Unicode request for Romance dialectology symbols

Nicolas Datua, nicolasdatua, gmail.com   2025 October 09
Kirk Miller, kirkmiller, gmail.com

The present proposal concerns three systems of phonetic notation used in nearly all French, Italian, and Spanish linguistic atlases of the 20th century,[1] as well as a number of more recent and ongoing publications.[2] The vast corpus of dialect data recorded in these atlases remains, on the whole, under-utilized by researchers for lack of accessibility. Numerous initiatives[3] aim to digitize these atlases, all of them encountering difficulties on account of various symbols missing from the Unicode standard. This compels every project to devise *ad hoc* solutions for the notation of its particular atlas,[4] an inconvenience for new projects and a complication for cross-project data sharing and comparison. We quote two relevant comments from the authors of such projects:

> 'We hope that future extensions of Unicode will fill these code point gaps for Romance transcription systems (as it has been done with the Teuthonista transcription system, see [Revised proposal to encode "Teuthonista" phonetic characters in the UCS](#)).' — Geisler et al. 2021: 25 (footnote 11).

> 'Unicode non ha la possibilità di rappresentare questo tipo di formattazione del testo, e il repertorio di lettere sotto- e sovrascritte che offre è troppo esiguo per soddisfare le esigenze dell'ALI' [Unicode offers no way to represent this sort of text formatting, and the inventory of sub– and superscript letters is too limited to satisfy the needs of the ALI (*Atlante linguistico italiano*).] — Cerruti et al. 2025: 44.

Our proposal aims, therefore, to facilitate current and future atlas digitization projects by adding the symbols in question to the Unicode standard.

We now discuss the three notation systems.

1. Rousselot-Gilliéron (RG)

The RG system was devised by Jean-Pierre Rousselot in the late 19th century and was used by Jules Gilliéron for the monumental *Atlas linguistique de la France* (ALF), the world's first linguistic atlas in cartographic form. The system was then adopted by the authors of the *Tableaux phonétiques des patois suisses romands* (TPh),[5] and later by the Centre national de la recherche scientifique (CNRS) for a series

---

[1]A linguistic atlas typically consists of a series of maps, each showing the responses of dialect speakers to a given prompt, such as 'How does one say "sky" in your dialect?' Traditionally the responses had to be transcribed *in situ* by the interviewer, in phonetic notation. It is only in the late 20th century that audio equipment became sufficiently accurate and portable to be relied on for field-work.

[2]See also [L2/14-169](#) *Proposal to encode additional dialectology Latin characters* by Denis Jacquerye.

[3]A partial list is provided in the bibliography below. Included are some atlases not cited in this proposal.

[4]Some avoid the issue by foregoing any direct representation of the original transcriptions, providing only conversions to the International Phonetic Alphabet (IPA). Unfortunately there is often no singular, unambiguous IPA equivalent for a given symbol or symbol-combination found in the atlases, and different projects may arrive at different approximations for what were (in reality) the same sounds, or vice-versa. In any case it is helpful to save a faithful representation of the original data for automatic conversion to IPA. This prevents continuity problems when one wants to modify the conversion rules, after having already processed a large body of transcriptions. It also allows projects to apply their own conversion rules to raw data from other projects, for increased comparability of phonetic data.

[5]The fieldwork for the TPh used a notation typical of German dialectology at the time, but during the publication process it was decided to convert the transcriptions to the RG system to increase comparability with the ALF. Incidentally, the TPh is not a linguistic atlas *sensu stricto*, since the data are published in tabular rather than cartographic form, but it had been planned as one, and its methodology

of atlases dedicated to specific regions, of which we cite eight: the ALAL, ALCe, ALJA, ALLy, ALN, ALO, ALPic and ALRé.

The base characters of the RG system are as follows.[6] On the second line we provide approximations, in the International Phonetic Alphabet (IPA), of the sounds that these symbols indicate in isolation.

RG:	a	b	c	ɛ	d	e	f	g	h	i	j	k	l	m	n	o	œ	p	r	s	t	u	ʋ	v	w	y	z
IPA:	ä	b	⁻[7]	ʃ	d	ɇ	f	g	h	i	ʒ	k	l	m	n	ǫ	œ	p	r	s	t	y	u	v	w	j	z

Many of the base characters, especially the vowels, may be given diacritics to denote other phonetic values. Nearly all the diacritics are already part of the Unicode standard.[8] The exception is the curled tilde ⟨ ̃ ⟩, which denotes partial nasalization. This is distinguished from the standard tilde ⟨ ̃ ⟩, which denotes full nasalization (cf. figure 2).

A key feature of the RG system, as well as the other systems discussed below, is the use of combining letters for intermediate sounds. For instance, ⟨å⟩ (IPA: [æ]) denotes a vowel that is acoustically between [a] and [e]. Similarly, ⟨š⟩ (IPA: [ʂ][9]) denotes a consonant with a degree of voicing between that of [s] and [z]. In principle any RG base character can be used as a combining letter, and indeed all of them are attested as such. The base characters ⟨ɛ j œ ʋ y⟩ lack combining equivalents in Unicode, so we propose them for inclusion (cf. figures 7, 12, 16, 22, 24).

Another feature of this system is the use of subscript letters[10] to denote weakly articulated sounds, as in ⟨ᵢ⟩ (extIPA: [i̯]). The base characters ⟨b c ɛ d f g œ ʋ⟩ lack subscript equivalents in Unicode, so we propose them for inclusion (cf. figures 44–54).

Some of the later CNRS atlases make use of spacing superscripts to denote weakly articulated sounds, either instead of subscripts (ALPic) or alongside them in specialized usages (ALJA, ALRé[11]). The base character ⟨ɛ⟩ lacks a superscript equivalent in Unicode, so we propose this for inclusion (Figure 35).

Of the atlases cited in this proposal, which use the RG system, the one with the most recent publication is the ALN, the fifth volume of which appeared in 2019. This was one of several volumes that had been left unpublished when the CNRS abruptly halted its regional atlas project in the 1990's. Many such volumes still await publication today.[12]

---

and scientific value are the same. We therefore include the TPh among the atlases in this proposal.

[6]By 'base characters', we mean the ones from which the rest may be built through combination or diacritic attachment. This is not a complete inventory of the basic sounds of the system, in which we would have included for instance [ė] and [n̪] (IPA: [ə], [ɲ]). The same is true of the other character inventories given below.

[7]⟨c⟩ never occurs in bare form, only in combinations such as ⟨ĉ⟩ and ⟨ç⟩ (IPA: [χ], [ç]). A similar situation holds for ⟨c⟩ in the other notations presented here.

[8]An overview of the diacritics already in Unicode, and their usage in Romance dialectology, is beyond the scope of this proposal.

[9]This IPA equivalence, like many of the others in this proposal, is only one of multiple possibilities. RG ⟨š⟩ could also be approximated as IPA ⟨ʐ⟩. Here it bears mentioning that the superpositions of Romance dialectological notation often imply values closer to the base character than to the combining one, hence our choice of the IPA ⟨ʂ⟩.

[10]The early RG notation, as it appears in the ALF and TPh, made use of small characters sitting on the baseline, and not subscripts in the strict sense. For the later CNRS series, it was decided to move the small characters below the line, so as to increase their visual distinction from regular characters. In light of this equivalence, we treat the early small characters as subscripts and do not suggest encoding them as a separate kind of letter.

[11]The ALJA uses ⟨ʰ⟩ to denote what appears to be a glottal stop, and the ALRé uses ⟨ʳ⟩ to denote vocalic outcomes of an original rhotic.

[12]Brun-Trigaud 2016 provides a useful overview of these.

2. Böhmer-Ascoli (BA)

The BA system is named for Eduard Böhmer and Graziadio Isaia Ascoli, and it takes after their phonetic notations. It was notably used for the *Atlante italo-svizzero* (AIS) with the following base characters:

The BA system was devised in the late 19th century by Eduard Böhmer, who took the phonetic notation of Graziadio Isaia Ascoli as a starting point. It was notably used for the *Atlante italo-svizzero* (AIS),[13] where it appears with the following base characters:

BA: a b c d e ǝ f g h i k l ł ⱡ ɫ m ɱ n ŋ o œ p r s t u v w y z α β γ δ θ φ χ ʕ
IPA: ä b -[14] d ɛ ə f g h i k l ʎ/lʲ l⁽ʲ⁾ ł m m͡ŋ n ŋ ɔ̯ œ p r s t u v w j z ɐ β ɣ ð θ ɸ ç ʔ

The diacritics missing from Unicode are:

- ⟨◌̃⟩, which denotes partial nasalization, as in the RG system (extIPA: [◌̃]; fig. 2).

- ⟨◌̋⟩, which is used in the combinations ⟨č⟩ and ⟨ǧ⟩ (IPA: [c], [ɟ]; fig. 3). *Cf* the spacing diacritic U+1FCE ɢʀᴇᴇᴋ ᴘsɪʟɪ ᴀɴᴅ ᴏxɪᴀ ⟨῎⟩.

