L2/25-251R

Unicode request for Romance dialectology symbols

Nicolas Datua, nicolasdatua@gmail.com
Kirk Miller, kirkmiller, gmail.com

2025 October 26

The present proposal covers four related systems of phonetic notation used in nearly all French, Italian, and Spanish linguistic atlases[1] of the 20th century, as well as in a number of more recent and ongoing publications. The vast corpus of dialect data recorded in these atlases remains, on the whole, under-utilized by researchers for lack of accessibility. Numerous initiatives[2] 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,[3] 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. It owes much to L2/14-169 *Proposal to encode additional dialectology Latin characters* by Denis Jacquerye, whose bibliography has saved us a great deal of time. In addition, some of our attestations are reproductions of his.

We thank Giacomo Ferrieri, Giorgio Nagy and Graziano Tisato for their feedback on various aspects of this proposal. We further thank Giorgio Nagy for placing at our disposal the data that he had used in his dialectometric study (Nagy 2024), which facilitated our search for attestations of certain characters.

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

[3] 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 combination of symbols found in the atlases, and different projects may arrive at different approximations for what are (in reality) the same sounds, or vice-versa. In any case it is helpful to archive a faithful representation of the original transcription 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.

We now discuss the four related notational systems in this family.

1. Rousselot-Gilliéron (RG)

The RG system was devised by Jean-Pierre Rousselot in the late 19th century and used by Jules Gilliéron for the monumental *Atlas linguistique de la France* (ALF), the world's first linguistic atlas in cartographic form. Thanks to this, the RG system became ubiquitous in French dialectology, adopted for instance by Charles Bruneau for his *Enquête linguistique sur les patois d'Ardenne* (ELPA), by the authors of the *Tableaux phonétiques des patois suisses romands* (TPh),[4] and by the Centre national de la recherche scientifique (CNRS) for a series of atlases dedicated to specific regions, of which we cite eight: the ALAL, ALCe, ALJA, ALLy, ALN, ALO, ALPic and ALRé.

The base letters of the RG system are as follows.[5] On the second line we provide approximations, in the International Phonetic Alphabet (IPA), of the sounds that these letters indicate in isolation.[6]

RG:	a	b	c	ϵ	d	e	f	g	h	i	j	k	l	m	n	o	œ	p	r	s	t	u	ʊ	v	w	y	z
IPA:[7]	ä	b	-[8]	ʃ	d	ę	f	g	h	i	ʒ	k	l	m	n	o̜	ø̜	p	r	s	t	y	u	v	w	j	z

Many of these base letters, especially the vowels, may be given diacritics to denote other phonetic values. Nearly all of the diacritics are already part of the Unicode standard.[9] The one 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 perceptually between [a] and [e]. Similarly, ⟨s̆⟩ (IPA: [ʂ][10]) denotes a consonant with a degree of voicing between that of [s] and [z]. In principle any RG base letter can also be used as a combining letter, and indeed all of them are attested as such. The base letters ⟨ϵ j œ ʊ y⟩ lack combining equivalents in Unicode, so we propose them for inclusion (cf. figures 7, 12, 16, 21, 23).

[4]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 and scientific value are the same. We therefore include the TPh among the atlases of this proposal.

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

[6]Exceptionally, the TPh have an additional base character ⟨ł⟩ (IPA: [ʟ]).

[7]We base this analysis on Gilliéron & Rousselot 1887: 1–17, Gilliéron & Edmont 1902: 19, and on the descriptions provided in the introduction to the other cited works. The IPA approximations that we provide here, and in the rest of this proposal, are often narrower than advisable for practical use. This high degree of precision is intended to illustrate certain details that would otherwise be obscured. For instance, we provide the IPA approximation of [ę] for the RG ⟨e⟩ to indicate that the latter represents, specifically, a mid front unrounded vowel, distinguished from the high-mid ⟨é⟩ (IPA: [e]). In practice, of course, in the IPA the mid vowel is often not distinguished from its high-mid and low-mid counterparts.

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

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

[10]This IPA equivalence, like many in this proposal, is only one of multiple possibilities. RG ⟨s̆⟩ could also be approximated as IPA ⟨z̝⟩. Here it bears mentioning that the superpositions of Romance dialectological notation often imply a value closer to the base letter than to the combining one, hence our choice of the IPA ⟨ʂ⟩.

Another feature of this system is the use of subscript letters[11] to denote weakly articulated sounds, as in ⟨ᵢ⟩ (extIPA: [i̤]). The base letters ⟨b c ɕ d f g œ u⟩ lack subscript equivalents in Unicode, so we propose them for inclusion (cf. figures 43–53).

Some of the later CNRS atlases make use of spacing superscripts, either instead of subscripts altogether (ALPic), or alongside them for more specialized functions (ALJA, ALRé[12]). The base letter ⟨ɕ⟩ lacks a superscript equivalent in Unicode, so we propose it for inclusion (Figure 34).

Of the atlases cited in this proposal that 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 1990s. Many volumes of other atlases await publication.[13]

2. Böhmer-Ascoli (BA)

The BA system was devised in the late 19th century by Eduard Böhmer, who took the phonetic notation of Graziadio Isaia Ascoli as his starting point. It was notably used for the *Atlante italo-svizzero* (AIS),[14] where it appears with the following base letters:

BA:	a	b	c	d	e	ə	f	g	h	i	k	l	ɬ	ɬ̈	ł	m	ɱ	n	ŋ	o	œ	p	r	s	t	u	v	w	y	z	α	β	γ	δ	θ	φ	χ	ʕ
IPA:[15]	ä	b	-[16]	d	ẹ	ə	f	g	h	i	k	l	ʎ	ʲ	ɭ	ʔ	m	m͡ɱ	n	ŋ	o̤	ø̤	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).

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 ⟨ɬ⟩
- Phonetic repurposing of ⟨ɱ⟩ (to IPA [ɱ])
- Addition of ⟨đ ø⟩ (IPA: [d̪͡ð], [ø̤])

ALD only:

- Addition of ⟨ɐ⟩ (IPA: [ɐ])[17]

VIVALDI only:

- Addition of ⟨ɹ ʁ ɬ χ⟩ (IPA: [ɹ], [ʁ], [t̪͡θ], [χ])

[11]The early RG notation, as it appears in the ALF, ELPA and TPh, made use of small type positioned on the baseline, and not subscripts in the modern sense. For the later CNRS series, it was decided to move the small letters below the line, so as to increase their visual distinction from the base letters. In light of this equivalence, we treat the early small letters as subscripts and do not suggest encoding them as distinct Unicode characters.
[12]The ALJA uses ⟨ʰ⟩ to denote what appears to be a glottal stop, and ALRé uses ⟨ʳ⟩ to denote vocalized outcomes of an original rhotic.
[13]Brun-Trigaud 2016 provides a useful overview of these.
[14]Also known by its original title, *Sprach- und Sachatlas Italiens und der Südschweiz*.
[15]We base this analysis on Jaberg & Jud 1928: 24–36.
[16]⟨c⟩ never occurs in bare form, only in combinations such as ⟨ć⟩ (IPA: [t͡ʃ]).
[17]Thus occupying the position held in the AIS and VIVALDI by ⟨α⟩, which is assigned a posterior position in the ALD (IPA: [ɐ̟]).

In all of these atlases, weak articulation is indicated with spacing superscript letters, while combining letters are used for intermediate sounds. Theoretically, any base letter might occur in superscript and combining forms. For now, we list only those that we have encountered.

Another new atlas that uses this notation is the *Atlante Multimediale dei Dialetti Veneti* (AMDV).

In all, we propose the following additions for this notation:

- Combining diacritics: ◌᷃ ◌̋ (figs. 2, 3)
- Combining letters: ɐ đ ɫ ŋ œ ɹ y γ δ θ φ χ (figs. 4, 8, 13, 15, 16, 18, 23, 25, 26, 28, 30, 31)

Both the ALD and VIVALDI atlases have had publications in the last fifteen years.

3. Pellis-Genre

The Pellis-Genre system, as we provisionally name it, was devised in the early 20th century by Ugo Pellis, the main field-worker of the *Atlante Linguistico Italiano* (ALI), and modified in the 1970s by Arturo Genre to resolve inconsistencies and to integrate variants employed by later field-workers. In the published volumes of the ALI, the base letters are as follows:[18]

ALI: a b c d e ə f g h i j k l ł m n o p r ꝛ ꞅ s ʃ t u v w y z ʒ ʔ α β γ δ ε ζ η θ λ ƛ ρ σ φ χ ω
IPA:[19] ä b – d ẹ ə f g h i j k l ɬ̟ m n ọ p r r̥ ɹ? s z t u v w ý t͜s d͜z ʔ ǟ̆ β ɣ ð ɛ̆ z̠ ŋ θ ɬ ɬ̟ ʀ s̠ ɸ x ð̆

We propose the following additions for this notation:

- Base letters: ꞅ (fig. 33)
- Modifier letters: ꝛ ꞅ ζ ρ σ (figs. 37, 38, 39, 40, 41)
- Combining letters: j ꝛ ꞅ ʒ y γ δ ζ θ σ φ χ (figs. 12, 24, 19, 20, 23, 25, 26, 27, 28, 29, 30, 31)

The publication of the long-awaited ALI continues to this day. The tenth volume appeared in 2023, and the eleventh is currently in preparation.

4. *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 letters are as follows:

RFE: a ɐ b ƀ ƀ̵ c d đ đ̵ e ə f g g̵ g̵̵ h ɦ i ị j k l ɫ ł m n ŋ ŋ̵ o ɔ p r ɹ s t u ᵾ v w x y z θ φ
IPA:[20] ä ǟ̆ b β β̞ - d̪ ð ð̞ ẹ ĕ f g ɣ ɣ̞ h ɦ i ĭ j k l ɫ ʟ? m n ŋ ŋ̞ ọ ŏ p ɾ ɹ s t̪ u ŭ v w x̠ ʝ z θ̟ ɸ

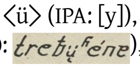
In the later series of regional atlases directed by Manuel Alvar, of which we cite two (the ALEA and ALEANR), we note the following differences:

- Absence of ⟨ƀ đ g̵ ɫ ŋ̵⟩
- ⟨ʉ⟩ in place of ⟨ᵾ⟩[21]

[18] ALI made a few graphic changes to the set of symbols established in Genre (1978): ⟨ξ⟩ was replaced with ⟨ζ⟩ and ⟨r̄, r̥⟩were replaced with ⟨ꝛ, ꞅ⟩. In addition, the maps feature ⟨ə⟩, a symbol used by field-workers other than Pellis.

[19] In light of certain complications in the officially published guide (cf. Ghia 2023: 29, footnote 19 and Canepari 2007: 105–107), we have based our phonetic analysis on Genre 1978.

[20] We base this analysis on the descriptions provided in the introduction to volume I of the ALPI and in Navarro Tomás 2004.

