims: safinas:** **tútas: trebegies: titúí: praistaklasa: posmúi**	IS TE 5*; Marin; 6:s

Fig. 1-10: (Rix 2002:68)

An example transcription of the first line of TE-5, supplemented with a PUA ✱, is:
✱𐌉𐌃𐌏𐌌:𐌔𐌀𐌚𐌉𐌍𐌞𐌔:𐌄𐌔𐌕𐌖𐌚:𐌄✱𐌄𐌋𐌔𐌝𐌕:𐌕𐌝𐌏𐌌:𐌐𐌏- <σidom:safinús:estuf:eσelsít:tíom:po->.

1.2. Allocation

The range U+10300-U+1032F is allocated to Old Italic, with positions U+10300-U+1031E assigned to letters and U+10320-U+10323 assigned to numerals.

My recommendation is to assign South Picene 𐌟 to U+1031F and Raetic 𐌿 to U+1032F. Thus, additional numerals may be assigned to the codepoints following U+10323, if necessary.

1.3. Character properties

```
1031F;OLD ITALIC LETTER ESS;Lo;0;L;;;;;N;;;;;
1032F;OLD ITALIC LETTER TTE;Lo;0;L;;;;;N;;;;;
```

1.4. Confusables

```
1031F OLD ITALIC LETTER ESS     ;       2731 HEAVY ASTERISK
```

1.5. Proposed code chart annotations

The following NamesList.txt annotations are proposed for the proposed characters:

```
1031F   OLD ITALIC LETTER ESS
        * South Picene
1032F   OLD ITALIC LETTER TTE
        * Raetic
```

The following additional annotations are proposed, based on a suggestion by Rex Wallace. These annotations acknowledge other names by which Old Italic ESH and SHE are called (samekh/ksi and tsadi/san respectively):

```
1030E   OLD ITALIC LETTER ESH
        = samekh
        = ksi

10311   OLD ITALIC LETTER SHE
        = tsadi
        = san
```

2. Survey of Old Italic script use across Italy

For the purpose of demonstrating the unifiability of the scripts of northern Italy with the scripts already unified in the Old Italic block, the glyph repertoires from each of the non-Greek, geographically-Italic writing systems presented in Bonfante (1996), Conway (1897), Diringer (1968), Faulmann (1880), and Jensen (1969) are collected and compared below. The writing systems under consideration include: Etruscan (Etr), Oscan (Osc), Umbrian (Umb), South Picene (SP), Faliscan (Fal), Archaic Latin (Lat), Venetic (Ven), Raetic (Rae), Gallic (Gal), and Lepontic (Lep).

In considering and enumerating the various glyphs of these languages, mirroring and minor variations in orientation will not be noted—all glyphs will be rendered in their left-to-right orientation, as Unicode does and as is common of modern scholarship. Differences in rounded versus angled letter

forms will not be taken as graphemic differences. The Old Italic glyphs that appear below are taken from David Perry's Cardo font with modifications and additions where it was lacking in glyph variants.

U+10300 𐌀 A ā <a>

The first letter of the alphabet is one of the most graphically diverse. Etruscan and southern Italic languages typically use easily recognizable forms such as 𐌀, 𐌁, and 𐌂. Latin uses less common forms such as 𐌛, 𐌜, 𐌝, and 𐌞. Faliscan uses the most dissimilar form of all: 𐌁.

Within northern Italic, 𐌞 (Ven, Rae, Lep), 𐌅 (Rae, Lep, Gal), and 𐌁 (Ven, Rae, Lep) are the most common forms, though 𐌄 (Ven), 𐌛 (Rae), and 𐌜 (Rae, Lep) also appear. The widespread northern Italic use of 𐌞 and 𐌅 (itself not elsewhere attested, though clearly related to the former) suggests the possibility that northern Italic constitutes a script distinct from Old Italic, but all forms retain the same general phonetic value and are clearly derived from a common model.

U+10301 𐌁 BE bē <b>

Throughout Italy, the form 𐌁/𐌂 was used, though many of the languages lacked a /b/ phoneme and thus lost the grapheme from their alphabets entirely. This letter is lacking in the northern Old Italic languages.

U+10302 𐌂 KE kē <c/g>

The most common form of this letter was simply 𐌂/𐌓. Venetic and Raetic attest this form. Etruscan attests a gimel-like form: 𐌌.

U+10303 𐌃 DE dē <d>

For dē, the most common form is again the form most recognizable in modern Latin: 𐌃/𐌄. R-like forms also appear in Oscan: 𐌐/𐌁. And Umbrian attests the novel form: 𐌕. Since northern Italic languages borrowed their alphabets from Etruscan after it had purged letters for phonemes it lacked, this letter is absent in the north.