This system is largely maintained in the later *Atlant linguistich dl ladin dolomitich y di dialec vejins* (ALD) and *Vivaio acustico delle lingue e dei dialetti d'Italia* (VIVALDI). We note the following modifications:

ALD and VIVALDI:

- Abandonment of ⟨◌̃⟩ and ⟨ⱡ⟩

- Phonetic repurposing of ⟨ɱ⟩ (IPA: [ɱ])

- Addition of ⟨đ ø⟩ (IPA: [d̪ð̠], [œ̽])

ALD only:

- Addition of ⟨ɐ⟩ (IPA: [ɐ])[15]

VIVALDI only:

- Addition of ⟨ʁ ɫ χ⟩ (IPA: [ʁ], [t̪θ], [χ])

On the other hand, the system in the *Atlante linguistico italiano* (ALI) shows core structural differences, to the extent that it requires a separate description:

ALI: a b c d e f g h i j k l ɫ ł m n o p r ɾ ʂ s ʃ t u v w y z ʒ α β γ δ ε ζ θ λ ƛ ρ σ φ χ ω
IPA:[16] ɑ b – d ɛ f g ç i j k l ļ ? m n ɔ̯ p r z̧ˀ ş̌ˀ s z t u v w i̧ ts d̠z ʒ β ɣ ð ə z̧ θ ɫ ɫ̧ˀ ʁ ş ɸ x ɐ

In all of these atlases, weak articulation is indicated with modifier letters (spacing superscripts), while combining letters are used for intermediate sounds. Theoretically, any base character might occur in superscript and combining forms. For now, we list those that we have encountered to date. In all, we propose the following additions for this notation:

---

[13]Also known by its original title, *Sprach- und Sachatlas Italiens und der Südschweiz*.
[14]⟨c⟩ never occurs in bare form, only in combinations such as ⟨ć⟩ (IPA: [t͡ʃ]).
[15]Thus occupying the position held in the AIS and ALD by ⟨α⟩, which is assigned a posterior position in the ALD (IPA: [ɐ]).
[16]Some of these interpretations are, for the moment, unconfirmed.

- Base characters: ꜱ (fig. 34)

- Combining diacritics: ◌̃ ◌̆ (figs. 2, 3)

- Combining letters: đ ɉ ł ŋ œ Ʒ ꜱ ʃ ӡ ʏ ɣ δ ζ θ φ χ σ (figs. 8, 12, 13, 15, 16, 21, 25, 19, 20, 24, 26, 27, 28, 29, 31, 32, 30)

- Modifier letters: Ʒ ꜱ ζ ρ σ (figs. 38, 39, 40, 41, 42)

Three of the four atlases cited here—namely, all but the AIS—have had publications in the last fifteen years. The latest is volume 10 of the ALI, published in 2023. Volume 11 is in preparation as of the time of writing.

3. *Revista de Filología Española* (RFE)

The RFE system was devised in the early 20th century by Tomás Navarro Tomás for use in the academic journal for which it is named. In the form that appears in the *Atlas Lingüístico de la Península Ibérica* (ALPI), the base characters are as follows. (The analysis of some is tentative.)

```
RFE: a b ƀ ƃ c d ḍ đ e ə f g g g̱ h ħ i j k l ꞁ ł m n ŋ ꞥ o ɔ p r ɾ ɹ s t u ᴜ v w x y z θ φ
IPA: ä b ḅ β - d ḍ ð ɛ ə f g g̱ ɣ h ħ i j k l ꞁ̣ ł m n ŋ ꞥ ọ ö̤ p r ɾ ɹ s t u ŭ v w x ʝ z θ ɸ
```

In the later series of regional atlases directed by Manuel Alvar, of which we cite two (the ALEA and ALEANR), we note the following differences:

- Absence of ⟨ƀ đ ǥ ꞁ ꞥ⟩

- ⟨ʉ⟩ in place of ⟨ᴜ⟩[17]

As in the BA system, weak articulation is indicated with superscripts, while combining letters are used to indicate intermediate phonetic values.

We propose the following additions:

- Base characters: ꞁ (fig. 33)

- Combining letters: ʚ ƀ đ ǥ ɉ ħ ꞁ ɔ ɹ ʉ y θ φ (figs. 5–6, 8, 9, 12, 10, 14, 17, 18, 23, 24, 29, 31)

- Modifier letters: ꞁ ꞥ (figs. 36, 37)

Two recent projects published in this notation are the *Atlas Lingüístico y etnográfico de Castilla – La Mancha* (AleCMAN) and the *Cartografía lingüística de Extremadura* (CLEx).

---

[17]Perhaps this substitution was intended to increase the visual distinction from ⟨ü⟩ (IPA: [y]), particularly in handwriting. In any case, we would suggest encoding Alvar's ⟨ʉ⟩ as ⟨ᴜ⟩ for consistency; some non-distinctive examples of ⟨ᴜ⟩ do appear in his atlases (cf. ALEA map 4, point Ca 100: *trɛbᴜ̯éne*).

<u>Particularities</u>

a. Pseudo-subscripts

Since the combination of a base character and combining letter may become rather tall, especially with diacritics, sometimes the entire combination is shifted downward to compensate, as if the base character were a subscript. This occurs systematically in the ALI, as in Figure 1b. Since these pseudo-subscripts are a typographical device, without phonetic significance, they should be encoded as normal baseline characters.

Figure 1a. ① ⟨š⟩ in ALPI map 27, point 569. The *s* is reduced in size to accommodate the tall *θ* but is set on the baseline. ② ⟨ŏ⟩ in ALI map 9, point 394. The *σ* is set as a subscript without semantic distinction from a baseline *σ*, as are all combined letters in the ALI,

b. Offset combining letters

When a base letter has an ascender, such as *k*, a combining letters may be offset to the right and may thus resemble a superscript modifier letter. This is the case with the *h* in ⟨bǫ́ᵏkả⟩ in Figure 1b, where it can only be a combining letter, but is also seen in Figure 9. Where possible, we have avoided such tokens in the illustrative figures.
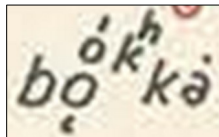
Figure 1b. ⟨bǫ́ᵏkả⟩ in ALPI map 2, point 532, with a laterally offset combining *h*.

Characters

Diacritics

◌̃ 1AF1 COMBINING CURLED TILDE / INVERTED LAZY S. Figure 2, Figure 35.

◌̋ 1AF2 COMBINING PSILI AND OXIA. Figure 3.

Combining letters

◌ɐ 1AF3 COMBINING LATIN SMALL LETTER TURNED A. Figure 4.

◌ƀ 1AF4 COMBINING LATIN SMALL LETTER B WITH STROKE. Figure 5 *ff*.

◌ȼ 1AF5 COMBINING LATIN SMALL LETTER C WITH BAR. Figure 7.

◌đ 1AF6 COMBINING LATIN SMALL LETTER D WITH STROKE. Figure 8.

◌ɦ 1AF7 COMBINING LATIN SMALL LETTER H WITH HOOK. Figure 10.

◌ı 1AF8 COMBINING LATIN SMALL LETTER DOTLESS I. Figure 11, Figure 62.

◌j 1AF9 COMBINING LATIN SMALL LETTER J. Figure 12.

◌ɫ 1AFA COMBINING LATIN SMALL LETTER L WITH MIDDLE TILDE. Figure 13.

◌ŋ 1AFB COMBINING LATIN SMALL LETTER ENG. Figure 15.

◌œ 1AFC COMBINING LATIN SMALL LIGATURE OE. Figure 16.

◌ʉ 1AFD COMBINING LATIN SMALL LETTER U WITH LEFT HOOK. Figure 22.

◌y 1AFE COMBINING LATIN SMALL LETTER Y. Figure 24.

◌ʒ 1AFF COMBINING LATIN SMALL LETTER EZH. Figure 25.

◌γ xxx0 COMBINING GREEK SMALL LETTER GAMMA. Figure 26.

◌δ xxx1 COMBINING GREEK SMALL LETTER DELTA. Figure 27.

◌ζ xxx2 COMBINING GREEK SMALL LETTER ZETA. Figure 28.

◌θ xxx3 COMBINING GREEK SMALL LETTER THETA. Figure 29.

◌σ xxx4 COMBINING GREEK SMALL LETTER SIGMA. Figure 30.

◌φ xxx5 COMBINING GREEK SMALL LETTER PHI. Figure 31.

◌χ xxx6 COMBINING GREEK SMALL LETTER CHI. Figure 32.

◌ɾ xxx7 COMBINING LATIN SMALL LETTER SCRIPT R. Figure 19.

◌ɹ xxx8 COMBINING LATIN SMALL LETTER REVERSED SCRIPT R. Figure 20.

◌ɹ xxx9 COMBINING LATIN SMALL LETTER TURNED R. Figure 18.

◌ʃ xxxA COMBINING LATIN SMALL LETTER ESH. Figure 21.

◌g xxxB COMBINING LATIN SMALL LETTER G WITH STROKE. Figure 9.

◌ł xxxC COMBINING LATIN SMALL LETTER L WITH STROKE. Figure 14.

◌ɔ xxxD COMBINING LATIN SMALL LETTER SPLIT O. Figure 17.

◌ü xxxE COMBINING LATIN SMALL LETTER SPLIT U. Figure 23.

Subscript letter

ʮ AB6F LATIN SUBSCRIPT SMALL LETTER U WITH LEFT HOOK. Figure 61.

Baseline letters

ł   1DF8B LATIN SMALL LETTER L WITH HALF STROKE. Figure 33.

ꟻ   1DF8C LATIN SMALL LETTER REVERSED SCRIPT R. Figure 34.

Modifier letters

ᶻ   1DFB6 MODIFIER LETTER SMALL ZETA. Figure 40.

ᵖ   1DFB7 MODIFIER LETTER SMALL RHO. Figure 41.

ᵟ   1DFB8 MODIFIER LETTER SMALL SIGMA. Figure 42.

ᶜ   1DFB9 MODIFIER LETTER SMALL C WITH BAR. Figure 35.

ˡ   1DFBA MODIFIER LETTER SMALL L WITH HALF STROKE. Figure 36.

ᵑ   1DFBB MODIFIER LETTER SMALL BARRED ENG. Figure 37.

ʳ   1DFBC MODIFIER LETTER SMALL SCRIPT R. Figure 38.

ʵ   1DFBD MODIFIER LETTER SMALL REVERSED SCRIPT R. Figure 39.

Subscript letters

ϲ   1DFBE LATIN SUBSCRIPT SMALL LETTER C WITH BAR. Figure 47.