[21] Perhaps this substitution was intended to increase the visual distinction from ⟨ü⟩ (IPA: [y]), particularly in handwriting. Some non-distinctive tokens of ⟨ᵾ⟩ do appear in his atlases (e.g. ALEA map 4, point Ca 100: ), and in any case we suggest encoding Alvar's ⟨ʉ⟩ as ⟨ᵾ⟩ for consistency with other RFE atlases.

As in the BA and Pellis-Genre systems, weak articulation is indicated with superscripts while combining letters are used to indicate intermediate phonetic values.

We propose the following additions:

- Base letters: ɬ (fig. 32)
- Modifier letters: ɬ ɲ (figs. 35, 36)
- Combining letters: ƀ đ ɠ j ɦ ɬ ɔ ɹ ᵾ y θ φ (figs. 5–6, 8, 9, 12, 10, 14, 17, 18, 22, 23, 28, 30)

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

Particularities: Pseudo-subscripts

Since the combination of a base and combining letter may become rather tall, especially when combined with diacritics, sometimes the entire combination is shifted downward to compensate, as if the base letter were a subscript. This occurs systematically in the ALI and ALPR, as in Figure 1a. Since these pseudo-subscripts are a typographical device, without phonetic significance, the supporting letters should be encoded as normal baseline characters and we do not request Unicode support.

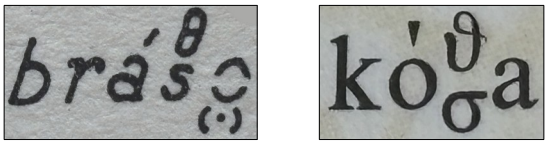


Figure 1a. ① ⟨s̰͡θ⟩ in ALPI map 27, point 569. The *s* is reduced in size to accommodate the tall *θ* but is set on the baseline. ② ⟨ǫ̰͡θ⟩ in ALI map 9, point 394. The *σ* is set as a subscript, as are all combined letters in the ALI, without semantic distinction from a baseline *σ*. In both cases the supporting letter should be encoded as a simple *s* or *σ*, and there is no need to add a subscript sigma to Unicode.

Offset combining letters

When a base letter has an ascender, such as *k*, a combining letter may be offset to the right and may thus resemble a spacing superscript. This is the case with the *h* in ⟨bǫ̰́ᵏʰkȧ⟩ in Figure 1b, where it can be seen that, because it is docked to a superscript, it can only be a combining letter. We have avoided such tokens in the illustrative figures where possible, but one can be seen in Figure 9.

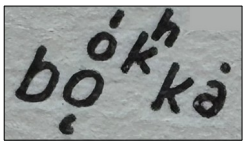


Figure 1b. ⟨bǫ̰́ᵏʰkȧ⟩ in ALPI map 2, point 532, with a laterally offset combining *h* on a superscript *k*.

Allography

Glyphs that correspond to distinct Unicode characters may be allographic in the context of these atlases. For example, the symbol for IPA [ŋ] may be ⟨ŋ⟩, ⟨ɳ⟩ or Greek ⟨η⟩ with no difference in meaning. This example is not unique to these atlases: modifier Greek η is well attested as a velar

nasal in other sources, but it has not been accepted as a Unicode character because it is always either a typographic hack for, or a conventionalized allograph of, the more generic ⟨ŋ⟩. Similarly, modifier Greek α can be substituted with Latin ⟨ᵅ⟩ without loss of meaning. More idiosyncratic to these atlases, tokens of Greek ξ in early atlases – including modifier and combining forms – are equivalent to ζ in later atlases, again with no contrastive use or difference in meaning (Figure 39). We propose that the later zeta form be taken as representative, and there is no need for a separate modifier or combining ξ in Unicode to support the Romance dialectological literature, unless an argument is made for diplomatic transcription.

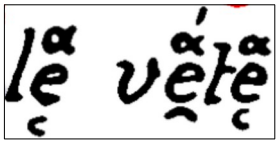


Figure 1c. Combining Greek α could be replaced with Latin ⟨◌ᵅ⟩ without loss of meaning. AIS map 131, point 439.

The allography of subscript letters with reduced-size baseline letters is noted in footnote 11.

Characters

◌᷍ 1AF5 COMBINING INVERTED LAZY S. Figure 2, Figure 34.

◌᷎ 1AF6 COMBINING COMMA ABOVE AND ACUTE. Figure 3.

Combining IPA letters

◌ᷓ 1AF7 COMBINING LATIN SMALL LETTER TURNED A. Figure 4.

◌ᷔ 1AF8 COMBINING LATIN SMALL LETTER H WITH HOOK. Figure 10.

◌ᷕ 1AF9 COMBINING LATIN SMALL LETTER DOTLESS I. Figure 11, Figure 61.

◌ᷖ 1AFA COMBINING LATIN SMALL LETTER J. Figure 12.

◌ᷗ 1AFC COMBINING LATIN SMALL LETTER ENG. Figure 15.

◌ᷘ 1AFD COMBINING LATIN SMALL LIGATURE OE. Figure 16.

◌ᷙ 107CE COMBINING LATIN SMALL LETTER TURNED R. Figure 18.

◌ᷚ 1AFE COMBINING LATIN SMALL LETTER Y. Figure 23.

◌ᷛ 1AFF COMBINING LATIN SMALL LETTER EZH. Figure 24.

Subscript letter

ᵤ AB6F LATIN SUBSCRIPT SMALL LETTER U WITH LEFT HOOK. Figure 60.

Combining Greek letters

◌ᷜ 107C0 COMBINING GREEK SMALL LETTER GAMMA. Figure 25.

◌ᷝ 107C1 COMBINING GREEK SMALL LETTER DELTA. Figure 26.

◌ᷞ 107C2 COMBINING GREEK SMALL LETTER ZETA. Figure 27.

◌ᷟ 107C3 COMBINING GREEK SMALL LETTER THETA. Figure 28.

◌ᷠ 107C4 COMBINING GREEK SMALL LETTER SIGMA. Figure 29.

◌ᷡ 107C5 COMBINING GREEK SMALL LETTER PHI. Figure 30.

◌ᷢ 107C6 COMBINING GREEK SMALL LETTER CHI. Figure 31.

Combining letters with stroke

◌ᵬ 107C7 COMBINING LATIN SMALL LETTER B WITH STROKE. Figure 5 *ff.*

◌ᶝ 107C8 COMBINING LATIN SMALL LETTER C WITH BAR. Figure 7.

◌ᶞ 107C9 COMBINING LATIN SMALL LETTER D WITH STROKE. Figure 8.

◌ᶢ 107CA COMBINING LATIN SMALL LETTER G WITH STROKE. Figure 9.

◌ᶩ 107CB COMBINING LATIN SMALL LETTER L WITH STROKE. Figure 14.

Combining dialectology letters

◌ᶫ 107CC COMBINING LATIN SMALL LETTER L WITH MIDDLE TILDE. Figure 13.

◌ᶻ 107CD COMBINING LATIN SMALL LETTER SCRIPT R. Figure 19.

◌ᶭ 107CE COMBINING LATIN SMALL LETTER REVERSED SCRIPT R. Figure 20.

◌ᶸ 107CF COMBINING LATIN SMALL LETTER U WITH LEFT HOOK. Figure 21.

◌ᶹ 107D0 COMBINING LATIN SMALL LETTER SPLIT O. Figure 17.

◌ᶺ 107D1 COMBINING LATIN SMALL LETTER SPLIT U. Figure 22.

ᶭ 1DF8B LATIN SMALL LETTER REVERSED SCRIPT R. Figure 33.

Letters with half stroke [to be continued]

ᶩ 1DF8C LATIN SMALL LETTER L WITH HALF STROKE. Figure 32.

Modifier letters

ᶻ 1DFB6 MODIFIER LETTER SMALL ZETA. Figure 39.

ᶿ 1DFB7 MODIFIER LETTER SMALL RHO. Figure 40.

ᶾ 1DFB8 MODIFIER LETTER SMALL SIGMA. Figure 41.

ᶝ 1DFB9 MODIFIER LETTER SMALL C WITH BAR. Figure 34.

ᶩ 1DFBA MODIFIER LETTER SMALL L WITH HALF STROKE. Figure 35.

ᶯ 1DFBB MODIFIER LETTER SMALL BARRED ENG. Figure 36.

ᶼ 1DFBC MODIFIER LETTER SMALL SCRIPT R. Figure 37.

ᶭ 1DFBD MODIFIER LETTER SMALL REVERSED SCRIPT R. Figure 38.

Subscript letters

ₔ 1DFBE LATIN SUBSCRIPT SMALL LETTER C WITH BAR. Figure 46.

ₕ 1DFBF LATIN SUBSCRIPT SMALL LIGATURE OE. Figure 59.

ᵦ 1DFC0 LATIN SUBSCRIPT SMALL LETTER B. Figure 43 *ff.*

ᵪ 1DFC1 LATIN SUBSCRIPT SMALL LETTER C. Figure 44, Figure 47 *ff.*

ᵨ 1DFC2 LATIN SUBSCRIPT SMALL LETTER D. Figure 44, Figure 51 *ff.*

ᵩ 1DFC3 LATIN SUBSCRIPT SMALL LETTER F. Figure 54 *ff.*

ᵧ 1DFC4 LATIN SUBSCRIPT SMALL LETTER G. Figure 57 *ff.*

Properties

We name some combining letters ‘Greek’ because they have a Greek rather than Latin form (for example, Greek ⟨ɣ⟩ and ⟨φ⟩ rather than Latin ⟨ɣ⟩ and ⟨ɸ⟩), but they are nonetheless intended for use in Latin text. Because the Unicode script property is determined by the supporting character, these do not need to be defined as either Latin or Greek.