U+10304 𐌄 E ē <e>

The form 𐌄 is most widespread throughout Italy, though 𐌄 also appears in Etruscan and the southern Italic languages. Latin and Faliscan, in addition to both of these forms, also attest a 𐌇 glyph. In northern Italy, 𐌄 appears for all languages and Raetic attests a unique 5-stroke form: 𐌄.

U+10305 𐌅 VE vē <f>

The letter vē is widely varied in Italy. The Unicode exemplar form, 𐌅, is typical of Etruscan, but otherwise attested only in Latin and South Picene in southern Italy. Oscan, Umbrian, and Etruscan demonstrate slightly varied forms such as 𐌆 and 𐌉. Latin presents a unique 𐌉 form, akin to its unique shape for ē. And Faliscan possesses a unique 𐌚 form.

In northern Italy, Venetic, Lepontic, and Raetic all use a shape identical to the Unicode exemplar form, 𐌅, suggesting that their unification with the Etruscan model alphabet is better warranted than the southern alphabets, at least on the basis of this letter.

U+10306 𐌆 ZE zē <z>

This letter is also widely varied in shape. The shape 𐌆/𐌉 is common in Etruscan, Oscan, and Faliscan. Other forms include 𐌊 (Etr, Fal, Umb), 𐌕 (Umb), 𐌄 (Fal, Umb), 𐌉 (Osc), 𐌆 (Lat), and 𐌅 (Etr, Fal). In northern Italy, the forms are no less varied. In common with southern Italy, 𐌊 (Ven, Rae, Lep) and 𐌄 (Ven) appear. Unique to the area are variants on the 𐌊 glyph: 𐌋 (Rae) and 𐌌 (Ven). Since these are clear derivatives with the same alphabetic position and similar phonetic values, they can easily be unified with the model form.

U+10307 𐌇 HE hē <h>

The letter hē appears in two major variants, 𐌇 (Etr, Osc, Fal, Lat) and 𐌈 (Etr). Circular versions of the former are common to Umbrian: 𐌏/𐌐. Other rectangular variants of the same form are rarely attested, usually unique to a single writing system: 𐌇 (Etr; probably only on the Marsiliana abecedarium), 𐌃 (Fal), 𐌋 (SP), 𐌇 (Fal, Lat), 𐌅 (Etr, SP), and 𐌆 (Etr). The Etruscan form 𐌈 is also common in Venetic and Raetic. Venetic also possesses the novel forms 𐌡, 𐌢, and 𐌣.

U+10308 𐌈 THE thē <θ>

The descendants of Greek θ appear in round, squared, and un-circumscribed varieties. Unicode's exemplar form, 𐌈, is common only in Etruscan. A square variety, 𐌆, is seen in South Picene. Circumscribed dots are seen more widely: 𐌏 (Etr, Umb, Fal); 𐌐 (Etr). Varieties with surrounded bars and crosses appear, chiefly in Etruscan: 𐌆, 𐌐, 𐌏. A few empty varieties also appear: 𐌏 (Etr, Fal), 𐌏 (SP), and 𐌏 (Etr). Oscan uses its glyphs for hē (𐌇) and tē (𐌕) to represent thē.

In Venetic and Lepontic, the common 𐌏 glyph is used. The most common glyphs used in Venetic are 𐌆 and its un-circumscribed form 𐌗. Its similarity to the letters eks and tē and their predecessors have led to suggestions that it is a unique letter that should be separately encoded, but it is, in fact, simply a derivative of western Greek θ, easily unified with the existing Old Italic character.

U+10309 𐌉 I i <i>

The basic 𐌉 shape is used in all Italic languages. The additional forms 𐌉 (Etr) and 𐌊 (Rae) are rare.

U+1030A 𐌊 KA kā <k>

The exemplar form 𐌊, sometimes with minor shape variations, is used in all Italic writing systems that have not dropped the letter (perhaps in favor of kē, as in Etruscan).

U+1030B 𐌋 EL el <l>

The exemplar form 𐌋, is used in all Italic languages. A Greek λ-like form (𐌌) is attested in Faliscan. A Λ-like form (𐌛) is seen in Lepontic. And a modern-type 𐌋 form is seen in Faliscan, Etruscan, and Lepontic. Raetic and Venetic also attest an inverted 𐌟 form.

U+1030C 𐌌 EM em <m>

The letter em, though widely varied throughout Italy, displays little unique variation in northern Italy. Common shapes include 𐌌 (Etr, Fal, Lat, Ven, Rae, Lep), 𐌑 (Etr, Osc, Umb, Fal), 𐌑 (Etr, Osc,

Umb, Fal), and 𐌌 (Umb, SP, Lat, Ven, Rae). Uncommon shapes include 𐌋 (Etr, Umb), 𐌎 (Etr), and the minor variations 𐌅 (Rae) and 𐌍 (Rae).