œ   1DFBF LATIN SUBSCRIPT SMALL LIGATURE OE. Figure 60.

b   1DFC0 LATIN SUBSCRIPT SMALL LETTER B. Figure 44 *ff.*

c   1DFC1 LATIN SUBSCRIPT SMALL LETTER C. Figure 45, Figure 48 *ff.*

d   1DFC2 LATIN SUBSCRIPT SMALL LETTER D. Figure 45, Figure 52 *ff.*

f   1DFC3 LATIN SUBSCRIPT SMALL LETTER F. Figure 55 *ff.*

g   1DFC4 LATIN SUBSCRIPT SMALL LETTER G. Figure 58 *ff.*

Properties

Some letters are named 'Greek' because they have a Greek rather than Latin form (for example, following Greek γ and φ rather than Latin ɣ and ɸ), but are nonetheless intended for use in Latin text.

The first character might be called either 'curled tilde' or 'inverted lazy s.' However, in one source it is not inverted.

Code points in red are proposed or in the pipeline and are subject to change.

1AF1;COMBINING CURLED TILDE / INVERTED LAZY;Mn;230;NSM;;;;;N;;;;;
1AF2;COMBINING PSILI AND OXIA;Mn;230;NSM;;;;;N;;;;;
1AF3;COMBINING LATIN SMALL LETTER TURNED A;Mn;230;NSM;;;;;N;;;;;
1AF4;COMBINING LATIN SMALL LETTER B WITH STROKE;Mn;230;NSM;;;;;N;;;;;
1AF5;COMBINING LATIN SMALL LETTER C WITH BAR;Mn;230;NSM;;;;;N;;;;;
1AF6;COMBINING LATIN SMALL LETTER D WITH STROKE;Mn;230;NSM;;;;;N;;;;;
1AF7;COMBINING LATIN SMALL LETTER H WITH HOOK;Mn;230;NSM;;;;;N;;;;;
1AF8;COMBINING LATIN SMALL LETTER DOTLESS I;Mn;230;NSM;;;;;N;;;;;
1AF9;COMBINING LATIN SMALL LETTER J;Mn;230;NSM;;;;;N;;;;;
1AFA;COMBINING LATIN SMALL LETTER L WITH MIDDLE TILDE;Mn;230;NSM;;;;;N;;;;;
1AFB;COMBINING LATIN SMALL LETTER ENG;Mn;230;NSM;;;;;N;;;;;
1AFC;COMBINING LATIN SMALL LIGATURE OE;Mn;230;NSM;;;;;N;;;;;
1AFD;COMBINING LATIN SMALL LETTER U WITH LEFT HOOK;Mn;230;NSM;;;;;N;;;;;
1AFE;COMBINING LATIN SMALL LETTER Y;Mn;230;NSM;;;;;N;;;;;
1AFF;COMBINING LATIN SMALL LETTER EZH;Mn;230;NSM;;;;;N;;;;;

xxx0;COMBINING GREEK SMALL LETTER GAMMA;Mn;230;NSM;;;;;N;;;;;
xxx1;COMBINING GREEK SMALL LETTER DELTA;Mn;230;NSM;;;;;N;;;;;
xxx2;COMBINING GREEK SMALL LETTER ZETA;Mn;230;NSM;;;;;N;;;;;
xxx3;COMBINING GREEK SMALL LETTER THETA;Mn;230;NSM;;;;;N;;;;;
xxx4 COMBINING GREEK SMALL LETTER SIGMA;Mn;230;NSM;;;;;N;;;;;
xxx5;COMBINING GREEK SMALL LETTER PHI;Mn;230;NSM;;;;;N;;;;;
xxx6;COMBINING GREEK SMALL LETTER CHI;Mn;230;NSM;;;;;N;;;;;
xxx7;COMBINING LATIN SMALL LETTER SCRIPT R;Mn;230;NSM;;;;;N;;;;;
xxx8;COMBINING LATIN SMALL LETTER REVERSED SCRIPT R;Mn;230;NSM;;;;;N;;;;;
xxx9;COMBINING LATIN SMALL LETTER TURNED R;Mn;230;NSM;;;;;N;;;;;
xxxA;COMBINING LATIN SMALL LETTER ESH;Mn;230;NSM;;;;;N;;;;;
xxxB;COMBINING LATIN SMALL LETTER G WITH STROKE;Mn;230;NSM;;;;;N;;;;;
xxxC;COMBINING LATIN SMALL LETTER L WITH STROKE;Mn;230;NSM;;;;;N;;;;;
xxxD;COMBINING LATIN SMALL LETTER SPLIT O;Mn;230;NSM;;;;;N;;;;;
xxxE;COMBINING LATIN SMALL LETTER SPLIT U;Mn;230;NSM;;;;;N;;;;;

AB6F;COMBINING LATIN SMALL LETTER U WITH LEFT HOOK;Mn;230;NSM;;;;;N;;;;;

1DF8B;LATIN SMALL LETTER L WITH HALF STROKE;Ll;0;L;;;;;N;;;;;
1DF8C;LATIN SMALL LETTER REVERSED SCRIPT R;Ll;0;L;;;;;N;;;;;

1DFB6;MODIFIER LETTER SMALL ZETA;Lm;0;L;<super> 03B6;;;;N;;;;;
1DFB7;MODIFIER LETTER SMALL RHO;Lm;0;L;<super> 03C1;;;;N;;;;;
1DFB8;MODIFIER LETTER SMALL SIGMA;Lm;0;L;<super> 03C3;;;;N;;;;;
1DFB9;MODIFIER LETTER SMALL C WITH BAR;Lm;0;L;<super> A793;;;;N;;;;;
1DFBA MODIFIER LETTER SMALL L WITH HALF STROKE;Lm;0;L;<super> 1DF8B;;;;N;;;;;
1DFBB MODIFIER LETTER SMALL BARRED ENG;Lm;0;L;<super> 1DF4F;;;;N;;;;;
1DFBC MODIFIER LETTER SMALL SCRIPT R;Lm;0;L;<super> AB4B;;;;N;;;;;
1DFBD MODIFIER LETTER SMALL REVERSED SCRIPT R;Lm;0;L;<super> 1DF8C;;;;N;;;;;

1DFBE;LATIN SUBSCRIPT SMALL LETTER C WITH BAR;Lm;0;L;<sub> A793;;;;N;;;;;
1DFBF;LATIN SUBSCRIPT SMALL LIGATURE OE;Lm;0;L;<sub> 0153;;;;N;;;;;
1DFC0;LATIN SUBSCRIPT SMALL LETTER B;Lm;0;L;<sub> 0062;;;;N;;;;;
1DFC1;LATIN SUBSCRIPT SMALL LETTER C;Lm;0;L;<sub> 0063;;;;N;;;;;
1DFC2;LATIN SUBSCRIPT SMALL LETTER D;Lm;0;L;<sub> 0064;;;;N;;;;;
1DFC3;LATIN SUBSCRIPT SMALL LETTER F;Lm;0;L;<sub> 0066;;;;N;;;;;
1DFC4;LATIN SUBSCRIPT SMALL LETTER G;Lm;0;L;<sub> 0067;;;;N;;;;;

Annotations

1AF2 COMBINING PSILI AND OXIA
→ 1FCE GREEK PSILI AND OXIA
xxx1 COMBINING GREEK SMALL LETTER DELTA
→ 1DD8 COMBINING LATIN SMALL LETTER INSULAR D
1DFC1 LATIN SUBSCRIPT SMALL LETTER C
• supports 0327 combining cedilla

Charts

Combining Diacritical Marks Extended +

| 1AB0 | | | | | | 1ADF |

	1AB	1AC	1AD	1AE	1AF	xxx	xxy
0	◌᪰	◌ᰀ	◌ᴠ	◌ꞁ	◌”	◌ϒ	
1	◌᪱	◌ᰁ	◌ᴎ	◌ᴕ	◌∞	◌δ	
2	◌᪲	◌ᰂ	◌ᴧ	◌⁼	◌ʺ	◌ζ	
3	◌᪳	◌ᰃ	◌ᴧ	◌ᴗ	◌ᵽ	◌θ	
4	◌᪴	◌ᰄ	◌ᴸ	◌□	◌ƀ	◌σ	
5	◌ˣˣ	◌[̣]	◌˩	◌ᵾ	◌ɛ	◌φ	
6	◌ᵚ	◌#	◌ᴝ	◌ᴈ	◌đ	◌χ	
7	◌ᵢ	◌ᵂ	◌ᴻ	◌ᴈ	◌ɦ	◌z	
8	◌ᵤ	◌+	◌̂	◌=	◌ˡ	◌ʒ	
9	◌ᵣ	◌++	◌#	◌ᴉ	◌j	◌ɹ	
A	◌ᵣ	◌₊₊	◌♭	◌↑	◌ł	◌ʃ	
B	◌◌	◌‴	◌ᴛ	◌→	◌ŋ	◌g	
C	◌«◌»	◌ᵹ	◌˙˙	◌ᵂ	◌œ	◌ɫ	
D	◌(◌)	◌ɳ	◌.◌	◌ᵂ	◌ɹ	◌ɔ	
E	◌(◌)	◌ᴛ	◌◌	◌ᵥ	◌y	◌ʊ	
F	◌ᵥ	◌ᵂ	◌˙◌	◌”	◌ʒ		

Latin Extended-E

	AB3	AB4	AB5	AB6
0	ꬰ	ꭀ	ꭐ	ꭠ
1	ꬱ	ꭁ	ꭑ	ꭡ
2	ꬲ	ꭂ	ꭒ	ꭢ
3	ꬳ	ꭃ	ꭓ	ꭣ
4	ꬴ	ꭄ	ꭔ	ꭤ
5	ꬵ	ꭅ	ꭕ	ꭥ
6	ꬶ	ꭆ	ꭖ	ꭦ
7	ꬷ	ꭇ	ꭗ	ꭧ
8	ꬸ	ꭈ	ꭘ	ꭨ
9	ꬹ	ꭉ	ꭙ	ꭩ
A	ꬺ	ꭊ	ꭚ	�
B	ꬻ	ꭋ	꭛	ꞔ
C	ꬼ	ꭌ	ꭜ	ꭜ
D	ꬽ	ꭍ	ꭝ	ꞕ
E	ꬾ	ꭎ	ꭞ	ꭰ
F	ꬿ	ꭏ	ꭟ	ꭱ