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

1AF5;COMBINING INVERTED LAZY S;Mn;230;NSM;;;;;N;;;;;
1AF6;COMBINING COMMA ABOVE AND ACUTE;Mn;230;NSM;;;;;N;;;;;
1AF7;COMBINING LATIN SMALL LETTER TURNED A;Mn;230;NSM;;;;;N;;;;;
1AF8;COMBINING LATIN SMALL LETTER H WITH HOOK;Mn;230;NSM;;;;;N;;;;;
1AF9;COMBINING LATIN SMALL LETTER DOTLESS I;Mn;230;NSM;;;;;N;;;;;
1AFA;COMBINING LATIN SMALL LETTER J;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 TURNED R;Mn;230;NSM;;;;;N;;;;;
1AFE;COMBINING LATIN SMALL LETTER Y;Mn;230;NSM;;;;;N;;;;;
1AFF;COMBINING LATIN SMALL LETTER EZH;Mn;230;NSM;;;;;N;;;;;

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

107C0;COMBINING GREEK SMALL LETTER GAMMA;Mn;230;NSM;;;;;N;;;;;
107C1;COMBINING GREEK SMALL LETTER DELTA;Mn;230;NSM;;;;;N;;;;;
107C2;COMBINING GREEK SMALL LETTER ZETA;Mn;230;NSM;;;;;N;;;;;
107C3;COMBINING GREEK SMALL LETTER THETA;Mn;230;NSM;;;;;N;;;;;
107C4 COMBINING GREEK SMALL LETTER SIGMA;Mn;230;NSM;;;;;N;;;;;
107C5;COMBINING GREEK SMALL LETTER PHI;Mn;230;NSM;;;;;N;;;;;
107C6;COMBINING GREEK SMALL LETTER CHI;Mn;230;NSM;;;;;N;;;;;
107C7;COMBINING LATIN SMALL LETTER B WITH STROKE;Mn;230;NSM;;;;;N;;;;;
107C8;COMBINING LATIN SMALL LETTER C WITH BAR;Mn;230;NSM;;;;;N;;;;;
107C9;COMBINING LATIN SMALL LETTER D WITH STROKE;Mn;230;NSM;;;;;N;;;;;
107CA;COMBINING LATIN SMALL LETTER G WITH STROKE;Mn;230;NSM;;;;;N;;;;;
107CB;COMBINING LATIN SMALL LETTER L WITH STROKE;Mn;230;NSM;;;;;N;;;;;
107CC;COMBINING LATIN SMALL LETTER L WITH MIDDLE TILDE;Mn;230;NSM;;;;;N;;;;;
107CD;COMBINING LATIN SMALL LETTER SCRIPT R;Mn;230;NSM;;;;;N;;;;;
107CE;COMBINING LATIN SMALL LETTER REVERSED SCRIPT R;Mn;230;NSM;;;;;N;;;;;
107CF;COMBINING LATIN SMALL LETTER U WITH LEFT HOOK;Mn;230;NSM;;;;;N;;;;;
107D0;COMBINING LATIN SMALL LETTER SPLIT O;Mn;230;NSM;;;;;N;;;;;
107D1;COMBINING LATIN SMALL LETTER SPLIT U;Mn;230;NSM;;;;;N;;;;;

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

Besides notes on look-alike characters, subscript *c* and *s* are attested with a combining cedilla.

02E2 MODIFIER LETTER SMALL S
- supports 0327 combining cedilla

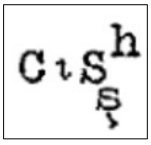
1AF1 COMBINING INVERTED LAZY S
→ 1DD1 COMBINING UR ABOVE
→ 1AB2 COMBINING INFINITY

1AF2 COMBINING COMMA ABOVE AND ACUTE
→ 1FCE GREEK PSILI AND OXIA

209B LATIN SUBSCRIPT SMALL LETTER S
- supports 0327 combining cedilla [see at right]

107C1 COMBINING GREEK SMALL LETTER DELTA
→ 1DD8 COMBINING LATIN SMALL LETTER INSULAR D

1DFC1 LATIN SUBSCRIPT SMALL LETTER C
- supports 0327 combining cedilla



Winnick (1972) uses subscript Turkish ş, as here in ⟨CıS$_{ş}^{h}$⟩, as well as subscript ç as in Figure 50.

Charts

This request completes the Combining Diacritical Marks Extended block and requires an additional block. SEW recommends a new block at 107C0..107FF named *Combining Diacritical Marks Extended-A*.

Combining Diacritical Marks Extended
1AB0 1ADF

	1AB	1AC	1AD	1AE	1AF
0					
1					
2					
3					
4					
5					
6					
7					
8					
9					
A					
B					
C					
D					
E					
F					

Combining Diacritical Marks Extended-A

107C0 107FF

	107C	107D	107E	107F
0	◌᷀	◌᷐		
1	◌᷁	◌᷑		
2	◌᷂			
3	◌᷃			
4	◌᷄			
5	◌᷅			
6	◌᷆			
7	◌᷇			
8	◌᷈			
9	◌᷉			
A	◌᷊			
B	◌᷋			
C	◌᷌			
D	◌᷍			
E	◌᷎			
F	◌᷏			

Latin Extended-E

AB30 **AB60**

	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	꭛	꭫
C	ꬼ	ꭌ	ꭜ	꭬
D	ꬽ	ꭍ	ꭝ	꭭
E	ꬾ	ꭎ	ꭞ	꭮
F	ꬿ	ꭏ	ꭟ	꭯

Latin Extended-G

1DF00 **1DFFF**

	1DF0	1DF1	1DF2	1DF3	1DF4	1DF5	1DF6	1DF7	1DF8	1DF9	1DFA	1DFB	1DFC	1DFD	1DFE	1DFF
0	𝼀	𝼐	ʤ	ɢ	A̶	ɹ̶	ω	j	A̅				b	ɣ	ʤ	ɭ
1	𝼁	𝼑	ʤ	ɣ	a̶	Ɐ̶	ʌr	j	E				c	ᴘ	ʤ	ɳ
2	𝼂	𝼒	ʨ	ɧ	ə	v̶	ʃh	ɵ	w̡				d	ƀ	ʆ	ŧ
3	𝼃	𝼓	ʨ	ɸ	ᴇ̶	ɷ̶	ʈh	ɵ	ɴ̶				f	đ	ʃ	ψ
4	𝼄	𝼔	tθ	ɋ	ɛ	ꭓ	ɟh	Ɯ	ɍ				g	ǥ	ʧ	ω
5	𝼅	𝼕	𝼥	ʀ	ɡ	y̶	wh	ɷ	ɪ				ʺ	ɨ	ʧ	ɖ
6	𝼆	𝼖	𝼦	ʞ	ħ	ʔ̶	ƨ	Ɒ	ɕ			ζ	ˏ	ɨ	ʒ	ɧ
7	𝼇	𝼗	𝼧	ɽ	ɦ̶	ɹ	ɮ	ɷ	X̶			ρ	ˎ	ɟ	ʒ	ɳ
8	𝼈	𝼘	𝼨	ʦ	ᴋ	ɔ	A	Ⱨ	x̶			σ	↗	ɥ	ᴅ	ʂ
9	𝼉	𝼙	𝼩	ɥ	ꝁ	ṳ	ᶏ	θ	ɳ			ϵ	˂˄	ɉ	ᴀ	ʐ
A	𝼊	𝼚	𝼪	ʕ	M̶	a	A	U	ɚ			ł	˃˄	ʉ	ᴇ	ɕ
B	𝼋	𝼛	ʤ	ß	m̶	ɖh	ƨ	ɥ	ƨ			ŋ	˂˅	w̶	ɿ	ƙ
C	𝼌	𝼜	ʧ	θ	ɱ̶	ɛɛ	Є	Ɯ	ɟ			z	˃˅	η	ɿ	ꞵ
D	𝼍	𝼝	ɖ	χ	N̶	ie	є	ɯ				ƨ	ɹ	ɼ	ɥ	ɗ
E	𝼎	𝼞	ʣ	ʮ	n̶	ɔi	ŧ	Ŋ				ϵ	ɔ	ȸ	ɥ	ƒ
F	𝼏	dð	ð	w̶	ŋ̶	ɔu	j	ɪj				œ	ü	φ	ɖ	f

References

Atlases

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

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

ALCe: Pierrette DUBUISSON (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 vols. Granada: Universidad de Granada-CSIC.

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

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

ALEIC: Gino Bottiglioni (1933–1942) *Atlante linguistico-etnografico italiano della Corsica*, 10 vols. Pisa.

ALF: Jules Gilliéron & Edmond Edmont (1902–1910) *Atlas linguistique de la France*, 12 vols. 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 vols. Paris: CNRS.

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

ALMex: Juan Manuel Lope Blanch (1990–2000) *Atlas lingüístico de México*, 6 vols. El Colegio de México.

ALN: Patrice Brasseur (1980–2019) *Atlas linguistique et ethnographique normand*, 5 vols. 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 vols. 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 vols. Paris: CNRS. http://anr-appi.univ-lille.fr/index.php/alpic

ALPR: Tomás Navarro Tomás (1948) Atlas Lingüístico de Puerto Rico. In *El español de Puerto Rico*, 253–327. Río Piedras: Universidad de Puerto Rico. http://www.alpr.info/

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

ALR-sn: Emil Petrovici (1956–1972) *Atlasul lingvistic român: serie nouă*, 7 vols. Bucharest: Editura Academiei.

AMDV: Graziano Tisato *et al.* (2014) *Atlante Multimediale dei Dialetti Veneti*. Fondazione Cariparo/University of Padua/ISTC-CNR. https://www3.pd.istc.cnr.it/amdv/

ELPA: Charles Bruneau (1914–1926) *Enquête linguistique sur les patois d'Ardenne*, 2 vols. Paris: Champion.

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

NALR-B: Petru Neiescu (1980–) *Noul atlas lingvistic român pe regiuni: Banat*, 5 vols. Bucharest: Editura Academiei.

NALR-MB: Vasile Arvinte, Stelian Dumistrăcel, Ion Florea, Ion Nuță & Adrian Turculeț (1987–2022) *Noul atlas lingvistic român pe regiuni: Moldova și Bucovina*, 5 vols. Bucharest: Editura Academiei (vols. 1–2) / Iași: Alexandru Ioan Cuza University Press (vols. 3–5).

NALR-O: Boris Cazacu (1967–1984) *Noul atlas lingvistic român pe regiuni: Oltenia*, 5 vols. Bucharest: Editura Academiei.

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

Digitization projects

Esther Baiwir & Cécile Kaisin (2022) L'ALPic en ligne et le projet d'un Atlas pan-picard informatisé (APPI). *Bien Dire et Bien Aprandre* 35: 47-60. https://anr-appi.univ-lille.fr/

Johnatan Estiven Bonilla, Ruth Yanira Rubio López, Andrea Lizeth Llanos, Julio Alexander Bernal-Chávez & Daniel Bejarano (2020) Proyecto de digitalización y nuevas perspectivas tecnológicas del Atlas Lingüístico-Etnográfico de Colombia. In Francesc Roca Urgell & Ángel J. Gallego (eds.), *Dialectología digital del español*, 13–28. Santiago de Compostela: Anexos de Verba 80. https://alec.caroycuervo.gov.co/

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/

Alberto Ghia (2023) Verso il DigitALEIC: Riflessioni sulla trasformazione dei materiali fonetici dalla grafia fonetica dell'ALEIC all'IPA. *Bolletino dell'Atlante Linguistico Italiano*, 3rd series 47: 109–141.

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/

Carolina Julià Luna (2021) Del atlas lingüístico tradicional al corpus geolingüístico digital: Diseño de un proyecto. *Scriptum Digital* 10: 109–147. https://corpat.es/

Christina Mutter & Aleksander Wiatr (2018) The virtual research environment of VerbaAlpina and its lexicographic function. In Krek, Čibej, Gorjanc & Kosem (eds.), *Proceedings of the XVIII EURALEX international congress: Lexicography in global contexts*, 775–785. Založba Univerze v Ljubljani. http://verba-alpina.gwi.uni-muenchen.de/

Xulio Sousa (2017) From field notebooks to automatic mapping: the 'Atlas Lingüístico Galego' database. *Dialectologia et Geolinguistica* 25 (1): 1–22.