U+1030D 𐌍 EN en <n>

The forms of en basically correlate to those of em, if distributed somewhat differently: 𐌍 (Etr, Fal, Lat, Ven, Rae, Lep, Gal), 𐌇 (Etr, Osc, Umb, Fal, Ven), 𐌐 (Umb, Lat, Rae, Lep), and 𐌑 (Etr, Osc, Umb, Fal, Lat, Lep).

U+1030E 𐌎 ESH eš <š>

The letter eš (𐌎) is limited to Etruscan abecedaria. This letter is derived from Phoenician samekh <𐤎> via Greek ksi <Ξ>.

U+1030F 𐌏 O o <o>

The only widely attested forms for o are 𐌏 (Etr, Fal, Lat, Ven, Rae, Lep, Gal) and the squared northern ◇ (Ven, Lep). Early Etruscan also demonstrates a dotted form: ⊙. South Picene uses a unique form: · (single punct).

U+10310 𐌐 PE pē <p>

The exemplar form 𐌐/𐌿 is widely attested, present in Etruscan, Umbrian, Faliscan, Latin, Raetic, Lepontic, and Gallic. Venetic uses a form with an extra stroke, 𐌿, also found in Raetic and Etruscan. Greek Π-shaped letters appear in a few languages: 𐌀 (Etr, Osc, Lat) and 𐌚 (Etr, Osc, SP). Two unique forms also exist: Etruscan 𐌋 and Raetic 𐌕.

U+10311 𐌑 SHE śē <ś>

The letter šē is most common in its original Greek form: 𐌑 (Etr, Umb, Ven, Rae). A common variant is 𐌎 (Etr, Fal, Rae, Lep, Gal). Minor northern variants include 𐌚 (Lep), 𐌑 (Rae, Lep), and 𐌟 (Lep). This letter is derived from Phoenician tsade <𐤑> via Greek san <Ϻ>.

U+10312 𐌒 KU kū <q>

This letter appears in three major forms: 𐌒 (Etr, Lat), 𐌘 (Etr, Fal), and 𐌦 (Etr, SP). Minor forms 𐌃 (Etr) and 𐌏 (Fal) are also attested. The letter is unattested north of Etruria.

U+10313 𐌓 ER er <r>

The letter er is most common in its Greek P-like form: 𐌓/𐌿 (Etr, SP, Fal, Lat, Rae). In some southern and all northern Italic languages, the 𐌃/𐌡 (Etr, Osc, Umb, Ven, Rae, Lep, Gal) form is used. The familiar 𐌓 form is attested only in Faliscan and Latin. And Lepontic exhibits a unique, distinctly þ-like form: 𐌡. In spite of the northern Italic languages favoring 𐌃 over the exemplar 𐌓 shape, the writing systems are easily unified with Old Italic, with respect to this letter, just as Oscan and Umbrian, which display the same affinity, are.

U+10314 𐌔 ES es <s>

The letter es appears in 3-, 4-, and 6-stroke varieties, all easily unified: 𐌔 (Etr, Osc, Umb, SP, Fal, Lat, Ven, Rae, Lep, Gal), 𐌢 (Etr, SP, Fal, Ven, Rae, Lep, Gal), 𐌣 (Fal).

U+10315 𐌕 TE tē <t>

This letter's common form, varying slightly in cross-bar position and angle, is 𐌕/𐌕/𐌕/𐌕 (Etr, Osc, Umb, SP, Fal, Lat, Ven, Rae, Lep). Etruscan, Umbrian, and Faliscan also have the form 𐌔. And Faliscan uses the novel form 𐌅.

In northern Italy, the unique forms 𐌟 (Rae) and 𐌑 (Ven) are found. However, by far, the most common and widespread version of the grapheme in northern Italy is the St. Andrew's cross variety: 𐌗 (Ven, Rae, Lep, Gal).

U+10316 𐌖 U ū <u>

The letter ū appears in three Y-type shapes: 𐌖 (Etr, Osc, Lat), 𐌔 (Etr, Lep), and 𐌘 (Etr). More widespread throughout Italy is 𐌅 (Etr, Osc, Umb, SP, Fal, Lat, Ven, Rae, Lep, Gal). Less common are its inverted form 𐌍 (Ven, Rae) and 𐌋 (Etr). Though none of the northern Italic languages use Unicode's exemplar shape, neither do many southern languages, but all of the languages use 𐌅, suggesting that if the southern languages can be unified with Etruscan, so can the northern.

U+10317 𐌗 EKS eks <x>

Eks appears only in southern Italic, most often as 𐌗 (Etr, Osc, Umb, Fal, Lat). Etruscan also evidences a 𐌟 form.