	1DF0	1DF1	1DF2	1DF3	1DF4	1DF5	1DF6	1DF7	1DF8	1DF9	1DFA	1DFB	1DFC	1DFD	1DFE	1DFF
0	ﬀŋ	ᴋ	ʣ	ɢ	ᴀ	ɫ	ω	ȷ	Ⱥ				b	ɣ	ʤ	ɬ
1	ȹ	ɬ	ʥ	ɣ	a	ⱱ	ɾ	ȷ	Ɇ				c	ᴘ	ʥ	ɳ
2	ɔ	ʤ	ʦ	ħ	ə	ⱴ	ʃh	❍	ʍ				d	ᴂ	ʆ	ɭ
3	ⱪ	ʧ	ʨ	ɸ	ᴇ	ⱷ	th	❍	Ⱨ				f	đ	ʄ	ψ
4	ⱡ	ŋ	tθ	ɋ	ε	χ	ʃh	❍	ɟ				g	g	ʧ	ω
5	ʐ	ɟ	ᶁ	ᴚ	ᶃ	ỿ	wh	ᴐ	ɺ				"	ɬ	ʧ	ɖ
6	ⱥ	ʃ	ᶀ	ᶄ	ɦ	ʔ	ʑ	℧	ç			ζ	،	ɫ	ʒ	ɧ
7	ᶙ	ʧ	ᴔ	ᶜ	ɦ	ⱹ	ʑ	Ⱦ				ρ	˅	ɫ	ʒ	ŋ
8	ⱡ	ʒ	ⱱ	ʧ	ᴋ	ᴐ	Ⱥ	Ⱨ	ӿ			σ	،	ᵾ	ᴅ	ʂ
9	ꜰ	ʤ	ɾ	ɥ	k	ᴂ	ɑ	θ	ɤ			ε	⸝	ʄ	ᴀ	ᴣ
A	ǃ	ị	ⱦ	ʃ	ᴍ	a	ᴀ	ᴜ	ɚ			₁	⸜	ʊ	ᴇ	ƈ
B	ꝭ	ᵱ	ʣ	ᴃ	ᴔ	ch	ᴀ	ᴜ	ɬ			ᴃ	⸜	w	ɭ	ᴋ
C	ꝭ	ʧ	ʨ	θ	ᴔ	ɘ	Ɛ	ᴝ	ꜱ			ᴢ	⸞	ɥ	ɭ	ƀ
D	ꝯ	ꜿ	ɟ	ꭕ	N	ie	ᴇ	ᴝ				ᴢ	⸝	ɽ	ᴜ	ɋ
E	ꝱ	s	ʣ	ⱨ	ᴎ	oi	ɬ	ᴜ				ᴇ	ɔ	ⱷ	ᴜ	ꜰ
F	ƈ	ꝰ	ᶚ	ᴡ	ᴎ	ou	ȷ	ʊ				œ	⸝	ⱷ	ɋ	ꜰ

References

Atlases

AIS: Karl Jaberg & Jakob Jud (1928–1940) *Sprach- und Sachatlas Italiens und der Südschweiz*, 8 vol. Zofingen: Ringier. www3.pd.istc.cnr.it/navigais-web

ALAL: Potte, Jean-Claude (1975–1992) *Atlas linguistique et ethnographique de l'Auvergne et du Limousin*, 3 vol. Paris: CNRS.

ALCe: Dubuisson, Pierrette (1971–1982) *Atlas linguistique et ethnographique du Centre*, 3 vol. Paris: CNRS.

ALD-II: Hans Goebl (2012) *Atlant linguistich dl ladin dolomitich y di dialec vejins, 2ª pert*. Strasbourg: ÉLiPhi. http://ald.gwi.uni-muenchen.de/

ALEA: Manuel Alvar, Antonio Llorente & Gregorio Salvador (1961–1973) *Atlas lingüístico y etnográfico de Andalucía*, 6 vol. Granada: Universidad de Granada-CSIC.

ALEANR: Manuel Alvar, Tomás Buesa & Antonio Llorente (1979–1983) *Atlas lingüístico y etnográfico de Aragón, Navarra y Rioja*, 12 vol. Madrid: La Muralla.

AleCMAN: Pilar Garcia Mouton & Moreno Fernández. *Atlas Lingüístico y etnográfico de Castilla - La Mancha. University of Alcalá.* https://alecman.web.uah.es/

ALF: Jules Gilliéron & Edmond Edmont (1902–1910) *Atlas linguistique de la France*, 12 vol. Paris : Champion. http://lig-tdcge.imag.fr/cartodialect5

ALI: Ugo Pellis, Lorenzo Massobrio & Matteo Bartoli (1995–) *Atlante linguistico italiano*. Rome: Istituto Poligrafico e Zecca dello Stato (vol. 1–8); Turin: Istituto dell'Atlante Linguistico Italiano (vol. 9–10). http://atlantelinguistico.it/

ALJA: Jean-Baptiste Martin & Gaston Tuaillon (1971–1978) *Atlas linguistique et ethnographique du Jura et des Alpes du Nord*, 3 vol. Paris: CNRS.

ALLy: Pierre Gardette (1950–1976) *Atlas linguistique et ethnographique du Lyonnais*, 5 vol. Paris : CNRS. http://ucly.fr/recherche/institut-chaire-projets-recherche/institut-pierre-gardette/ally-en-ligne

ALN: Patrice Brasseur (1980–2019) *Atlas linguistique et ethnographique normand*, 5 vol. Paris: CNRS (vol. 1–3) / Caen: Presses universitaires de Caen (vol. 4) / Strasbourg: ÉLiPhi (vol. 5).

ALO: Geneviève Massignon & Brigitte Horiot (1971–1983) *Atlas linguistique et ethnographique de l'Ouest*, 3 vol. Paris: CNRS.

ALPI: Tomás Navarro Tomás *et al.* (1962) *Atlas lingüístico de la la Península Ibérica*, vol. 1 Fonética. Madrid: CSIC. http://alpi.csic.es/es/publicaciones

ALPic: Fernand Carton & Maurice Lebègue (1989–1997) *Atlas linguistique et ethnographique picard*, 2 vol. Paris: CNRS. http://anr-appi.univ-lille.fr/index.php/alpic

ALRé: Michel Carayol, Robert Chaudenson & Christian Barat (1984–1995) *Atlas linguistique et ethnographique de la Réunion*, 3 vol. Paris: CNRS.

CLEx: González Salgado & José Antonio. *Cartografía lingüística de Extremadura*. https://geolectos.com/

SBS: Werner König et al. (1997–2009) *Sprachatlas von Bayerisch Schwaben*, 14 vol. Heidelberg.

TPh: Louis Gauchat, Jules Jeanjaquet & Ernst Tappolet (1925) *Tableaux phonétiques des patois suisses romands*. Neuchâtel: Attinger. https://archive.org/details/gauchat-et-al-1925-tppsr/mode/2up

VIVALDI: Dieter Kattenbusch et al. (1998–2016) *Vivaio Acustico delle Lingue e Dialetti d'Italia*. Berlin, Humboldt-Universität, Institut für Romanistik. http://www2.hu-berlin.de/vivaldi

Digitization projects

Massimo Cerruti, Lorenzo Ferrarotti, Stefano Fiori, Matteo Rivoira & Barbara Turchetta (2025) Il progetto DigitALI: problemi e prospettive della digitalizzazione dell'Atlante Linguistico Italiano (ALI). *Linguistik Online* 137 (5): 37–67.

Claudine Fréchet, Fabio Armand & Michel Bert (2022) Mise à disposition en ligne de l'*Atlas linguistique du Lyonnais*: avancement, contraintes et perspectives. *Géolinguistique* 22.

Pilar García Mouton (coord.), Inés Fernández-Ordóñez, David Heap, & Maria Pilar Perea, João Saramago & Xulio Sousa (2016) ALPI-CSIC. http://www.alpi.csic.es/

Hans-Joachim Geisler, Robert Forkel & Johann-Mattis List (2021) A digital, retro-standardized edition of the Tableaux phonétiques des patois suisses romands (TPPSR). In André, Avanzi, Lo Vecchio & Millour (eds.) *Nouveaux regards sur la variation dialectale*, 13–36. Strasbourg: ÉliPhi. https://tppsr.clld.org/

Michele Loporcaro, & Stephan Schmid, & Chiara Zanini, & Diego Pescarini, & Giulia Donzelli, & Stefano Negrinelli & Graziano Tisato (2021) *AIS, reloaded: A digital dialect atlas of Italy and southern Switzerland.* In André, Avanzi, Lo Vecchio & Millour (eds.) *Nouveaux regards sur la variation dialectale*, 111–136. Strasbourg: ÉliPhi. https://www.ais-reloaded.uzh.ch/

Other

Vytautas Ambrazas, ed. (2nd edition, 2006) *Lithuanian Grammar.* Institute of the Lithuanian Language.

Boas, Goddard, Sapir & Kroeber (1916) *Phonetic Transcription of Indian Languages: Report of Committee of American Anthropological Association* (Smithsonian Miscellaneous Collections 66.6)

Frasz Boas (1892) Notes on the Chemakum Language. *American Anthropologist* 5(1): 37-44.

Guylaine Brun-Trigaud (2016) Vers un renouveau des atlas linguistiques régionaux? *Géolinguistique* 16. https://doi.org/10.4000/geolinguistique.444

Kalnyn' (1973): Л.Э. Калнынь, *Опыт моделирования системы украинского диалектного языка.* Наука.