Graziano Tisato (2014) Metodologia e strumenti di sviluppo dell'Atlante Multimediale dei Dialetti Veneti. *Atti Del Sodalizio Glottologico Milanese*, V. 2012 (7): 172–192.

Ricardo Tranquilli Navarro (2019) Bases para la creación del Atlas lingüístico y etnográfico de la provincia de Zaragoza (ALEPZ). *Archivo de filología aragonesa* 75: 189–212. ifc.dpz.es/index/alepz/

Martijn Wieling, Eva Sassolini, Sebastiana Cucurullo & Simonetta Montemagni (2016) ALT explored: Integrating an online dialectometric tool and an online dialect atlas. In *Proceedings of the tenth international conference on language resources and evaluation* (LREC'16), 3265–3272. Portorož, Slovenia: European Language Resources Association (ELRA).

Other

Vytautas Ambrazas, ed. (2006) [1977] *Lithuanian Grammar*, 2nd ed. Vilnius: Baltos Lankos.

Frasz Boas, Pliny Goddard, Edward Sapir & Alfred Kroeber (1916) *Phonetic Transcription of Indian Languages: Report of Committee of American Anthropological Association*. Washington: Smithsonian. http://hdl.handle.net/10088/23549

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

Oliver Buesa Tomás & Luis Flórez (1954) *El Atlas lingüístico-etnográfico de Colombi (ALEC)*. Bogota: Instituto Caro y Cuervo.

Luciano Canepari (2007) *Natural phonetics and tonetics*. Munich: LINCOM Europa.

Arturo Genre (1978) Tavola di unificazione dei segni di trascrizione di Ugo Pellis. *Bolletino dell'Atlante Linguistico Italiano*, 3rd series, 2: 57–86.

Jules Gilliéron & Jean-Pierre Rousselot (1887) *Revue des patois gallo-romans*, volume I. Paris: Champion.

Jules Gilliéron & Edmond Edmont (1902) *Notice servant à l'intelligence des cartes*. Paris: Champion.

Karl Jaberg & Jakob Jud (1928) *Der Sprachatlas als Forschungsinstrument: Kritische Grundlegung und Einführung in den Sprach- und Sachatlas Italiens und der Südschweitz*. Halle: Niemeyer.

Lyudmila Kalnyn' [Л.Э. Калнынь] (1973) *Опыт моделирования системы украинского диалектного языка*. Moscow: Nauka.

Lyudmila Kalnyn' & Lyudmila Maslennikova [Л.Э. Калнынь & Л.И. Масленникова] (1981) *Сопоставительная модель фонологической системы славянских диалектов*. Moscow: Nauka.

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

John Kelly & John Local (1989) *Doing Phonology: Observing, Recording, Interpreting*. Manchester University Press.

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

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

Giorgio Nagy (2024) *Varietà linguistiche romanze in Italia e Svizzera: un'analisi statistica*. University of Padua, thesis. https://thesis.unipd.it/handle/20.500.12608/84088

Tomás Navarro Tomás (2004) [1918] *Manual de pronunciación española*, 28th ed. Madrid: CSIC.

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.

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

Bengisu Winnick (1972) *The phonetics and phonology of Istanbul Turkish*. SOAS University of London, thesis. https://doi.org/10.25501/SOAS.00028654

Figures[22]

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 121, 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), ④ ALEIC map 2 point 4.

[22]In addition to the aforementioned atlases, we cite some German (SBS), Italian (ALEIC), and Romanian atlases (ALR-sn, NALR-B, NALR-MB, NALR-O) that use closely related phonetic notations. In some cases we also provide supporting examples from other linguistic publications.

Figure 4. Combining ɐ. ① ALD-II map 17 point 37, ② ALEA map 4 point Gr 201, ③ ALPI questionnaire 236 point 544, ④ *Revista de Filología Española* 33: 24, from Jacquerye, ⑤ Buesa Oliver & Flórez (1954: *cuadro de las vocales españolas*). In the last example we have a turned ⟨a̟⟩, such that the combining letter is *under* the baseline letter. However, this difference in placement is allographic elsewhere, for example in the ALPI field notes.

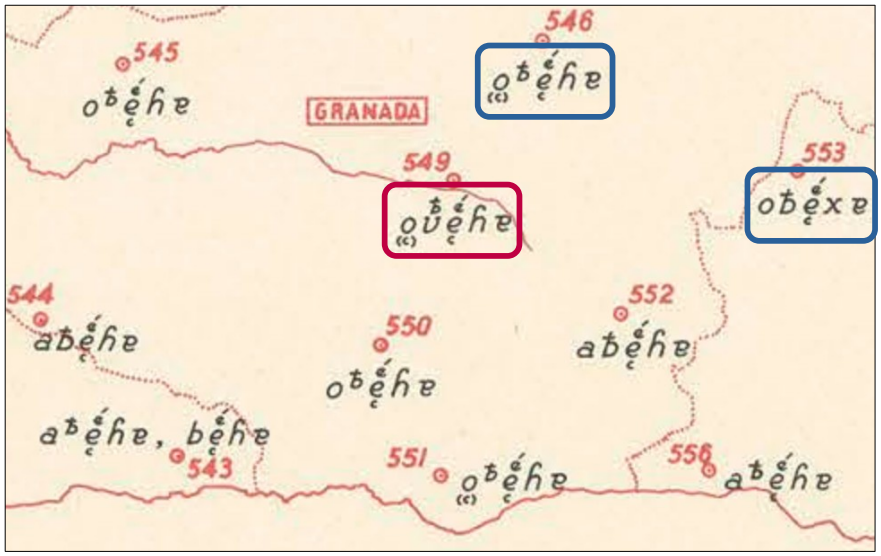


Figure 5. Combining ƀ. ALPI map 6 point 549. Note that combining ƀ (red) contrasts with modifier ƀ, which is currently in the Unicode pipeline, as well as baseline ƀ (blue).

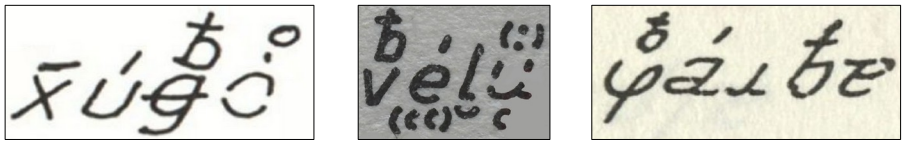


Figure 6. Combining ƀ. ① ALEANR map 329 point Hu 107, ② ALPI map 2 point 245, ③ ALEA map 616 point H 402.

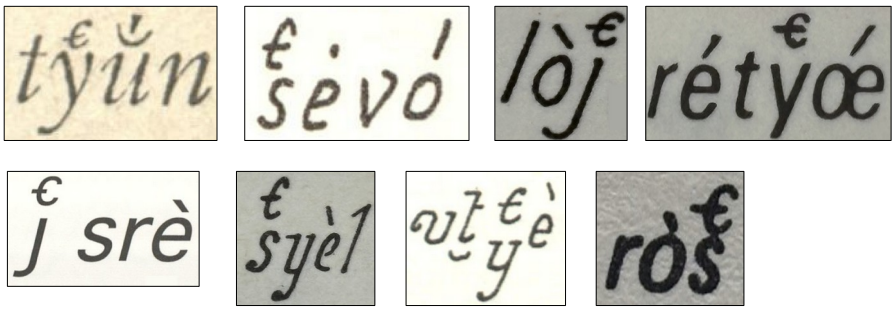


Figure 7. Combining ϵ. ① ALF map 9 point 791, ② ALLy map 311 point 15, ③ ALO map 122 point 66, ④ ALPic map 57 point 51, ⑤ ALN map 1460 point 33, ⑥ ALAL map 1 point 21, ⑦ ALJA 1590 point 5, ⑧ ALRé map 7 point 11.

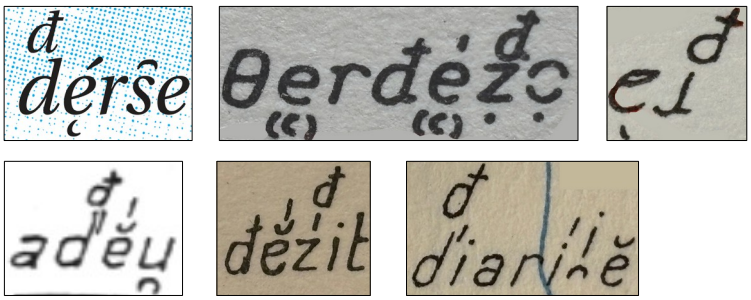


Figure 8. Combining đ. ① ALD-II map 60 point 93, ② ALPI map 2 point 141, ③ ALEA map 823 point J 501, ④ ALR-sn map 1466, from Jacquerye, ⑤ NALR-B map 124 point 15, ⑥ NALR-O map 136 point 936.

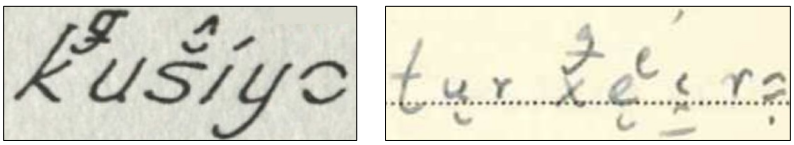


Figure 9. Combining g̰. ① ALEA map 562 point Al 405, ② ALPI questionnaire 479a point 150. It is a bit difficult to distinguish from a spacing diacritic in (1), but a comparison with similar symbol combinations on the same map, such as ⟨g̰⟩ at http://imgur.com/a/HyJAbJa, clarifies the situation.

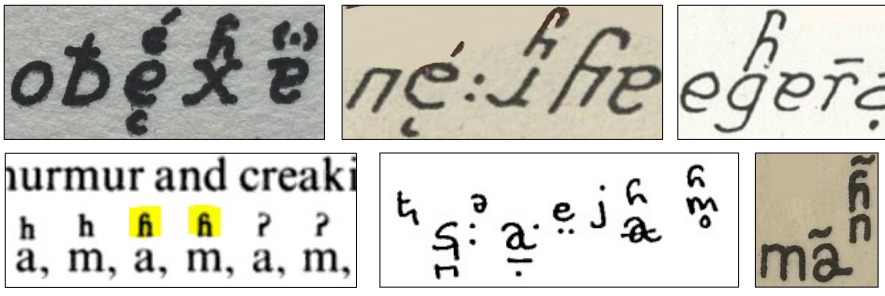


Figure 10. Combining ɦ. ① ALPI map 20 point 452, ② ALEA map 4 point Ma 405, ③ ALEANR map 336 point Z 502, ④ Pulleyblank (1984: 34), ⑤ Kelly & Local (1989: 88), ⑥ Buesa Oliver & Flórez (1954: *cuadro de las consonantes españolas*).