U+10318 𐌘 PHE phē <φ>

Phē appears only in northern Italic and Etruscan, usually in the similar forms 𐌘/𐌦 (Etr, Ven, Rae) and 𐌆/𐌦 (Etr, Ven, Rae, Lep). Single-language northern variants include 𐌧 (Ven) and 𐌐 (Rae).

U+10319 𐌙 KHE khē <χ>

Khē appears only in northern Italic, Etruscan, and Faliscan, usually in the similar forms 𐌙 (Etr, Fal, Ven, Rae, Lep) and 𐌗/𐌉 (Etr, Fal, Ven, Rae, Lep). The inverted form 𐌑 is limited to Raetic.

U+1031A 𐌚 EF ef <f>

The Etruscan-invented letter ef, 𐌚, appears without much graphic variation in Etruscan, Oscan, and Umbrian. Faliscan appears to have invented its own form, 𐌑, for the same sound. South Picene simplified 𐌚 to : (double puncts). (Cf. South Picene's simplification of 𐌏 to · (single punct), noted above.) The letter is absent from northern Italic.

U+1031B 𐌛 ERS eř <ř>

This letter eř, 𐌛, is unique to Umbrian, without significant graphic variation.

U+1031C 𐌜 CHE çē <ç>

This letter çē, 𐌜, is unique to Umbrian, without significant graphic variation.

U+1031D 𐌝 II í <í>

Signs for í are present only in Oscan (𐌉/·𐌉) and, by independent invention, in South Picene (⋈).

U+1031E 𐌖 UU ú <ú>

Signs for ú are present only in Oscan (𐌖) and South Picene (𐌿).

3. Against the disunification of northern and southern Old Italic

The alternative position on the encoding of northern Old Italic is to disunify northern and southern Old Italic scripts. The latter would thus be represented by Unicode's existing Old Italic script while a new northern Old Italic script would be created to encode those characters used north of Etruria. Based on the text of TUS, in which Old Italic was identified as inappropriate for encoding the languages north of Etruria, this was the implicit position adopted by Unicode prior to the submission of this proposal.

However, although some of the northern Old Italic languages employ far fewer characters than the 26 known from the Marsiliana abecedarium (e.g. the 17 characters employed by Lepontic), there is only one character employed by the northern Old Italic languages that cannot be unified with Unicode's existing Old Italic script: the Raetic character 𐌡, proposed above. Raetic is thus the only northern Old Italic script that does not employ a strict subset of the characters present in the Old Italic block.
The table below identifies the variety of basic glyph shapes employed by the Venetic, Raetic, Gallic, and Lepontic writing systems, along with their most common Etruscan forms and other Etruscan forms that match glyphs seen north of Etruria. Yellow-highlighted Etruscan glyphs are not evidenced north of Etruria, and yellow-highlighted northern Old Italic glyphs are not evidenced in Etruscan. The table demonstrates that, although there are many northern Old Italic glyphs not seen in Etruscan, the set of Etruscan characters that have no corresponding glyph within the northern Old Italic languages consists of exactly those Etruscan characters that were dropped from northern Old Italic.

The case for disunification of the northern and southern Old Italic scripts is extremely weak. Two of the stronger cases for disunification are the glyph shapes of northern Old Italic 𐌀 <a> and 𐌗 <t> versus their typical Etruscan and southern Old Italic forms: 𐌀 <a> and 𐌕 <t>. The glyph 𐌀, however, continues to be used for <a> in the Venetic-speaking region after it transitions to the use of Latin. To suggest that Venetic 𐌀 and Etruscan 𐌀 cannot be unified would imply that Latin needs a new 𐌀-shaped character for Latin written in the area around Este. The character form 𐌗 <t> is unique to the languages north of Etruria and present in all of their writing systems, suggesting it may deserve independent encoding. However, it is clearly the basic form 𐌕 in a rotated orientation. Since its alphabetic position is identical to <t> in other Old Italic writing systems, it is most reasonable to unify the rotated northern Old Italic form 𐌗 with the unrotated forms 𐌕 and 𐌔, seen in the southern writing systems.

In contrast to this situation, South Picene has already been unified with the other Old Italic scripts although it has many more significant divergences from the forms seen in Etruscan and the rest of Old Italic. Typical South Picene glyphs (cf. TE-5 above) include 𐌆 for 𐌅 <v>, · for 𐌏 <o>, : for 𐌈 <f>, and 𐌉 for 𐌕 <t> in addition to characters invented for South Picene: 𐌋 <í>, 𐌖 <ú>, and 𐌘 <σ>. Thus the South Picene writing system is considerably more atypical and eccentric from the perspective of Etruscan or other Sabellian languages. Nevertheless, it was deemed unifiable with the Old Italic script.