Kalnyn' & Maslennikova (1981): Л.Э. Калнынь & Л.И. Масленникова, *Сопоставительная модель фонологической системы славянских диалектов.* Наука, Moscow.

Patricia Keating (2018) Some challenges for the IPA from languages of China. *13th Phonetic Conference of China*, Jinan University, Guangzhou.

Uli Kozok (2015) *A 14th Century Malay Code of Laws.* ISEAS, Singapore.

Andrea de Leeuw van Weenen (2000) *A Grammar of Möðruvallabók.* Leiden University.

Ling Feng (2007) The articulatory and acoustic study of fricative vowels in Suzhou Chinese. *Proceedings of ICPhS XVI*, International Congress of Phonetic Sciences.

MENOTA: AM 677 4to – De xii abusivis saeculi, Prosper's Epigrams, Gregory's Homilies on the Gospels, Gregory's Dialogues v. 1.0.

Edwin Pulleyblank (1962) The Consonantal System of Old Chinese. *Asia Major* 9: 2.

—— (1984) *Middle Chinese: A Study in Historical Phonology.* UBC Press.

George Trager (1972) *Language and languages.* Chandler Publishing Company.

Bengisu Winnick (1972) *The phonetics and phonology of Istanbul Turkish.*

Figures[18]

Figure 2. The curled tilde (lazy *s*), with regular tilde from the same sources for comparison. Sources of lazy *s*: ① ALJA map 73 point 82, ② ALLy map 56 point 56, ③ ALPic map 82 point 15 (closed variant), ④ TPh column 177 point 6, ⑤ AIS map 852 point 466, ⑥ ALN map 1458 point 2 (reversed variant, found in vol. 5), ⑦ ALN map 823 point 1, ⑧ ALAL map 9 point 30, ⑨ ALF map 9 point 694.

Figure 3. Combining ̇ + ̇. ① AIS map 889 point 544, ② ALD-II map 10 point 81, ③ VIVALDI «l'angelo» (Ceppomorelli).
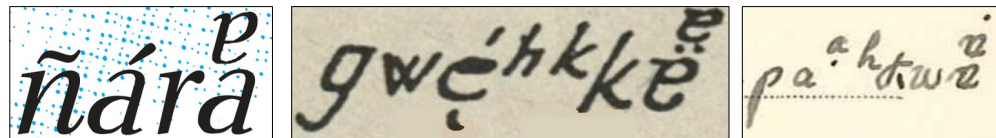
Figure 4. Combining ɐ. ① ALD-II map 17 point 37, ② ALEA map 4 point Gr 201, ③ ALPI questionnaire 236 point 544.

---

[18]For some characters we cite supporting non-atlas examples.
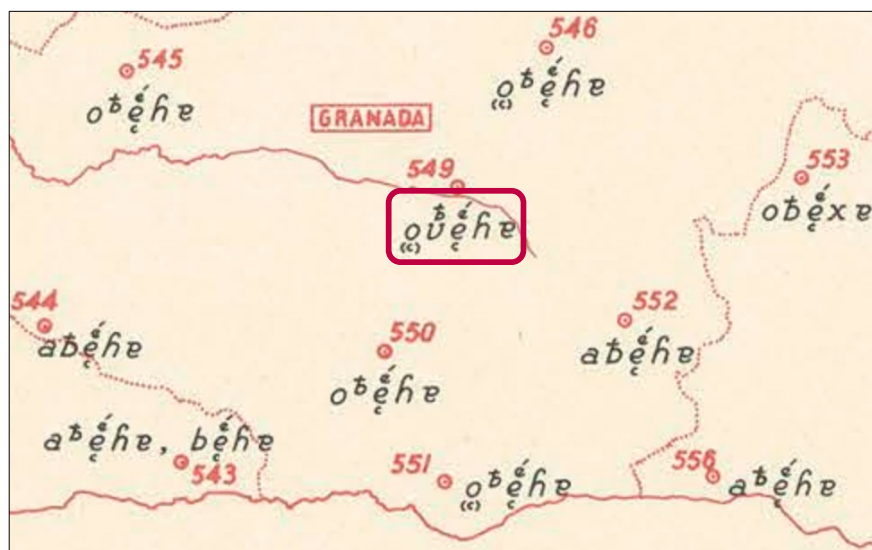
Figure 5. Combining ƀ. ALPI map 20 point 549. Note that combining ƀ contrasts with modifier ƀ, which is currently in the Unicode pipeline.
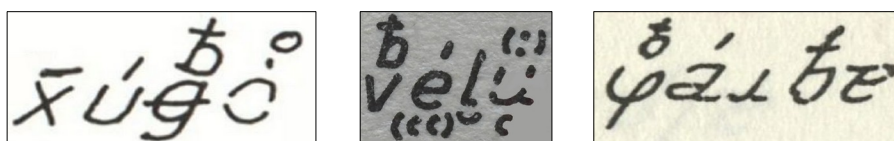
Figure 6. Combining ƀ. ① ALEANR map 329 point Hu 107, ② ALPI map 2 point 245, ③ ALEA map 616 point H 402.
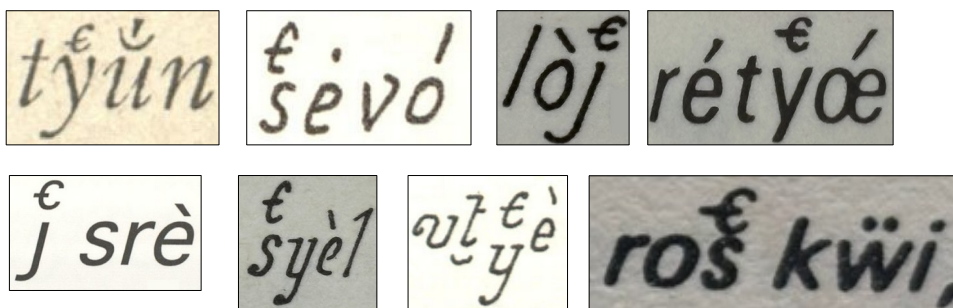
Figure 7. Combining ɛ. ① ALF map 9 point 791, ② ALLy map 311 point 15, ③ ALO map 122 point 66, ④ ALPic map 57 point 51, ⑤ ALN map 1460 point 33, ⑥ ALAL map 1 point 21, ⑦ ALJA 1590 point 5, ⑧ ALRé map 7 point 11.
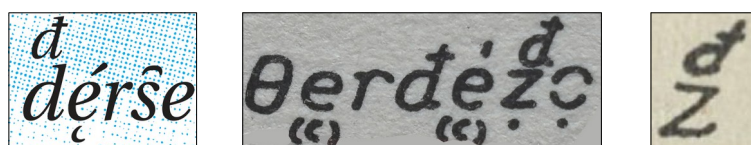
Figure 8. Combining đ. ① ALD-II map 60 point 93, ② ALPI map 2 point 141, ③ ALEA vol. 1, *transcripción fonética*, §1.

16
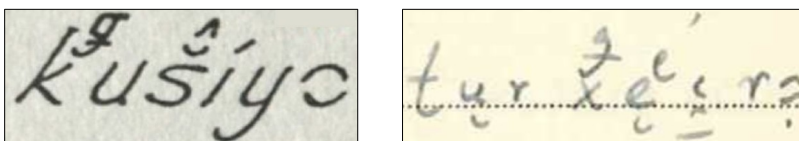
Figure 9. Combining *g.* ① ALEA map 562 point Al 405, ② ALPI questionnaire 479a point 150. It is a bit difficult to distinguish from a spacing diacritic in (1), but a comparison with similar symbol combinations on the same map, such as ⟨ǧ⟩ at http://imgur.com/a/HyJAbJa, clarifies the situation.
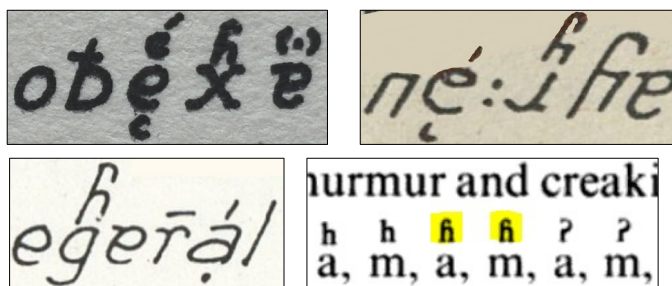
Figure 10. Combining *ɦ.* ① ALPI map 20 point 452, ② ALEA map 4 point Ma 405, ③ ALEANR map 336 point Z 502, ④ Pulleyblank (1984: 34).
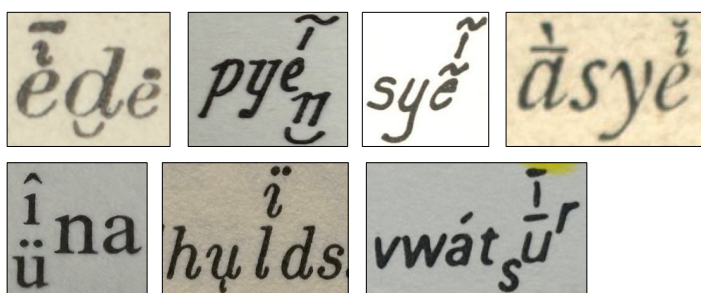
Figure 11. Combining dotless *i* with macron, breve, tilde, circumflex and trema. ① TPh column 236 line 11, ② ALAL map 273 point 26, ③ ALCe map 1438 point 13, ④ ALF map 8 point 156, ⑤ ALI map 45 point 51, ⑥ SBS 7.2: 388, ⑦ ALRé map 249 point 35. This character is only needed when a diacritic suppresses the dot of a combining *i*, U+0365. On inquiry to SEW about the possibility of simply using U+0365, we got the response, *I don't think U+0365 can become soft-dotted; consider proposing a combining dotless i.* The alternative would be multiple atomic characters for combining *i* + diacritic. See also Figure 62.
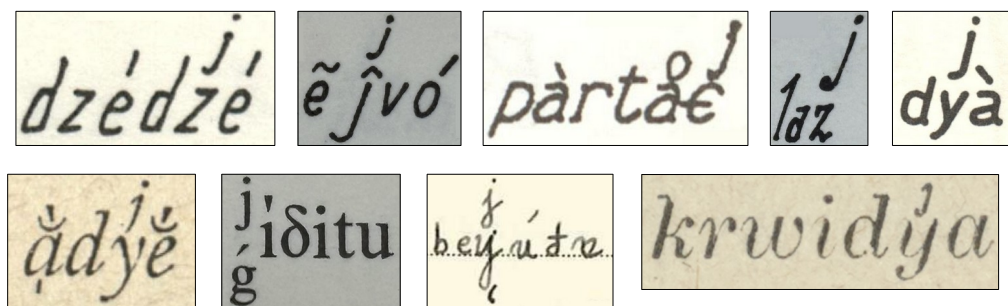
Figure 12. Combining *j*. ① ALLy map 496 point 17, ② ALO map 576 point 117, ③ ALN map 784 point 26, ④ ALAL map VI point 27, ⑤ ALCe map 516 zone B, ⑥ ALF map 9 point 895, ⑦ ALI map 47 point 945, ⑧ ALPI questionnaire 140 point 352, ⑨ TPh column 54 point 29 [combining *j* is always without tittle in TPh].

Figure 13. Combining *ł*. ALD-II map 7 point 211.

Figure 14. Combining *ł*. ALPI map 38 point 245.
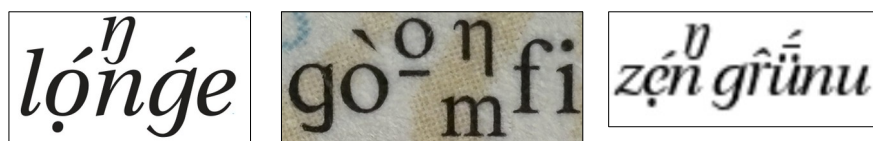
Figure 15. Combining *ŋ*. ① ALD-II map 60 point 211, ② ALI map 153 point 445, ③ VIVALDI «i prati sono verdi» (Issime).

18
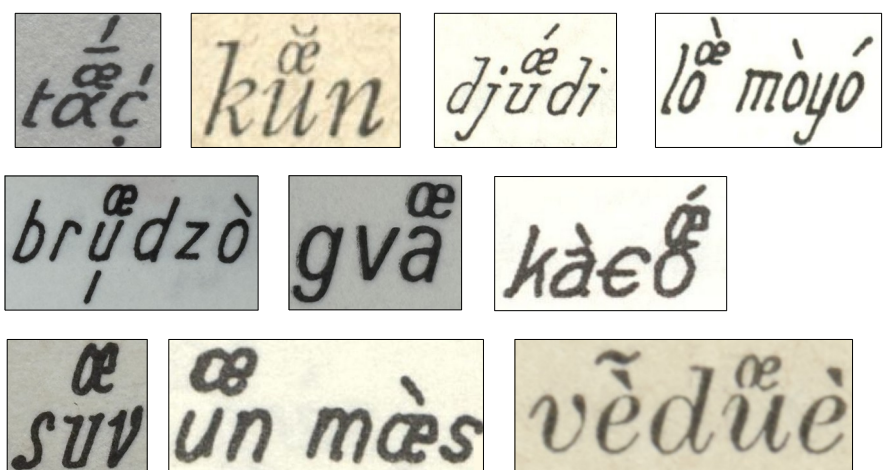
Figure 16. Combining æ. ① AIS map 864 point 149, ② ALF map 9 point 776, ③ ALJA map 92 point 13, ④ ALLy map 176 point 49, ⑤ ALO map 383 point 94, ⑥ ALPic map 158 point 9, ⑦ ALN map 782 point 5, ⑧ ALAL map 2 point 10, ⑨ ALCe map 65 point 11, ⑩ TPh column 179 line 20, notes.
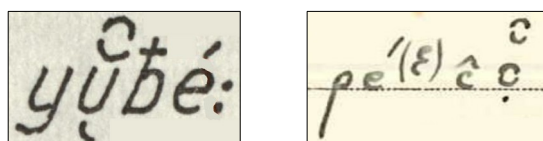
Figure 17. Combining ɔ. ① ALEA map 1838 point H 402 , ② ALPI questionnaire 22 point 544.
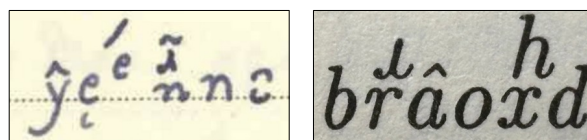
Figure 18. Combining ɹ. ① ALPI questionnaire 134 point 486, ② SBS 7.2: 169.
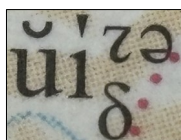
Figure 19. Combining ɾ. ALI map 94 point 858.
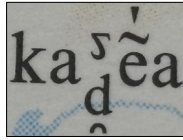
Figure 20. Combining reversed ɿ. ALI map 1 point 747. The reflection marks a difference in voicing.
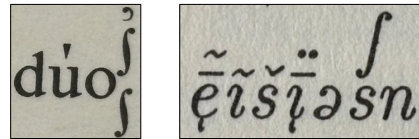
Figure 21. Combining ʃ. ① ALI map 2 point 740, ② SBS 7.2: 68.
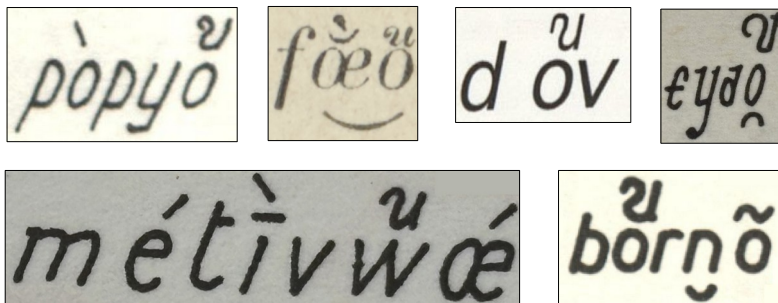
Figure 22. Combining ʋ. ① ALLy map 383 point 53, ② TPh column 50 point 13, ③ ALN map 1516 point 2, ④ ALAL map 1 point 61, ⑤ ALO map 61 point 62, ⑥ ALCe map 532 point 48.
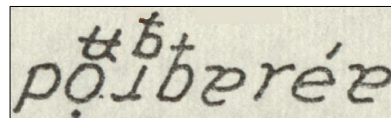
Figure 23. Combining ʮ. ALEA map 1588 point J 404. The original split ʮ is replaced by barred ʉ in both the ALEA and the ALEANR, which are handwritten, presumably because it's difficult to handwrite a split ʮ in a way that is clearly distinct from ü. All traditional linguistic atlases are based on quickly handwritten field notes, made by the linguist interviewing a dialect speaker, so this was a significant issue. In print, however, split ʮ is used.
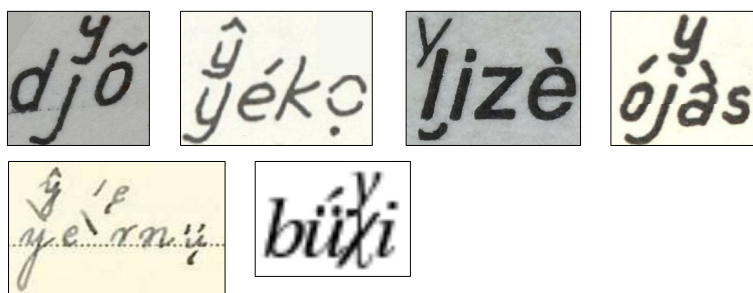
Figure 24. Combining *y*. ① ALO map 354 point 90, ② ALEANR map 11 point Lo 605, ③ ALAL map 74 point 9 (enquête D), ④ ALCe map 96 point 51, ⑤ ALPI questionnaire 134 point 339, ⑥ VIVALDI «ho letto questo libro» (Issime). Note that the *y* suppresses the dot of the supporting *j* in 1 but not in 4; other combining letters behave the same.
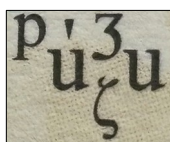
Figure 25. Combining ʒ. ALI map 18 point 577.
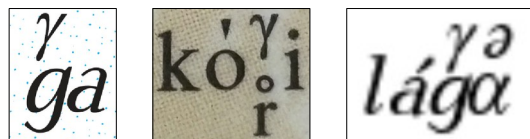
Figure 26. Combining γ. ① ALD-II map 24 point 185, ② ALI map II point 943, ③ VIVALDI «l'acqua» (Padola).
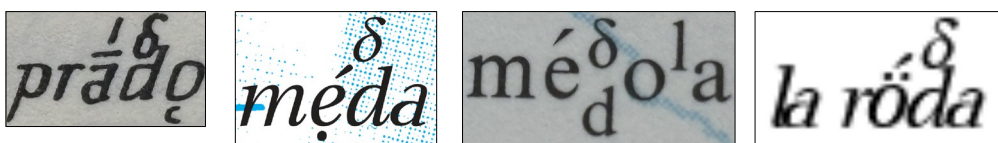
Figure 27. Combining δ. ① AIS map 1415 point 511, ② ALD-II map 5 point 145, ③ ALI map 3 point 253, ④ VIVALDI «la ruota» (Isola Sant'Antonio).
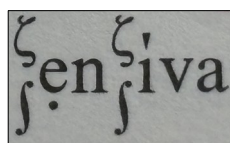
Figure 28. Combining ζ. ALI map 32 point 354.
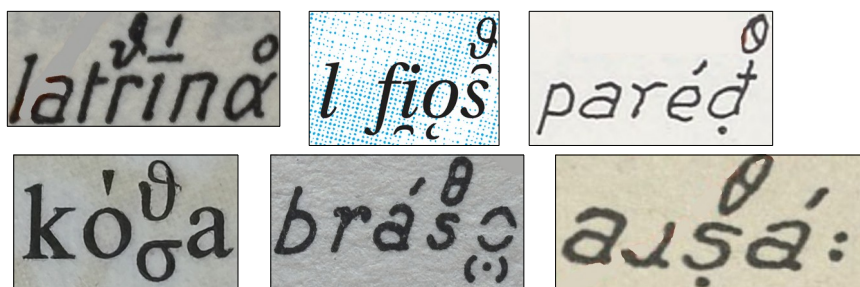
Figure 29. Combining θ. ① AIS map 871 point 222, ② ALD-II map 8 point 109, ③ ALEANR map 97 point 305, ④ ALI map 9 point 394, ⑤ ALPI map 27 point 569, ⑥ ALEA map 18 point Se 602. This should be identified as a normal theta θ rather than as the theta symbol ϑ.
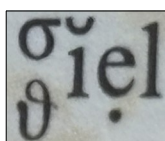
Figure 30. Combining σ. ALI map 33 point 349.
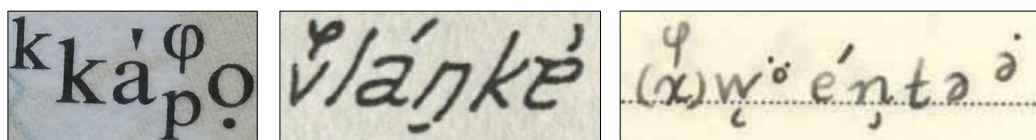
Figure 31. Combining φ. ① ALI map 8 point 524, ② ALEA map 315 point Al 403, ③ ALPI questionnaire 92 point 559.
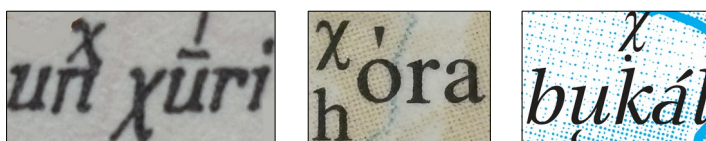
Figure 32. Combining χ. ① AIS map 1357 point 818, ② ALI map II point 969, ③ ALD-II map 90 point 76. The dot-below in 3 might be set above the *k*. However, this won't work for all situations, like combining ⟨ę⟩ which will need a way to tell the font that the diacritic belongs on the combining letter.
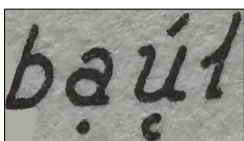
Figure 33. L with half stroke, ⱡ. ALPI map 24 point 615.
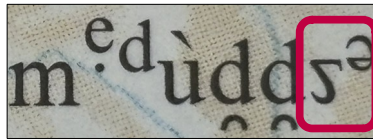
Figure 34. Reversed ɿ. ALI map 3 point 923.
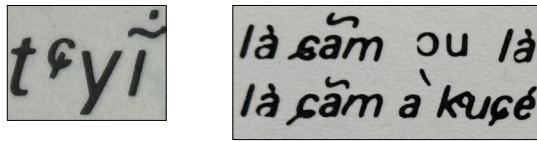
Figure 35. Modifier ɛ. ① ALPic map 188 point 20. The stroke is set at an angle; at right are several tokens from ALRé map 252, caption, of baseline ɛ showing similar variation in bar positioning.
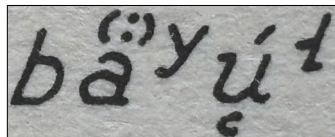
Figure 36. Modifier ł. ALPI map 24 point 135.
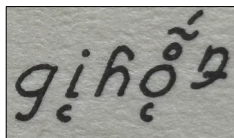
Figure 37. Modifier ŋ. ALPI map 11 point 519.
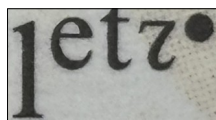
Figure 38. Modifier ɿ. ALI map 1 point 943.