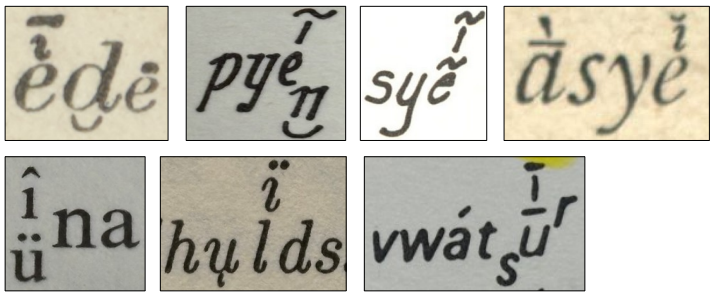


Figure 11. Combining dotless *i* with macron, breve, tilde, circumflex and trema. ① TPh column 236 line 11, ② ALAL map 273 point 26, ③ ALCe map 1438 point 13, ④ ALF map 8 point 156, ⑤ ALI map 45 point 51, ⑥ SBS 7.2: 388, ⑦ ALRé map 249 point 35. This character is only needed when a diacritic suppresses 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 61.

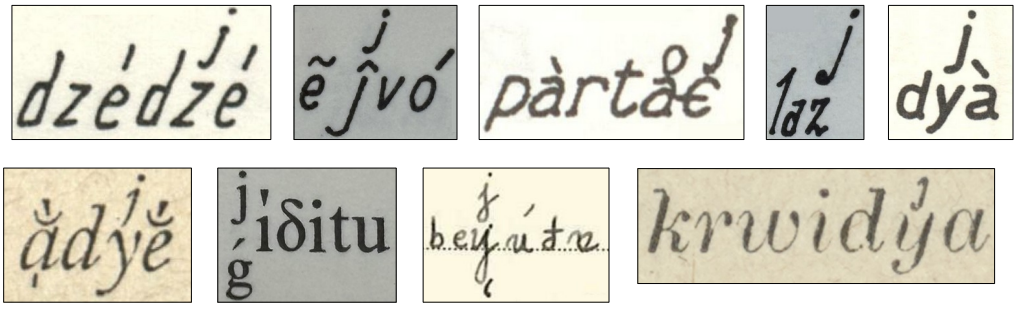


Figure 12. Combining *j*. ① ALLy map 496 point 17, ② ALO map 576 point 117, ③ ALN map 784 point 26, ④ ALAL map VI point 27, ⑤ ALCe map 516 zone B, ⑥ ALF map 9 point 895, ⑦ ALI map 47 point 945, ⑧ ALPI questionnaire 140 point 352, ⑨ TPh column 54 point 29 [combining *j* is always without tittle in TPh].

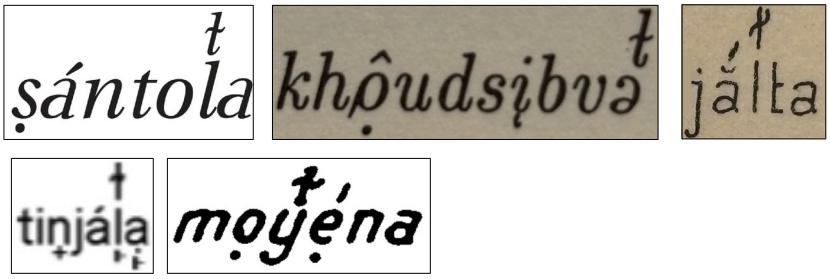


Figure 13. Combining *ł*. ① ALD-II map 7 point 211, ② SBS 11: 66, ③ ALR-sn map 11 point 48, ④ NALR-MB 2005: 36, from Jacquerye, ⑤ AIS map 989 point 393.

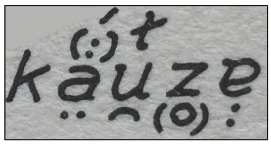


Figure 14. Combining *ɬ*. ALPI map 38 point 245.

Figure 15. Combining *ŋ*. ① ALD-II map 60 point 211, ② ALI map 153 point 445 (*ŋ* allograph), ③ VIVALDI «i prati sono verdi» (Issime), ④ ALR-sn map 1459 point 791, from Jacquerye, ⑤ AIS map 128 point 14 (*ŋ* allograph), ⑥ ALPI questionnaire 404 point 338, ⑦ Buesa Oliver & Flórez (1954: *cuadro de las consonantes españolas*).

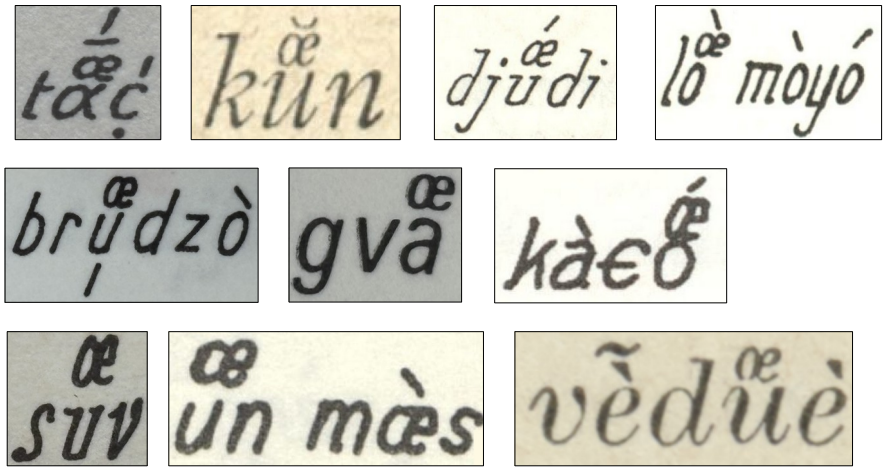


Figure 16. Combining *œ*. ① AIS map 864 point 149, ② ALF map 9 point 776, ③ ALJA map 92 point 13, ④ ALLy map 176 point 49, ⑤ ALO map 383 point 94, ⑥ ALPic map 158 point 9, ⑦ ALN map 782 point 5, ⑧ ALAL map 2 point 10, ⑨ ALCe map 65 point 11, ⑩ TPh column 179 line 20, notes.

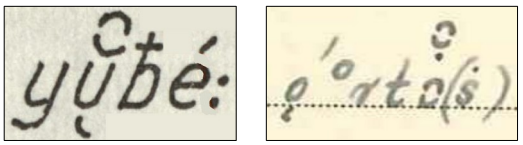


Figure 17. Combining ◌̚. ① ALEA map 1838 point H 402, ② ALPI questionnaire 247 point 145.

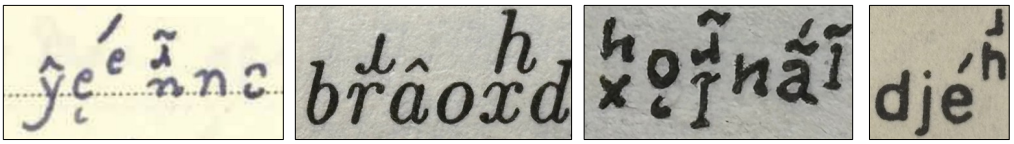


Figure 18. Combining ◌ʲ. ① ALPI questionnaire 134 point 486, ② SBS 7.2: 169, ③ ALPR map 20 point 14, ④ Buesa Tomás & Flórez (1954: *cuadro de las consonantes españolas*).

Figure 19. Combining ◌ᵋ. ALI map 94 point 858.

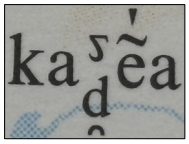


Figure 20. Combining reversed ◌ᵋ. ALI map 1 point 747. The reflection marks a difference in voicing.

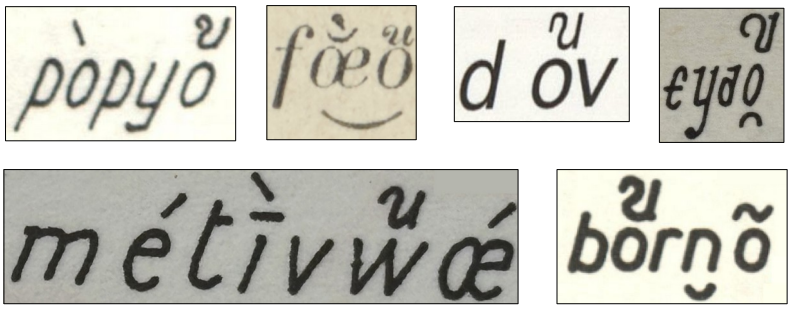


Figure 21. Combining ◌ᵾ. ① ALLy map 383 point 53, ② TPh column 50 point 13, ③ ALN map 1516 point 2, ④ ALAL map 1 point 61, ⑤ ALO map 61 point 62, ⑥ ALCe map 532 point 48.

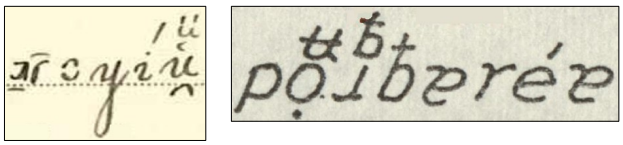


Figure 22. Combining ʉ. ① ALPI questionnaire 61 point 316, ② ALEA map 1588 point J 404 – the split ʉ is here replaced by barred ʉ̶, presumably because it's difficult to hand write a split ʉ in a way that is clearly distinct from ü (see footnote 21).

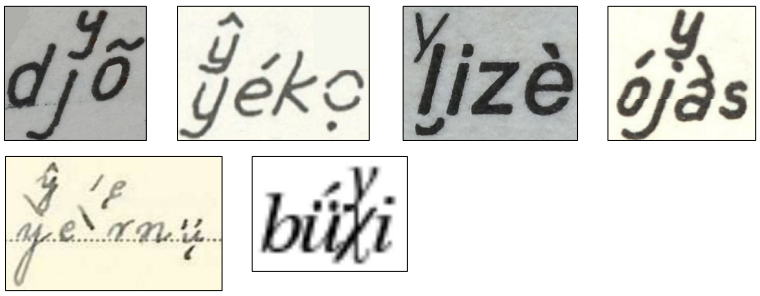


Figure 23. Combining y. ① ALO map 354 point 90, ② ALEANR map 11 point Lo 605, ③ ALAL map 74 point 9 (enquête D), ④ ALCe map 96 point 51, ⑤ ALPI questionnaire 134 point 339, ⑥ VIVALDI «ho letto questo libro» (Issime). Note that the *y* suppresses the dot of the supporting *j* in 1 but not in 4; other combining letters behave the same.

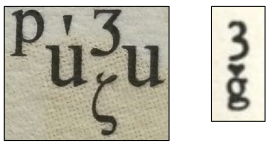


Figure 24. Combining ʒ. ① ALI map 18 point 577, ② Genre (1978: 61).