Most arguments against unifying northern Old Italic scripts with the existing Old Italic Unicode block are based on the notion that the Old Italic block is not sufficient to the needs of scholars working on

Venetic, Raetic, Gallic, and/or Lepontic. Those scholars who are chiefly concerned with recording epigraphic data, such as numerous glyph variants, will not find their needs completely met by any proposal to encode the characters of the northern Old Italic scripts, whether they are unified with Old Italic, unified as a new Northern Old Italic script, or disunified as separate Venetic, Raetic, Gallic, and Lepontic scripts. Those users needing to encode distinct glyphs could be served by separate fonts implementing distinct glyph styles, as the standard currently suggests for users of southern Old Italic, or by adding variation selectors to Unicode. Both of these solutions fall outside the scope of the present proposal.

Etruscan	transliteration	Venetic	Raetic	Gallic	Lepontic
𐌀	a	𐌀 𐌀 𐌀	𐌀 𐌅 𐌀 𐌀 𐌀	𐌅	𐌀 𐌅 𐌀 𐌀
𐌁	b				
𐌂	c	𐌂	𐌂		
𐌃	d				
𐌄	e	𐌄	𐌄 𐌄	𐌄	𐌄
𐌅	v	𐌅	𐌅		𐌅
𐌉 𐌆	z	𐌆 𐌋 𐌗	𐌆 𐌆		𐌆
𐌇 𐌈	h	𐌈 𐌇 𐌇 𐌇	𐌈		
𐌗 𐌘	θ	𐌘 𐌗 𐌗			𐌘
𐌉	i	𐌉	𐌉 𐌉	𐌉	𐌉
𐌊	k	𐌊	𐌊	𐌊	𐌊
𐌋 𐌋	l	𐌋 𐌋	𐌋 𐌋	𐌋	𐌋 𐌋 𐌋
𐌌	m	𐌌 𐌌	𐌌 𐌌 𐌌 𐌌		𐌌
𐌍 𐌇 𐌍	n	𐌍 𐌇	𐌍 𐌍	𐌍	𐌍 𐌍 𐌍
𐌎	š				
𐌏 𐌄	o	𐌏 𐌄	𐌏	𐌏	𐌏 𐌄
𐌐 𐌐	p	𐌐	𐌐 𐌐 𐌕	𐌐	𐌐
𐌑 𐌑	ś	𐌑	𐌑 𐌑 𐌑	𐌑	𐌑 𐌑 𐌑 𐌑
𐌒	q				
𐌓 𐌃	r	𐌃	𐌓 𐌃	𐌃	𐌃 𐌃
𐌔 𐌔	s	𐌔 𐌔	𐌔 𐌔	𐌔 𐌔	𐌔 𐌔
𐌕 𐌕 𐌕	t	𐌗 𐌕 𐌕 𐌕	𐌗 𐌕 𐌕	𐌗	𐌗 𐌕
𐌖 𐌅 𐌉	u	𐌅 𐌖	𐌅 𐌖	𐌅	𐌅 𐌉
𐌗	x				
𐌚 𐌘	φ	𐌘 𐌚 𐌘	𐌘 𐌚 𐌓		
𐌘 𐌅	χ	𐌘 𐌅	𐌘 𐌅 𐌕		𐌘 𐌅
𐌆	f				
	t'		𐌁		

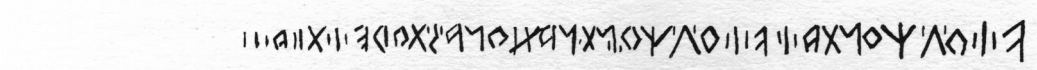
Fig. 3-1: Glyph variants in northern Old Italic compared to Etruscan

4. Encoded examples of inscriptions from northern Italy

The following section employs the existing (as of Unicode 6.2) Old Italic character repertoire in order to demonstrate its sufficiency for the encoding of Venetic, Raetic, Lepontic, and Gallic inscriptions, along with the Germanic Negau (Negova) helmet inscription.

4.1. Venetic

PELLEGRINI-PROSDOCIMI, *Lingua venetica*, *cit.*, Es 40, p. 143; LEJEUNE, *Manuel*, *cit.*, p. 204 e *passim*.