Figure 39. Modifier reversed ɿ. ALI map 1 point 46.
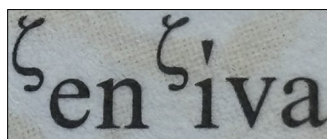
Figure 40. Modifier ʒ. ALI map 32 point 349.
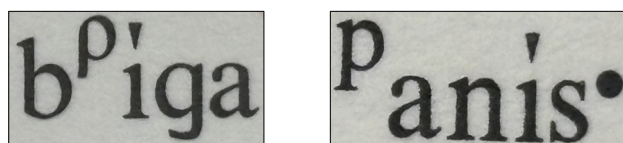
Figure 41. Modifier *ρ*. ALI map II point 94, with modifier *p* from elsewhere on the same map for comparison.
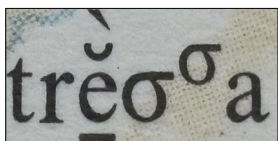
Figure 42. Modifier *σ*. ALI map 14 point 446.
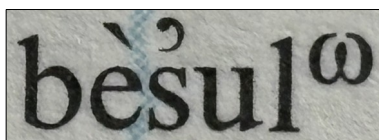
Figure 43. Modifier *ω*. ALI map 36 point 142. This is in the pipeline at U+1DFF4, but is not yet scheduled for Unicode 18.
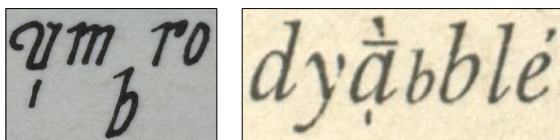
Figure 44. Subscript *b*. ① ALAL map 10 point 20, ② ALF map 403 point 699. The first generation of atlases set 'subscript' letters in a smaller type size, as at right, but the modern convention is to use true subscripts, as at left, and that is how the data is digitized.

1 ma:li˜ bu:bu , bubu ditu(mᵇ)unₙi #³¹ pa
2 di sipanuḥ ña , jaka tida tarisi #³²
3 lima: maş(d)aⁿ(d)uña // bara˜ ma:#uba#
4 pañ(c)uwida , didaⁿ(d)u lima ta:hil⌐
5 sapaha: // bara˜ bahila˜ °ura˜ ma:ta
6 kaʳja ya˜ puʳwa , sa:kati lima daⁿ(d)u

3 tap(d)in(d)i˜ lanᵢay ra: ŋo, lima
   maş(d)an(d)a
4 ña // puṇara:pi jaka bahu:ta˜ mas⌐
5 pirak⌐ riti ra(ⁿc)u˜ ka˜śa ta(mᵇ)uga , si
6 la:maña batiga puhu:n⌐ //⁵³ siⁿgaṇ⌐

Figure 45. Subscript *b*, *c* and *d*. Kozok (2015: 69, 71). Conjuncts transliterated as subscripts. They commonly occur after nasals, but note ⟨s_d, p_d⟩.

Figure 46. Subscript *b*. Kalnyn' & Maslennikova (1981: 396). Equivalent Cyrillic and Latin transcriptions. Cyrillic is already supported. Figure 59 shows that the superscript minus sign does not need to appear directly over the subscript, and that this can be digitized ⟨п₆⁻'/p_b⁻'⟩.

Figure 47. Subscript *ɛ*. ① ALLy map 914 point 31, ② ALN map 823 point 3, ③ ALAL map 3 point 53, ④ ALCe map 1344 point 30.
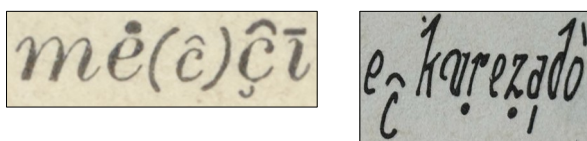
Figure 48. Subscript *c* with turned breve. ① TPh column 381 line 13, ② ALAL map 42 point 27. Other tokens have a cedilla, as next. As is the case with superscript ⟨ˢ⟩, a subscript ⟨ᵪ⟩ should be rendered as subscript *c* plus a combining cedilla.

Figure 49. Subscript *c* with cedilla. Boas (1892: 37 fn, 38, 39, 41).

Figure 50. Subscript *c*. Kalnyn' & Maslennikova (1981: 360). The alignment of the apostrophe needs to be handled by the font; U+0315 should work.

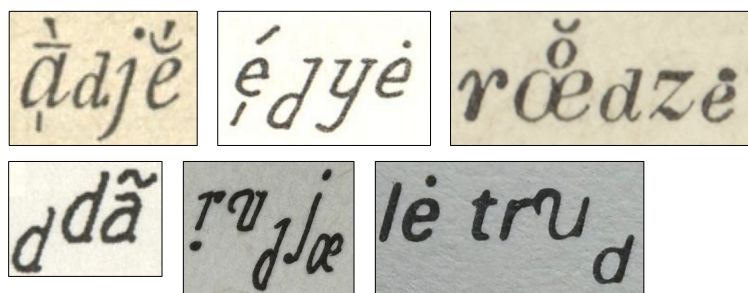Figure 51. Subscript *c* with cedilla. Winnick (1972: 163).

Figure 52. Subscript *d.* ① ALF map 9 point 794, ② ALJA map 135 point 41, ③ TPh column 248 line 17, ④ ALN map 783 point 21, ⑤ ALAL map 8 point 33, ⑥ ALRé map 8 point 37.

Figure 53. Subscript *d.* Pulleyblank (1962: 215). Later on the page Pulleyblank says that all three final plosives appear to have been implosive. ('Implosive' here seems to mean unreleased.) In this notation the three finals would be ⟨-ᵦ, -d, -g⟩.