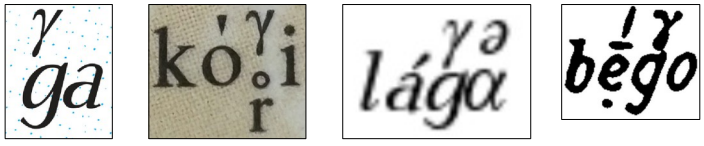


Figure 25. Combining ɣ. ① ALD-II map 24 point 185, ② ALI map II point 943, ③ VIVALDI «l'acqua» (Padola), ④ AIS map 457 point 511.

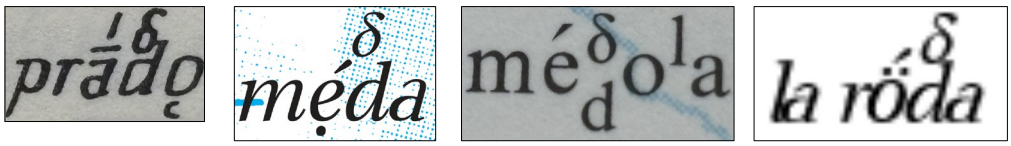


Figure 26. Combining ð. ① AIS map 1415 point 511, ② ALD-II map 5 point 145, ③ ALI map 3 point 253, ④ VIVALDI «la ruota» (Isola Sant'Antonio).

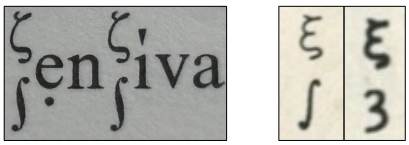


Figure 27. Combining ζ. ① ALI map 32 point 354, ② Genre (1978: 79, 61). ⟨ζ⟩ and ⟨ξ⟩ are allographic in this context; no atlas contrasts them.

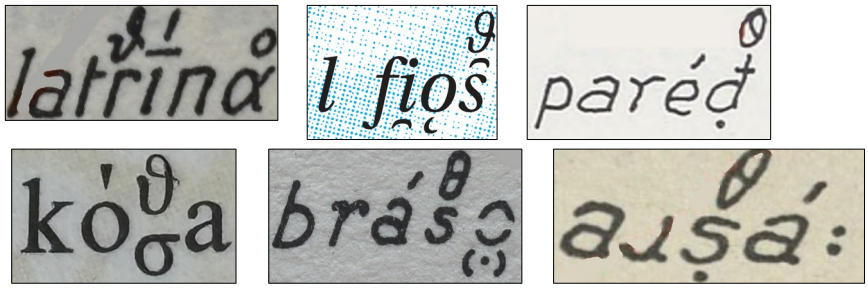


Figure 28. Combining θ. ① AIS map 871 point 222, ② ALD-II map 8 point 109, ③ ALEANR map 97 point 305, ④ ALI map 9 point 394, ⑤ ALPI map 27 point 569, ⑥ ALEA map 18 point Se 602. This should be identified as a normal theta θ rather than as the theta symbol ϑ.

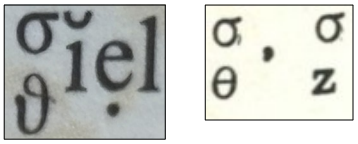


Figure 29. Combining σ. ① ALI map 33 point 349, ② Genre (1978: 61).

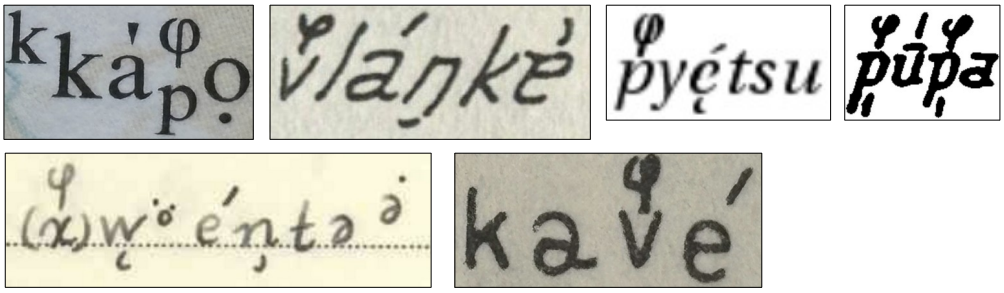


Figure 30. Combining φ. ① ALI map 8 point 524, ② ALEA map 315 point Al 403, ③ VIVALDI «dammi un altro pezzo» (Castelfondo), ④ AIS map 750 point 575, ⑤ ALPI questionnaire 92 point 559, ⑥ Buesa Oliver & Flórez (1954: *cuadro de las consonantes españolas*).

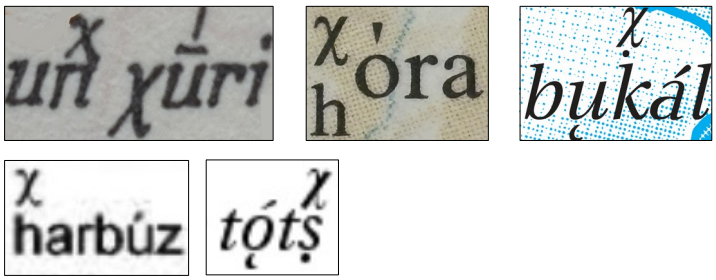


Figure 31. Combining χ. ① AIS map 1357 point 818, ② ALI map II point 969, ③ ALD-II map 90 point 76, ④ NALR-MB map 13 west of point 477, ⑤ VIVALDI «dammi un altro pezzo» (Castelfondo), from Jacquerye. The dot-below in 3 might be set above the *k*. However, this won't work for all situations, like combining ⟨e̯⟩ which will need a way to tell the font that the diacritic belongs on the combining letter.

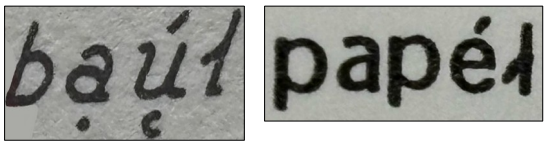


Figure 32. L with half stroke, ɬ. ① ALPI map 24 point 615, ② ALMEx map 81 point 189.

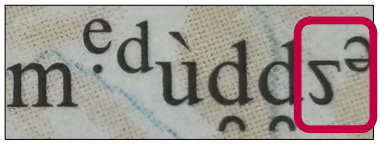


Figure 33. Reversed ɾ. ALI map 3 point 923.

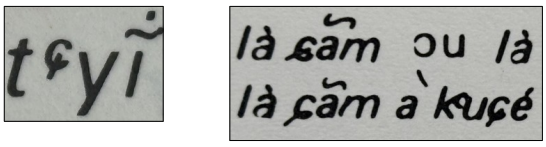


Figure 34. Modifier ϵ. ① ALPic map 188 point 20. The stroke is set at an angle; at right are several tokens from ALRé map 252, caption, of baseline ϵ showing similar variation in bar positioning.

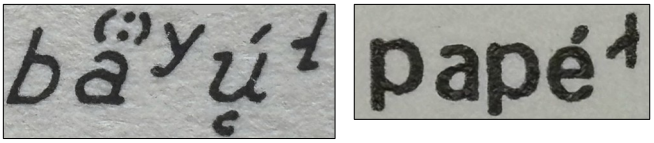


Figure 35. Modifier *l*. ① ALPI map 24 point 135, ② ALMEx map 81 point 183.

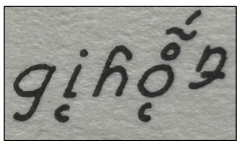


Figure 36. Modifier ɻ. ALPI map 11 point 519.

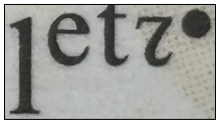


Figure 37. Modifier ɾ. ALI map 1 point 943.

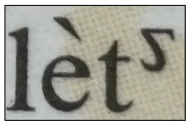


Figure 38. Modifier reversed ɾ. ALI map 1 point 46.

Figure 39. Modifier ʒ. ① ALI map 32 point 349, ② Genre (1978: 71). ⟨ʒ⟩ and ⟨ǯ⟩ are allographic in this context; no atlas contrasts them.

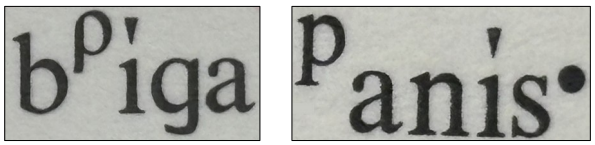


Figure 40. Modifier ρ. ALI map II point 94, with modifier *p* from elsewhere on the same map for comparison.

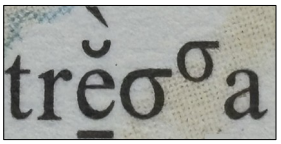


Figure 41. Modifier σ. ALI map 14 point 446.

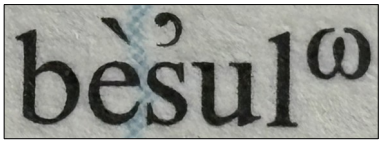


Figure 42. Modifier ω. ALI map 36 point 142. This is in the pipeline at U+1DFF4, but is not yet scheduled for Unicode 18.

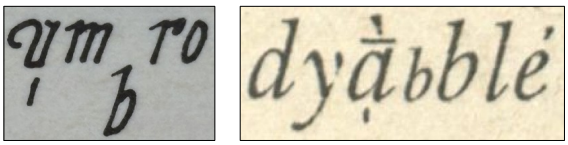


Figure 43. Subscript *b*. ① ALAL map 10 point 20, ② ALF map 403 point 699. The first generation of atlases set 'subscript' letters in a smaller type size, as at right, but the modern convention is to use true subscripts, as at left, and that is how we suggest digitizing the data.[11]