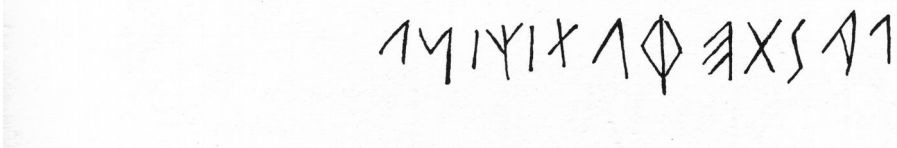
vho.u.χo.n.ta.i.vho.u.χo.n.tnazona.s.tore.i.tiia.i. =
Fougontai Fougontna donasto Reitiai

Fig. 4-1: (Morandi 1982:183)

Example encoding: 𐤆𐤇𐤏.𐤉.𐤉𐤏.𐤌.𐤕𐤀.𐤉.𐤆𐤇𐤏.𐤉.𐤉𐤏.𐤌.𐤕𐤌𐤀𐤆𐤏𐤍𐤀.𐤎.𐤕𐤏𐤓𐤄.𐤉.𐤕𐤉𐤉𐤀.𐤉.
<vho.u.χo.n.ta.i.vho.u.χo.n.tnazona.s.tore.i.tiia.i.>

4.2. Raetic

G. PELLEGRINI, in *NSc,* 1918, p. 183; *PID* 239.



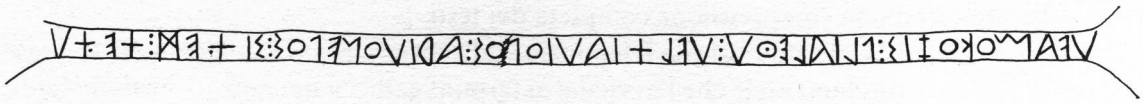
A. *laste φutiχinu*
B. *XI*(?)

Fig. 4-2: (Morandi 1982:199)

Example encoding: 𐤋𐤀𐤎𐤕𐤄 𐤘𐤕𐤉𐤉𐤉𐤌𐤉
<laste φutiχinu>

4.3. Lepontic

A. L. PROSDOCIMI, in *StEtr*, XXXV, 1967, p. 199 e segg.; LEJEUNE, *Lepontica, cit.*, p. 96 e segg.; TIBILETTI BRUNO, in *Lingue e dialetti*, p. 141.



uvamokozis plialeθu uvi̯tiauiopos ariuonepos siteś tetu

Fig. 4-3: (Morandi 1982:188)

Example encoding: 𐌅𐌚𐌰𐌼𐌾𐌺𐌾𐌾𐌹𐌹𐍃:𐌿𐌻𐌹𐌰𐌻𐌴𐍈𐌘:𐌅𐌚𐌹𐌸𐌹𐌰𐌿𐌹𐌾𐌾𐌿𐍃:𐌰𐌡𐌹𐌅𐌾𐌽𐌴𐌿𐍃:𐍃𐌹𐌸𐌴𐌼:𐌸𐌴𐌸𐌘
<uvamokozis:plialeθu:uvitiauiopos:ariuonepos:siteś:tetu>

4.4. Gallic

PID 337; M. LEJEUNE, in *Hommages à M. Niedermann, Collect. Latomus,* XXIII, 1956, pp. 206-215; LEJEUNE, *Lepontica, cit.,* p. 39 e segg.; TIBILETTI BRUNO, in *Lingue e dialetti,* pp. 155-156.

[1] *tanotaliknoi* [2] *kuitos* [3] *lekatos* [4] *anokopokios* [5] *setupokios* [6] *esanekoti* [7] *anareuiśeos* [8] *tanotalos* [9] *karnitus / takos toutas* [---] *pu* [---] / [---] *ṭeṭaso ṗoiḳan* (?)

Fig. 4-4: (Morandi 1982:192)

Example enncoding:

𐌕𐌀𐌍𐌏𐌕𐌀𐌋𐌉𐌊𐌍𐌏𐌉	<tanotaliknoi>
𐌊𐌖𐌉𐌕𐌏𐍃	<kuitos>
𐌋𐌄𐌊𐌀𐌕𐌏𐍃	<lekatos>
𐌀𐌍𐌏𐌊𐌏𐌐𐌏𐌊𐌉𐌏𐍃	<anokopokios>
𐍃𐌄𐌕𐌖𐌐𐌏𐌊𐌉𐌏𐍃	<setupokios>
𐌄𐍃𐌀𐌍𐌄𐌊𐌏𐌕𐌉	<esanekoti>
𐌀𐌍𐌀𐌓𐌄𐌖𐌉𐌌𐌄𐌏𐍃	<anareuiśeos>
𐌕𐌀𐌍𐌏𐌕𐌀𐌋𐌏𐍃	<tanotalos>
𐌊𐌀𐌓𐌍𐌉𐌕𐌖𐍃	<karnitus>

𐌕𐌀𐌊𐌏𐍃 𐌕𐌏𐌖𐌕𐌀𐍃 𐌐𐌖 𐌕𐌄𐌕𐌀𐍃𐌏 𐌐𐌏𐌉𐌊𐌀𐌍 <takos toutas pu tetaso poikan>

4.5. Germanic

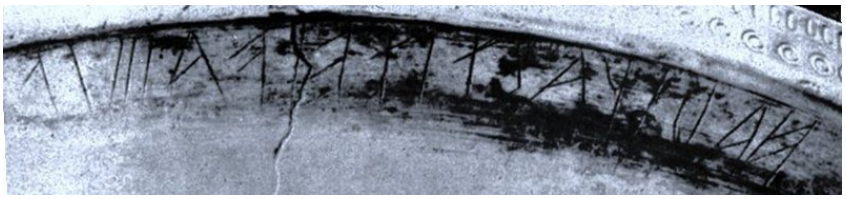


Fig. 4-5: Negau Helm B (from Negova, Slovenia; currently housed at the Kunst Historisches Museum, Vienna)

Example encoding: 𐍈𐌰𐍁𐌉𐍇𐌰𐍃𐌈𐌉 𐌈𐌴𐌹𐍆𐌰 𐌹𐌹𐌹 𐌹𐌻
<harix̣asti teiva iii il>
(traditionally rendered with <g> for <χ>, thus <harigasti teiva iii il>)

5. Recommendation for revised text of section 14.2 of the standard

The following revised text of section 14.2 of the Unicode Standard, with additions signaled via underline and removals signaled via strikethrough, incorporates the changes recommended above:

> **14.2 Old Italic**
>
> ***Old Italic: U+10300–U+1032F***
>
> The Old Italic script unifies a number of related historical alphabets located on the Italian peninsula. Some of these were used for non-Indo-European languages (Etruscan, Raetic, and probably North Picene), and some for various Indo-European languages belonging to the Italic branch (Faliscan and members of the Sabellian group, including Oscan, Umbrian, and South Picene) or other branches (Venetic, Lepontic, and Gallic). The ultimate source for the alphabets

in ancient Italy is Euboean Greek used at Ischia and Cumae in the bay of Naples in the eighth century BCE. Unfortunately, no Greek abecedaries from southern Italy have survived. Faliscan, Oscan, Umbrian, North Picene, ~~and~~ South Picene, Raetic, Venetic, Lepontic, and Gallic all derive from an Etruscan form of the alphabet.

There are some 10,000 inscriptions in Etruscan. By the time of the earliest Etruscan inscriptions, circa 700 BCE, local distinctions are already found in the use of the alphabet. Three major stylistic divisions are identified: the Northern, Southern, and Caere/Veii. Use of Etruscan can be divided into two stages, owing largely to the phonological changes that occurred: the "archaic Etruscan alphabet," used from the seventh to the fifth centuries BCE, and the "neo-Etruscan alphabet," used from the fourth to the first centuries BCE. Glyphs for eight of the letters differ between the two periods; additionally, neo-Etruscan abandoned the letters KA, KU, and EKS.

The unification of these alphabets into a single Old Italic script requires language-specific fonts because the glyphs most commonly used may differ somewhat depending on the language being represented.

Most of the languages have added characters to the common repertoire: Etruscan and Faliscan add LETTER EF; Oscan adds LETTER EF, LETTER II, and LETTER UU; Umbrian adds LETTER EF, LETTER ERS, and LETTER CHE; North Picene adds LETTER UU; ~~and~~ South Picene adds LETTER II, ~~and~~ LETTER UU, and LETTER ESS; and Raetic adds LETTER TTE.

The Latin script itself derives from a south Etruscan model, probably from Caere or Veii, around the mid-seventh century BCE or a bit earlier. However, because there are significant differences between Latin and Faliscan of the seventh and sixth centuries BCE in terms of formal differences (glyph shapes, directionality) and differences in the repertoire of letters used, this warrants a distinctive character block. Fonts for early Latin should use the uppercase code positions U+0041..U+005A. ~~The unified Alpine script, which includes the Venetic, Rhaetic, Lepontic, and Gallic alphabets, has not yet been proposed for addition to the Unicode Standard but is considered to differ enough from both Old Italic and Latin to warrant independent encoding. The Alpine script is thought to be the source for Runic, which is encoded at U+16A0..U+16FF. (*See Section 14.3, Runic*.)~~

Character names assigned to the Old Italic block are unattested but have been reconstructed according to the analysis made by Sampson (1985). While the Greek character names (alpha, beta, gamma, and so on) were borrowed directly from the Phoenician names (modified to Greek phonology), the Etruscans are thought to have abandoned the Greek names in favor of a phonetically based nomenclature, where stops were pronounced with a following -e sound, and liquids and sibilants (which can be pronounced more or less on their own) were pronounced with a leading e- sound (so [k], [d] became [ke:], [de:] ~~became~~while [l:], [m:] became [el], [em]). It is these names, according to Sampson, which were borrowed by the Romans when they took their script from the Etruscans.

Directionality. Most early Etruscan texts have right-to-left directionality. From the third century BCE, left-to-right texts appear, showing the influence of Latin. ~~Oscan, Umbrian, and~~

~~Faliscan~~The other scripts unified under Old Italic also generally have right-to-left directionality. Boustrophedon appears rarely, and not especially early (for instance, the Forum inscription dates to 550–500 BCE). Despite this, for reasons of implementation simplicity, many scholars prefer left-to-right presentation of texts, as this is also their practice when transcribing the texts into Latin script. Accordingly, the Old Italic script has a default directionality of strong left-to-right in this standard. If the default directionality of the script is overridden to produce a right-to-left presentation, the glyphs in Old Italic fonts should also be mirrored from the representative glyphs shown in the code charts. This kind of behavior is not uncommon in archaic scripts; for example, archaic Greek letters may be mirrored when written from right to left in boustrophedon.

Punctuation. The earliest inscriptions are written with no space between words in what is called scriptio continua. There are numerous Etruscan inscriptions with dots separating word forms, attested as early as the second quarter of the seventh century BCE. This punctuation is sometimes, but only rarely, used to separate syllables rather than words. From the sixth century BCE, words were often separated by one, two, or three dots spaced vertically above each other. Venetic inscriptions typically indicate syllable-initial vowels and syllable-final consonants with vertical strokes or dots on both sides of the letter.

Numerals. Etruscan numerals are not well attested in the available materials, but are employed in the same fashion as Roman numerals. Several additional numerals are attested, but as their use is at present uncertain, they are not yet encoded in the Unicode Standard.

Glyphs. The default glyphs in the code charts are based on the most common shapes found for each letter. Most of these are similar to the Marsiliana abecedary (mid-seventh century BCE). Note that the phonetic values for U+10317 old italic letter eks [ks] and U+10319 OLD ITALIC LETTER KHE [kh] show the influence of western, Euboean Greek; eastern Greek has U+03A7 GREEK CAPITAL LETTER CHI [x] and U+03A8 GREEK CAPITAL LETTER PSI [ps] instead.

The geographic distribution of the Old Italic script is shown in *Figure 14-1*. In the figure, the approximate distribution of the ancient languages that used Old Italic alphabets is shown in white. Areas for the ancient languages that used other scripts are shown in gray, and the labels for those languages are shown in oblique type. In particular, note that the ancient Greek colonies of the southern Italian and Sicilian coasts used the Greek script proper. ~~Also, languages such as Ligurian, Venetic, and so on, of the far north of Italy made use of alphabets of the Alpine script.~~ Rome, of course, is shown in gray, because Latin was written with the Latin alphabet, now encoded in the Latin script.

In addition to the revised text, the map appearing at the end of this section, on page 469 of the standard, should be revised to remove the gray shading of the Gallic, Lepontic, Raetic, and Venetic territories. The spelling of *Rhaetic* should also be corrected to *Raetic*, or *Raetic* should be made to conform to this spelling in the text supplied above.

This map also implies that Old Italic should be employed for the encoding of the languages of Sicily (Elymian, Sicanian, and Siculan). However, there is no suggestion or evidence in the original proposal (L2/00-140) to indicate that these are not written in a Greek script. It seems advisable to gray these

Sicilian language territories (or omit the islands from the map altogether). Minimally, the spelling of Elymian should be corrected.

References

Bonfante, Larissa. 1996. “The Scripts of Italy.” In *The World's Writing Systems*, eds. Peter T. Daniels and William Bright. Oxford: Oxford University Press.

Conway, R. S. 1897. *The Italic Dialects*. Cambridge: Cambridge University Press.

Diringer, David. 1968. *The Alphabet: A Key to the History of Mankind.* London: Hutchinson & Co.

Faulmann, Carl. 1880. *Das Buch der Schrift.* Vienna: Kaiserlich-Königlichen Hof- und Staatsdruckerei.

Jensen, Hans. 1969. *Die Schrift in Vergangenheit und Gegenwart*. Berlin: Veb.

Marinetti, Anna. 1985. *Le iscrizioni sudpicene. I. Testi.* Firenze: Leo S. Olschki.

Morandi, Alessandro. 1982. *Epigrafia italica.* Roma: L'Erma di Bretschneider.

Rix, Helmut. 2002. *Sabellische Texte.* Heidelberg: C. Winter.

Schumacher, Stefan. 1992. *Die Rätischen Inscriften: Geschichte und heutiger Stand der Forschung.* Innsbruck: Verlag der Instituts für Sprachwissenschaft der Universität Innsbruck.