Figure 54. Subscript *d.* Kalnyn' (1973: 327). Illustrations specific to Ukrainian are presented in Cyrillic, ⟨т⟩.
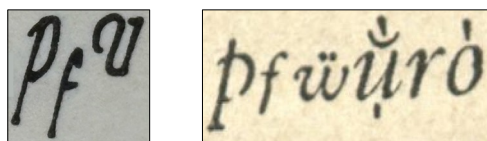
Figure 55. Subscript *f.* ① ALAL map 317 point 11, ② ALF map 1683 point 805.

Figure 56. Subscript *f*. Winnick (1972: 163).

Figure 57. Subscript *f*. Kalnyn' & Maslennikova (1981: 396). Equivalent Cyrillic and Latin transcriptions. Figure 59 shows that the superscript minus sign does not need to appear directly over the subscript, and that this example can be digitized ⟨ʙ_ɸ⁻'/w_f'⟩.
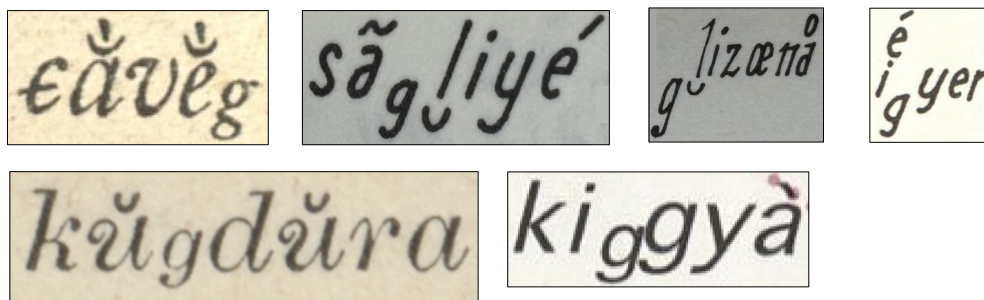

Figure 58. Subscript *g*. ① ALF map 1200 point 988, ③ ALO map 396 point 38, ④ ALAL map 74 point 17, ⑤ ALCe map 230 point 30, ⑥ TPh column 145 line 31, ⑦ ALN map 1410 point 30. This is a regular *g* set in italics, not a script *g*.

Figure 59. Subscript *g*. Kalnyn' & Maslennikova (1981: 350, 372). The second example shows (a) that there is no distinction between print ⟨g⟩ and script ⟨g⟩, and (b) that the C' superscript notation can be shifted to the right of the C_g subscript notation, and does not need to appear directly above it – that would appear to be merely a space-saving measure.
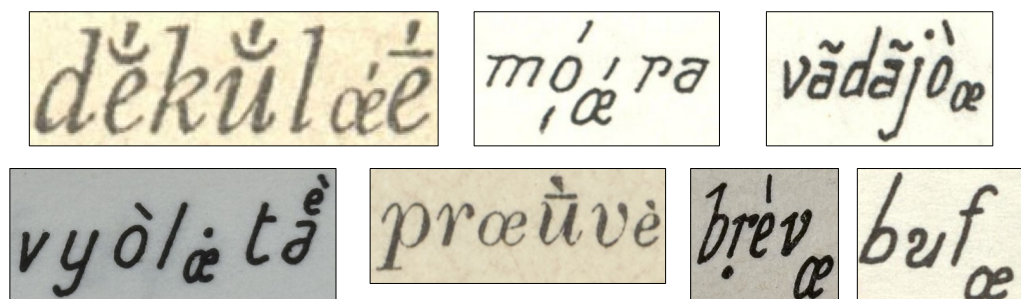
Figure 60. Subscript *æ*. ① ALF map 1138 point 275, ② ALJA map 440 point 49, ③ ALLy map 205 point 2, ④ALO map 362 point 84, ⑤ TPh column 318 line 5, ⑥ ALAL map 3 point 74, ⑦ ALCe map 18 point 65.
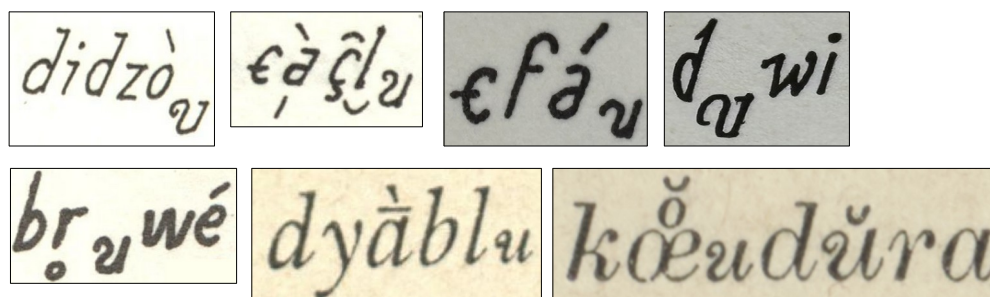
Figure 61. Subscript *u*. ① ALJA map 92 point 81, ② ALLy map 216 point 8, ③ ALO map 576 point 58, ④ ALAL map 118 point 70, ⑤ ALCe map 13 point 42, ⑥ ALF map 403 point 691, ⑦ TPh column 144 point 25.

Figure 62. ① MENOTA p. 32v, ② Leeuw van Weenen (2000: 68). Combining dotless *i* without a diacritic. Although the lack of a tittle is not semantically distinct, the proposed combining letter will be useful for diplomatic transcription.

Figure 63. The baseline letters *d* and *g* with a half stroke. ALPI, *introducción*. These and the analogous *b* are not requested for now, pending further attestation.

<table>
<tr><td colspan="2">

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646[19]
Please fill all the sections A, B and C below.
Please read Principles and Procedures Document (P & P) from std.dkuug.dk/JTC1/SC2/WG2/docs/principles.html for guidelines and details before filling this form.
Please ensure you are using the latest Form from std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html.
See also std.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html for latest *Roadmaps*.

</td></tr>
</table>

A. Administrative

1. **Title:**	*Romance dialectology symbols*
2. Requester's name:	*Nicolas Datua and Kirk Miller*
3. Requester type (Member body/Liaison/Individual contribution):	*individual*
4. Submission date:	*2025 October 09*
5. Requester's reference (if applicable):	
6. Choose one of the following:	
This is a complete proposal:	*yes*
(or) More information will be provided later:	

B. Technical – General

1. Choose one of the following:	
a. This proposal is for a new script (set of characters):	*no*
Proposed name of script:	
b. The proposal is for addition of character(s) to an existing block:	*yes*
Name of the existing block:	*Combining Diacritical Marks Extended, Latin Extended-E, Latin Extended-G*
2. Number of characters in proposal:	*48*

3. Proposed category (select one from below - see section 2.2 of P&P document):

A-Contemporary   **x**   B.1-Specialized (small collection)   B.2-Specialized (large collection)
C-Major extinct   D-Attested extinct   E-Minor extinct
F-Archaic Hieroglyphic or Ideographic   G-Obscure or questionable usage symbols

4. Is a repertoire including character names provided?	*yes*
a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?	*yes*
b. Are the character shapes attached in a legible form suitable for review?	*yes*

5. Fonts related:
    a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?
        Kirk Miller
    b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):
        SIL (Gentium Release)

6. References:	
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?	*yes*
b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?	*yes*
7. Special encoding issues:	
Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)?	*no*

8. Additional Information:

Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at www.unicode.org for such information on other scripts. Also see Unicode Character Database (www.unicode.org/reports/tr44/) and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before? _ _ _*yes*_ _ _

    If YES explain _ _ _ *Requests for subscripts c–g, with less evidence, were made in L2/21-043 and 21-207R.* _ _ _

2. Has contact been made to members of the user community (for example: National Body,

    user groups of the script or characters, other experts, etc.)? *yes*

        If YES, with whom? _ _ _ _ _ *Authors are members of the user community.* _ _

        If YES, available relevant documents: _ _ _ _ _ _ _ _ _ _ _ _ _ _ _

3. Information on the user community for the proposed characters (for example:

    size, demographics, information technology use, or publishing use) is included? _ _ _ _ _ _

    Reference: _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _

4. The context of use for the proposed characters (type of use; common or rare) *transcription*

    Reference: _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _

5. Are the proposed characters in current use by the user community? _ _ _*yes*_ _ _

    If YES, where? Reference: _ _ _ _ _ _ _ _ _ _ _ _ _ _ _

6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely

    in the BMP? *no*

        If YES, is a rationale provided? _ _ _ _ _ _

           If YES, reference: _ _ _ _ _ _ _ _ _ _ _ _

7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)? _ _ *no*

8. Can any of the proposed characters be considered a presentation form of an existing

    character or character sequence? *no*

        If YES, is a rationale for its inclusion provided? _ _ _ _ _ _

           If YES, reference: _ _ _ _ _ _ _ _ _ _ _ _

9. Can any of the proposed characters be encoded using a composed character sequence of either

    existing characters or other proposed characters? *no*

        If YES, is a rationale for its inclusion provided? _ _ _ _ _ _

           If YES, reference: _ _ _ _ _ _

10. Can any of the proposed character(s) be considered to be similar (in appearance or function)

    to**,** or could be confused with, an existing character? *no*

        If YES, is a rationale for its inclusion provided? _ _ _ _ _ _

           If YES, reference: _ _ _ _ _ _ _ _ _ _ _ _

11. Does the proposal include use of combining characters and/or use of composite sequences? _ _ *no* _ _

    If YES, is a rationale for such use provided? _ _ _ _ _ _

        If YES, reference: _ _ _ _ _ _ _ _ _ _ _ _

    Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? _ *no*

        If YES, reference: _ _ _ _ _ _ _ _ _ _ _ _

12. Does the proposal contain characters with any special properties such as

    control function or similar semantics? *no*

        If YES, describe in detail (include attachment if necessary) _ _ _ _ _ _

_ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _

_ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _

13. Does the proposal contain any Ideographic compatibility characters? _ _ *no* _ _

    If YES, are the equivalent corresponding unified ideographic characters identified? _ _ _ _ _ _

        If YES, reference: _ _ _ _ _ _ _ _ _ _ _ _