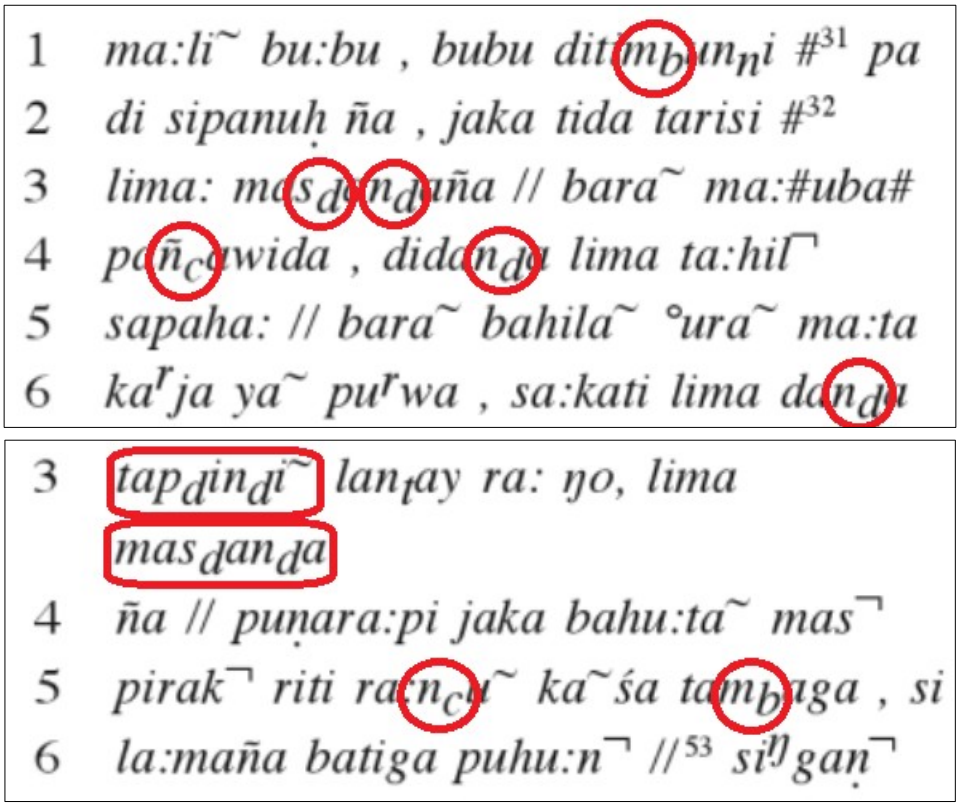


Figure 44. Subscript *b*, *c* and *d*. Kozok (2015: 69, 71). Conjuncts transliterated as subscripts. They commonly occur after nasals, but note ⟨s_d, p_d⟩. Subscripting in the diplomatic transliteration is not predictable from the normalized transliteration. For example, in Kozok (2015: P/5/6), normalized *geggah rabutti rampassi* is diplomatic ⟨gaggaḥ rabut$_t$irampas$_s$i⟩, with neither the second *g* in *gg* nor the *p* in *mp* subscript.

Figure 45. Subscript *b*. Kalnyn' & Maslennikova (1981: 396). Equivalent Cyrillic and Latin transcriptions. Cyrillic is already supported. Figure 58 shows that the superscript minus sign does not need to appear directly over the subscript, and that this can be digitized ⟨п$_б$˙’/p$_b$˙’⟩.

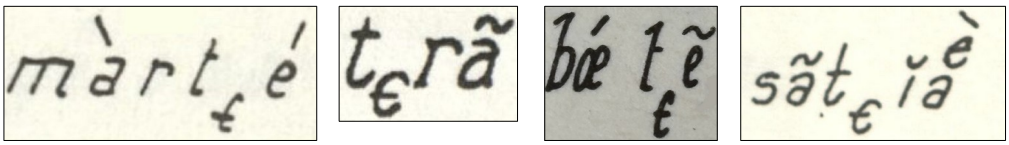


Figure 46. Subscript *є*. ① ALLy map 914 point 31, ② ALN map 823 point 3, ③ ALAL map 3 point 53, ④ ALCe map 1344 point 30.

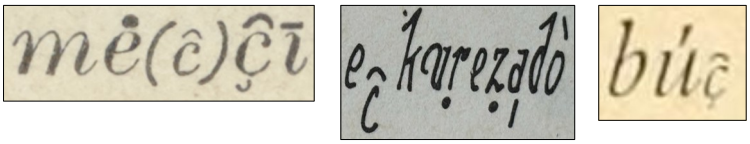
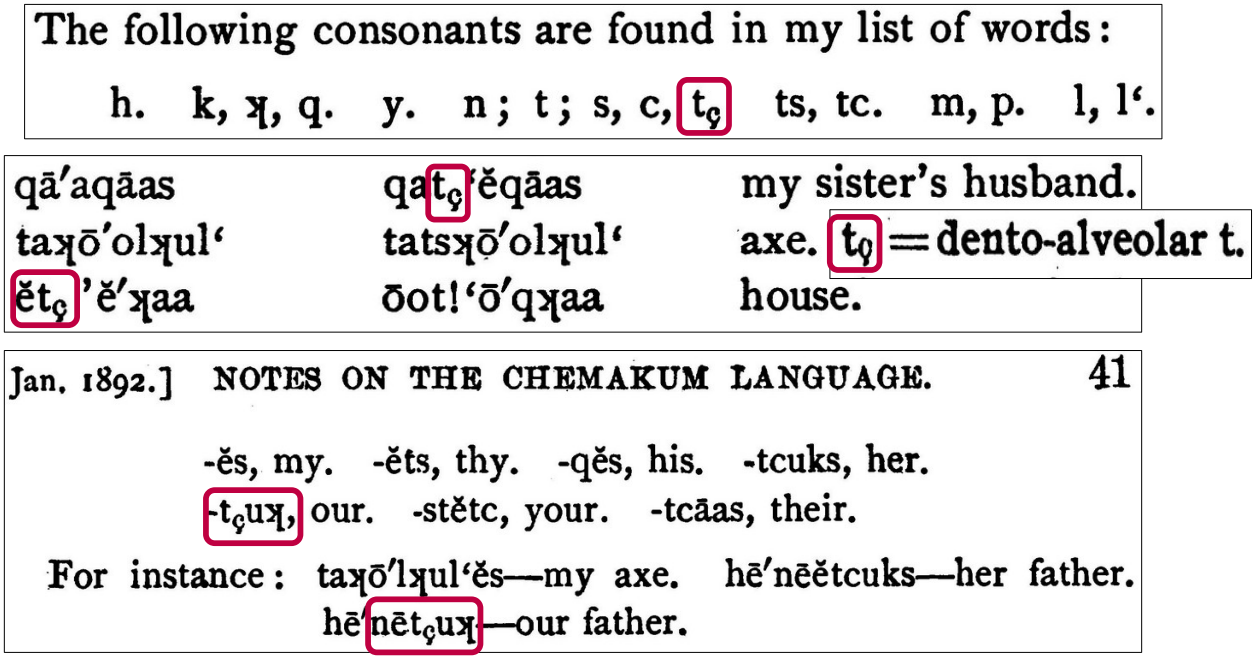


Figure 47. Subscript *c* with turned breve and cedilla. ① TPh column 381 line 13, ② ALAL map 42 point 27, ③ ELPA questionnaire 38. As is the case with superscript ⟨ᶜ̧⟩, a subscript ⟨$_{ç}$⟩ should be rendered as subscript *c* plus a combining cedilla.



The following consonants are found in my list of words:

h. k, ɬ, q. y. n; t; s, c, $t_{ç}$ ts, tc. m, p. l, l‘.

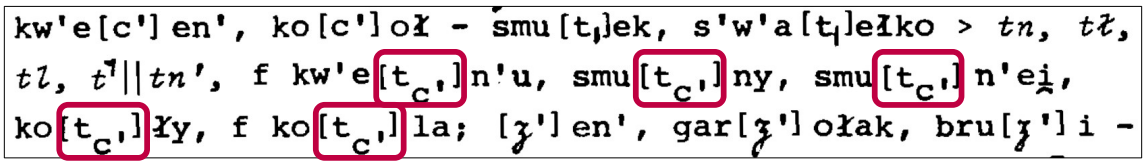
qā′aqāas	qa$t_{ç}$′ĕqāas	my sister’s husband.
taɬō′olɬul‘	tatsɬō′olɬul‘	axe. $t_{ǫ}$ = dento-alveolar t.
ĕ$t_{ç}$’ĕ′ɬaa	ōot!‘ō′qɬaa	house.

Jan. 1892.] NOTES ON THE CHEMAKUM LANGUAGE. 41

-ĕs, my. -ĕts, thy. -qĕs, his. -tcuks, her.
-$t_{ç}$uɬ, our. -stĕtc, your. -tcāas, their.

For instance: taɬō′lɬul‘ĕs—my axe. hē′nēĕtcuks—her father.
hē‘nēt$_{ç}$uɬ—our father.

Figure 48. Subscript *c* with cedilla. Boas (1892: 37 fn, 38, 39, 41).



kw'e[c'] en', ko[c'] oɬ - smu[$t_{ç}$]ek, s'w'a[$t_{ç}$]eɬko > *tn*, *tɬ*, *tl*, *t'||tn'*, f kw'e[$t_{ç}$'] n'u, smu[$t_{ç}$'] ny, smu[$t_{ç}$'] n'eɨ, ko[$t_{ç}$'] ɬy, f ko[$t_{ç}$'] la; [ʒ'] en', gar[ʒ'] oɬak, bru[ʒ'] i -

Figure 49. Subscript *c*. Kalnyn’ & Maslennikova (1981: 360). The alignment of the apostrophe needs to be handled by the font; U+0315 should work.

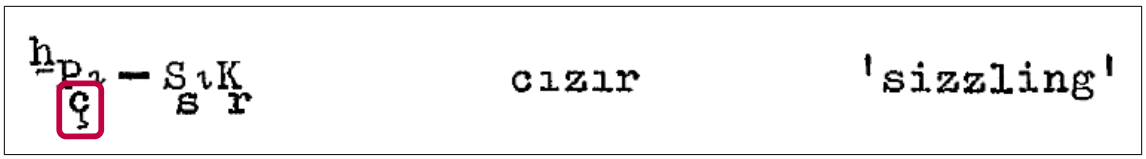


Figure 50. Subscript *c* with cedilla. Winnick (1972: 163).

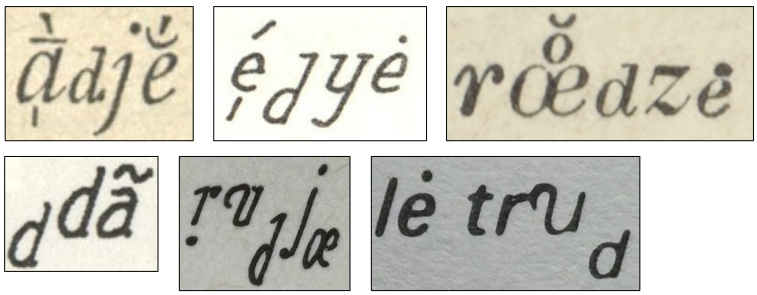
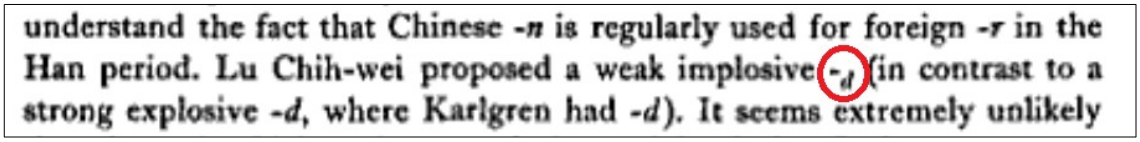
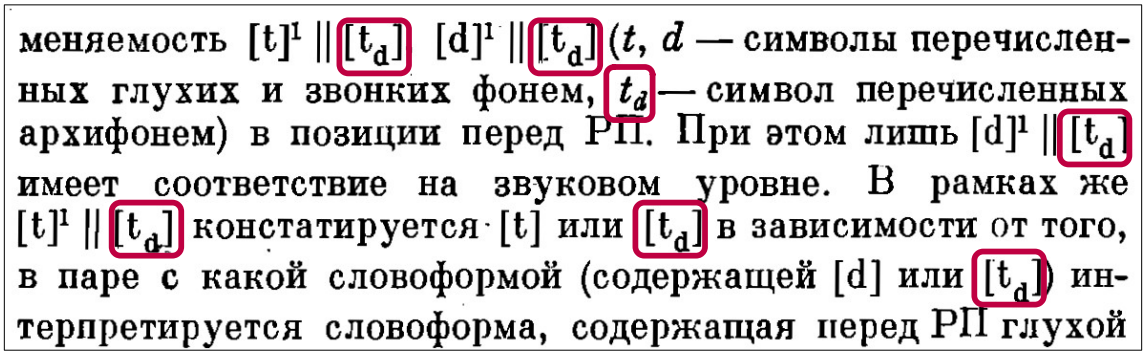


Figure 51. Subscript *d*. ① ALF map 9 point 794, ② ALJA map 135 point 41, ③ TPh column 248 line 17, ④ ALN map 783 point 21, ⑤ ALAL map 8 point 33, ⑥ ALRé map 8 point 37.



understand the fact that Chinese *-n* is regularly used for foreign *-r* in the Han period. Lu Chih-wei proposed a weak implosive *-d* (in contrast to a strong explosive *-d*, where Karlgren had *-d*). It seems extremely unlikely

Figure 52. 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 ⟨-b, -d, -g⟩.



меняемость [t]¹ || [t_d] [d]¹ || [t_d] (*t*, *d* — символы перечисленных глухих и звонких фонем, *t_d* — символ перечисленных архифонем) в позиции перед РП. При этом лишь [d]¹ || [t_d] имеет соответствие на звуковом уровне. В рамках же [t]¹ || [t_d] констатируется [t] или [t_d] в зависимости от того, в паре с какой словоформой (содержащей [d] или [t_d]) интерпретируется словоформа, содержащая перед РП глухой

Figure 53. Subscript *d*. Kalnyn' (1973: 327). Illustrations specific to Ukrainian are presented in Cyrillic, ⟨т⟩.

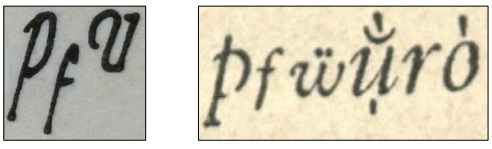


Figure 54. Subscript *f*. ① ALAL map 317 point 11, ② ALF map 1683 point 805.

Figure 55. Subscript *f* in ⟨$^{h}K_{f}tK_{r}$⟩. Winnick (1972: 163).

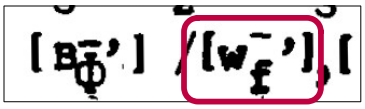


Figure 56. Subscript *f*. Kalnyn' & Maslennikova (1981: 396). Equivalent Cyrillic and Latin transcriptions. Figure 58 shows that the superscript minus sign does not need to appear directly over the subscript, and that this example can be digitized ⟨$в_{ф}^{-}$'/w_{f}^{-}'⟩.

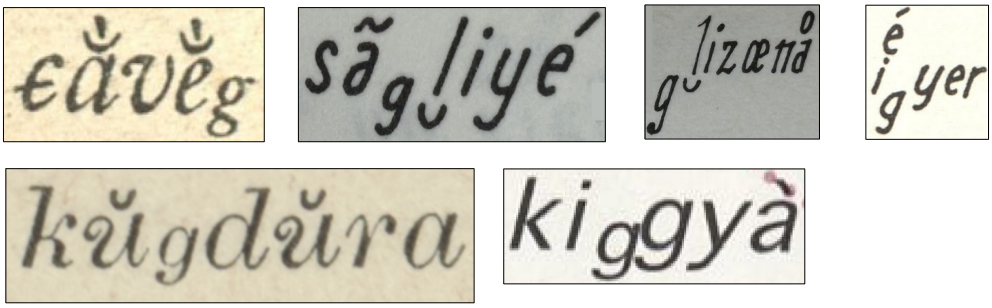


Figure 57. Subscript *g*. ① ALF map 1200 point 988, ③ ALO map 396 point 38, ④ ALAL map 74 point 17, ⑤ ALCe map 230 point 30, ⑥ TPh column 145 line 31, ⑦ ALN map 1410 point 30. This is a regular *g* set in italics, not a script *g*.

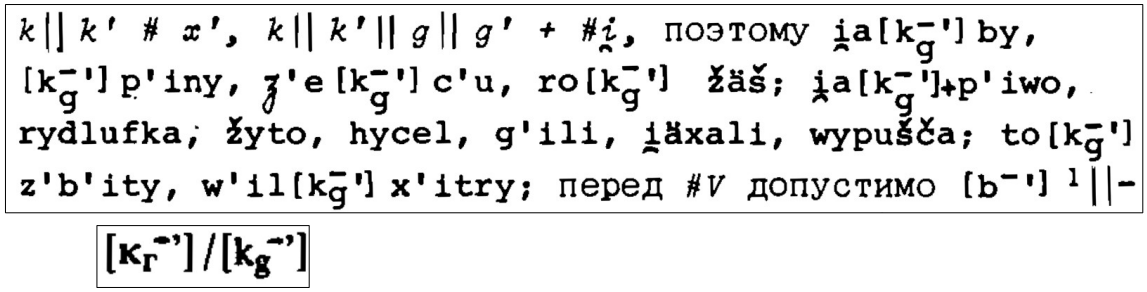


Figure 58. Subscript *g*. Kalnyn' & Maslennikova (1981: 350, 372). The second example shows (a) that there is no distinction between print ⟨$_{g}$⟩ and script ⟨$_{g}$⟩, and (b) that the C^{-}' superscript notation can be shifted to the right of the C_{g} subscript notation, and does not need to appear directly above it – that would appear to be merely a space-saving measure.

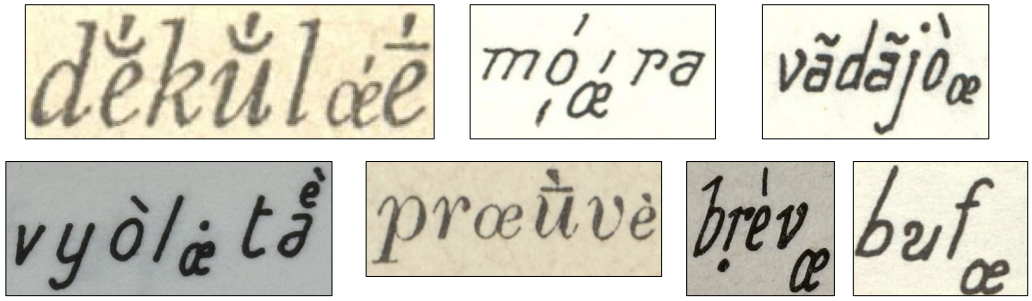


Figure 59. Subscript œ. ① ALF map 1138 point 275, ② ALJA map 440 point 49, ③ ALLy map 205 point 2, ④ALO map 362 point 84, ⑤ TPh column 318 line 5, ⑥ ALAL map 3 point 74, ⑦ ALCe map 18 point 65.

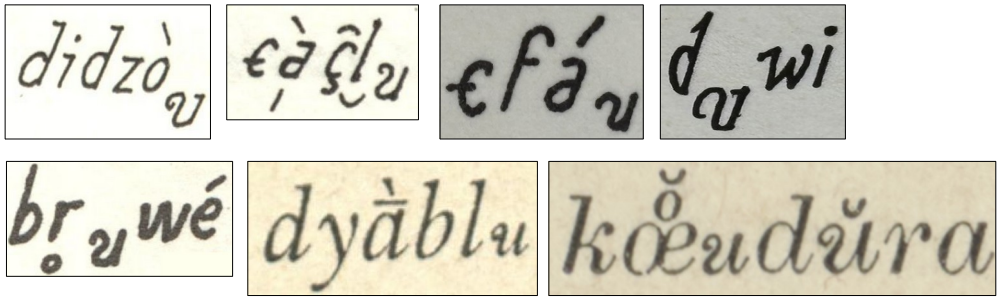


Figure 60. Subscript u. ① ALJA map 92 point 81, ② ALLy map 216 point 8, ③ ALO map 576 point 58, ④ ALAL map 118 point 70, ⑤ ALCe map 13 point 42, ⑥ ALF map 403 point 691, ⑦ TPh column 144 point 25.

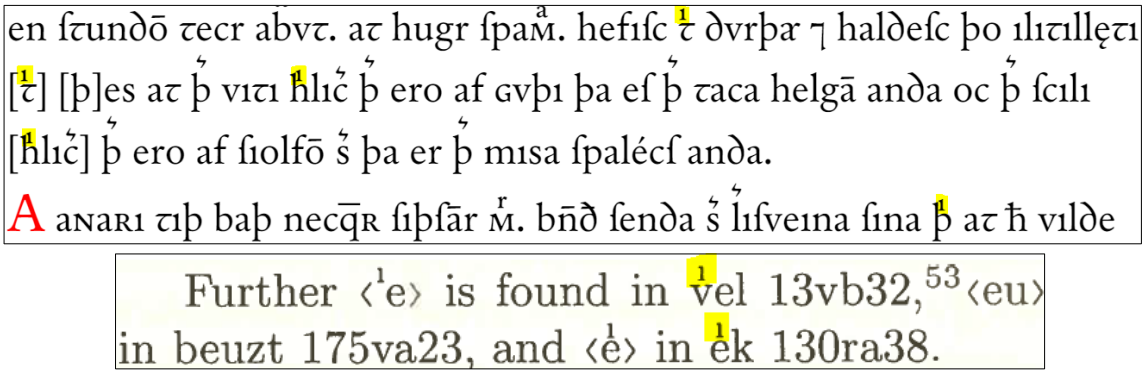


Figure 61. ① MENOTA p. 32v, ② van Weenen (2000: 68). Combining dotless *i* without a diacritic. Although the lack of a tittle is not semantically distinct, the proposed combining letter will be useful for diplomatic transcription.

Figure 62. The baseline letters *b*, *d* and *g* with a half stroke. ALPI *introducción*. These will be requested in an addendum to this proposal.

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646[23]
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*.

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 26*
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): *yes*
      Proposed name of script: *Combining Diacritical Marks Extended-A*
   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: *47*
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.

[23] Form number: N4502-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11, 2005-01, 2005-09, 2005-10, 2007-03, 2008-05, 2009-11, 2011-03, 2012-01)

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before? *yes*
   If YES explain *Requests for subscripts c, d, f, g were made in L2/21-043 and 21-207R with less evidence; some characters were proposed in L2/13-172 and its replacement L2/14-169r some examples were copied from them*
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: