

Unicode request for Romance dialectology symbols

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

2025 December 05

This proposal covers a family of phonetic alphabets used in nearly all French, Italian, and Spanish linguistic atlases¹ of the 20th century, and many of the 21st. These atlases form a vast dialect corpus that remains under-utilized by researchers due to a lack of accessibility. There are numerous initiatives to digitize these atlases,² but all encounter difficulties in rendering symbols missing from Unicode. Researchers are forced to devise a variety of *ad hoc* compromises, which seriously hinder new projects and complicate the comparison of data from existing ones.³ We quote two relevant comments:

‘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.

We propose adding the missing baseline symbols to Unicode to facilitate current and future digitization projects. Our proposal 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 work. In addition, some of our attestations are reproductions of his.

Because of discussion at SEW over whether Unicode should continue to encode combining letters such as those accepted for Teuthonista and Cyrillic (in the combining diacritical marks and Cyrillic extended blocks), or should instead decide that encoding those characters was a mistake, the combining letters found in these atlases have been split off as an addendum to this proposal.

We thank Giacomo Ferrieri and Graziano Tisato for their feedback on various aspects of this proposal, Matteo Rivoira for elucidating details of the Pellis-Genre alphabet, and 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.

We now introduce the phonetic alphabets in question.

¹A linguistic atlas typically contains a series of maps that show how speakers of local dialects responded to prompts such as *How do you say ‘sky’?* Traditionally the responses were transcribed *in situ* by the interviewer, in the phonetic alphabet of the atlas.

²See the *Digitization projects* section of the bibliography.

³Just as publishers in the era of metal type made sometimes inconsistent substitutions and hacks for unavailable symbols, so today publishers often make different choices when digitizing data that does not have Unicode support. This results in conflicting encodings of the same symbol (e.g. Œ, œ or medievalist Œ for original œ), which is a nuisance for dialectometric work. A similar problem occurs when one converts the data to IPA, because different authors make different and sometimes conflicting IPA approximations.

1. Rousselot-Gilliéron (RG) alphabet

The RG alphabet was developed in the late 19th century by Jean-Pierre Rousselot and Jules Gilliéron. The latter went on to use it in the *Atlas linguistique de le France* (ALF), the world's first linguistic atlas in cartographic form. This inspired virtually all French and Gallo-Romance dialectologists to adopt their alphabet, for instance Charles Bruneau for his *Enquête linguistique sur les patois d'Ardenne* (ELPA), Louis Gauchat *et al.* for their *Tableaux phonétiques des patois suisses romands* (TPh), and the CNRS (Centre national de la recherche scientifique) for a series of regional atlases: the ALAL, ALCe, ALJA, ALLy, ALN, ALO, ALPic, ALRé, and many others not cited here.

The most recently published of these atlases is the ALN, volume 5 of which appeared in 2019. The ALN is one of several atlases that had been only partly published when the CNRS abruptly halted its regional atlas program in the 1990s. Some of the abandoned material has since been published by others, but much still remains.⁴

The base letters of the RG alphabet 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.

RG: ⁶	a	b	c	ç	d	e	f	g	h	i	j	k	l	m	n	o	œ	p	r	s	t	u	ü	v	w	y	z
IPA: ⁷	ä	b	⁻⁸	ɸ	d	ɛ	f	g	h	i	ɹ	k	l	m	n	ɔ	ø	p	r	s	t	y	u	v	w	j	z

There is also a series of diacritics,⁹ most of which already exist in Unicode. The one exception is <̥>, which denotes partial nasalization (extIPA: [̥̃]) and is distinguished—as can be seen in Figure 1—from the tilde <̃>, which denotes full nasalization (IPA: [̃]).

A key feature of the RG alphabet, and of the others discussed below, is the use of combining letters to mark intermediate sounds, as explained in Figure A. For instance, <ũ> (IPA: [ũ]) denotes a vowel that is perceptually between [u] and [ó] (IPA: [u] and [o]). Similarly, <š> (IPA: [ʃ̺]) denotes a consonant with a degree of voicing between that of [s] and [z].

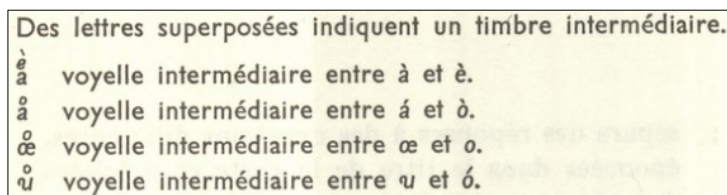


Figure A. ALJA 1: *introduction*. Explanation that combining letters indicate intermediate values.

⁴Brun-Trigaud 2016.

⁵By ‘base letters’ we mean those from which the rest may be built through combination (as in $g + d \rightarrow \textcircled{g}$) or the addition of diacritics. This is not an inventory of the basic *sounds* of the RG alphabet, which would include for instance [ɛ], [ŋ], [ʒ] (IPA: [ə], [ɲ], [θ]) and many others.

⁶There is an additional base letter <ɭ> in the TPh, which denotes a 'velar l' (IPA: [ɭ]) as it does also in the Böhmer-Ascoli and RFE alphabets discussed below.

⁷We base this analysis on Gilliéron & Rousselot 1887: 1–17, Gilliéron & Edmont 1902: 19, and on the descriptions in the introductions to the other cited works that use this alphabet. The IPA approximations that we provide, both here and in the rest of our proposal, are often made rather narrow to convey certain details, such as that RG <e> represents a mid vowel (IPA: [ɛ]), which is distinguished from the high-mid vowel <é> (IPA: [e]).

⁸ $\langle c \rangle$ never occurs in isolation, only in the combinations $\langle \tilde{c} \rangle$ and $\langle \tilde{\zeta} \rangle$ (IPA: $[\chi]$, $[\zeta]$). A similar situation holds for $\langle c \rangle$ in the other alphabets discussed below.

[illegible]

3. Pellis-Genre (PG) alphabet

The Pellis-Genre alphabet, as we call it, was devised in the early 20th century by Ugo Pellis—the main surveyor of the *Atlante Linguistico Italiano* (ALI)—and reformed in the 1970s by Arturo Genre, who resolved certain inconsistencies and integrated variant notations employed by later surveyors. In the published volumes of the ALI, there are the following base letters:¹⁹

PG: a b c d e ə f g h i j k l l̥ m n o p r z ʒ s ʃ t u v w y z ʒ ? α β γ δ ε ζ η θ λ λ̥ ρ σ φ χ ω
IPA:²⁰ ä b - d̥ ə f g h i j k l ɬ m n ɒ p r ð̥ θ̥ s z t u v w ý ts̥ dz̥ ʔ ä̃ β γ ð̃ ɛ̃ z̃ η̃ θ̃ ɬ̃ ɬ̃ ɾ̃ s̃ ɸ̃ x̃ ʃ̃

The diacritics²¹ are all found in Unicode. As in the BA alphabet, superscripts denote weakly articulated sounds and combining letters denote intermediate sounds.

In all, we propose the following characters to support the PG alphabet:

- Base letter: $\mathfrak{s}$ (figs. 6–7)
- Modifier letters: $\mathfrak{z}$ $\mathfrak{s}$ $\mathfrak{z}$ $\mathfrak{z}$ $\mathfrak{p}$ $\mathfrak{o}$ (figs. 23, 24, 12, 13, 14, 15)

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) alphabet

The RFE alphabet was devised in the early 20th century by Tomás Navarro Tomás for use in the academic journal after which it is named. Navarro Tomás went on to use this alphabet for his experimental *Atlas Lingüístico de Puerto Rico* (ALPR) and then for the monumental *Atlas Lingüístico de la Península Ibérica* (ALPI), where the following base letters appear:

RFE: a e b b̃ b̄ c d d̃ d̄ e ə f g g̃ ḡ h ħ i j k l l̃ l̄ m n ŋ ɲ o ɔ p r ɹ s t u ũ v w x y z θ φ
IPA:²² ä å ß ß̥ - ɖ ɗ ɗ̥ ɗ̄ ẽ f g ɣ ɣ̃ h ħ i j k l l̃ l̄ ɭ m n ŋ ɲ ɣ̥ ɔ̃ p r ɹ s t̥ ũ v w x̣ ɹ̣ z ʈ ɸ

The diacritics²³ are supported by Unicode, apart from <◌◌̥> and its turned variant <◌◌̥̥>, which indicate central (‘mixed’) vowels. These are unitary diacritics like <◌◌̥̥> and <◌◌̥̥̥̥>: a single dot is undefined on these vowel letters.

As in the BA and PG alphabets, superscript letters denote weakly articulated sounds, and intermediate sounds are denoted with combining letters.

The RFE alphabet was nearly universal in Spanish dialectology until the end of the 20th century. It continues to be used in some 21st-century publications such as the *Atlas Lingüístico y etnográfico de Castilla - La Mancha* (ALeCMan) and the *Cartografía lingüística de Extremadura* (CLEx).

Weak articulation is indicated with superscripts while combining letters are used to indicate intermediate phonetic values.

¹⁹The ALI substitutes some of the symbols in Genre (1978): ⟨Ǥ⟩ replaces ⟨Ǻ⟩, while ⟨ǣ⟩, ⟨ǟ⟩ replace ⟨ǣ̄⟩, ⟨ǣ̆⟩. The phonetic interpretations of ⟨ǣ⟩ and ⟨ǟ⟩ are not clear to us—perhaps IPA [ǣ] and [ǟ]; the surveyors were not entirely consistent with each other in how they used these two symbols. In addition, the published maps feature ⟨ǣ̄⟩, a symbol used by surveyors other than Pellis.

²⁰In 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.

[illegible]

²²We base this analysis on the descriptions provided in the introduction to volume I of the ALPI and in Navarro Tomás 2004.

[illegible]

In all, we propose the following characters to support the RFE alphabet:

- Combining diacritics: $\acute{\circ}$ $\grave{\circ}$ (figs. 4, 5)
- Base letters: $\mathfrak{b}$ $\mathfrak{d}$ $\mathfrak{g}$ $\mathfrak{l}$ (fig. 8, 9, 10, 11)
- Modifier letters: $\mathfrak{l}$ $\mathfrak{h}$ $\mathfrak{r}$ (figs. 19, 20, 23)
- Subscript letter: $\mathfrak{d}$ (fig. 37)

Distinct Unicode characters that function as allographs in these atlases

Glyphs that correspond to distinct Unicode characters may be allographic in the context of these atlases. For example, the symbol for IPA $[\eta]$ may be left-tail $\langle\eta\rangle$, right-tail $\langle\eta\rangle$, or Greek $\langle\eta\rangle$ with no difference in meaning (). This variation is not unique to these atlases: Greek η is well attested as a velar nasal in other source. A modifier $\langle\mathfrak{h}\rangle$ 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 modifier $\langle\mathfrak{h}\rangle$.



Figure B. ① AIS 553: 715 with straight-tail η , ② AIS 827: 715 with straight-tail η , ③ AIS phonetic symbol list, p 49, where η and η have strongly right-curved tails.

Idiosyncratically, Greek ξ in the original RG notation is replaced by ζ in the ALI. SEW recommends that these symbols be kept distinct; we therefore propose both the Genre form $\langle\xi\rangle$ (Figure 14) and the ALI form $\langle\xi\rangle$ (Figure 13).

Characters

- ∞ 1AF1 COMBINING INVERTED LAZY S. Figure 1 ff.
- ◌̣ 1AF2 COMBINING HOOK-ACUTE. Figure 3.
- ◌̇ 1AF4 COMBINING COLON ABOVE. Figure 4, 6.
- ◌̣̇ 1AF5 COMBINING COLON BELOW. Figure 5 ff.

Subscript letter

- u AB6F LATIN SUBSCRIPT SMALL LETTER U WITH LEFT HOOK. Figure 44.

[continue preceding subheading in Latin Extended-G]

- Ɔ 1DF8B LATIN SMALL LETTER REVERSED SCRIPT R. Figure 7 ff.

Letters with half stroke

- ḃ 1DF8C LATIN SMALL LETTER B WITH HALF STROKE. Figure 9.
- ḋ 1DF8D LATIN SMALL LETTER D WITH HALF STROKE. Figure 10.
- ḡ 1DF8E LATIN SMALL LETTER G WITH HALF STROKE. Figure 11.
- ḥ 1DF8F LATIN SMALL LETTER L WITH HALF STROKE. Figure 12.

Modifier letters

- ζ 1DFB1 MODIFIER LETTER SMALL ZETA. Figure 13.
- ξ 1DFB2 MODIFIER LETTER SMALL XI. Figure 14.
- ρ 1DFB3 MODIFIER LETTER SMALL RHO. Figure 15.
- σ 1DFB4 MODIFIER LETTER SMALL SIGMA. Figure 16.
- χ 1DFB5 MODIFIER SMALL LETTER CHI WITH STROKE. Figure 17.
- € 1DFB6 MODIFIER LETTER SMALL C WITH BAR. Figure 18.
- ł 1DFB7 MODIFIER LETTER SMALL L WITH HALF STROKE. Figure 19.
- ʒ 1DFB8 MODIFIER LETTER SMALL BARRED ENG. Figure 20.
- ƚ 1DFB9 MODIFIER SMALL LETTER R WITH STROKE. Figure 22 .
- Ʒ 1DFBA MODIFIER LETTER SMALL SCRIPT R. Figure 23.
- Ǝ 1DFBB MODIFIER LETTER SMALL REVERSED SCRIPT R. Figure 24.
- ʌ 1DFBC MODIFIER SMALL LETTER STIRRUP R. Figure 25.

Subscript letters

- € 1DFBD LATIN SUBSCRIPT SMALL LETTER C WITH BAR. Figure 29.
- ḋ 1DFBE LATIN SUBSCRIPT SMALL LETTER D WITH STROKE. Figure 37.
- œ 1DFBF LATIN SUBSCRIPT SMALL LIGATURE OE. Figure 43.
- ḃ 1DFC0 LATIN SUBSCRIPT SMALL LETTER B. Figure 26 ff.
- ḥ 1DFC1 LATIN SUBSCRIPT SMALL LETTER C. Figure 27, Figure 30 ff.
- ḋ 1DFC2 LATIN SUBSCRIPT SMALL LETTER D. Figure 27, Figure 34 ff.
- ḡ 1DFC3 LATIN SUBSCRIPT SMALL LETTER F. Figure 38 ff.
- ḥ 1DFC4 LATIN SUBSCRIPT SMALL LETTER G. Figure 41 ff.

Properties

1AF1;COMBINING INVERTED LAZY S;Mn;230;NSM;;;;;N;;;;;

1AF2;COMBINING HOOK-ACUTE;Mn;230;NSM;;;;;N;;;;;

1AF4;COMBINING COLON ABOVE;Mn;230;NSM;;;;;N;;;;;

1AF5;COMBINING COLON BELOW;Mn;230;NSM;;;;;N;;;;;

AB6F;LATIN SUBSCRIPT SMALL LETTER U WITH LEFT HOOK;Lm;0;L;<sub> AB52;;;;;N;;;;;

1DF8B;LATIN SMALL LETTER REVERSED SCRIPT R;Ll;0;L;;;;;N;;;;;

1DF8C;LATIN SMALL LETTER B WITH HALF STROKE;Ll;0;L;;;;;N;;;;;

1DF8D;LATIN SMALL LETTER D WITH HALF STROKE;Ll;0;L;;;;;N;;;;;

1DF8E;LATIN SMALL LETTER G WITH HALF STROKE;Ll;0;L;;;;;N;;;;;

1DF8F;LATIN SMALL LETTER L WITH HALF STROKE;Ll;0;L;;;;;N;;;;;

1DFB1;MODIFIER LETTER SMALL ZETA;Lm;0;L;<super> 03B6;;;;;N;;;;;

1DFB2;MODIFIER LETTER SMALL XI;Lm;0;L;<super> 03BE;;;;;N;;;;;

1DFB3;MODIFIER LETTER SMALL RHO;Lm;0;L;<super> 03C1;;;;;N;;;;;

1DFB4;MODIFIER LETTER SMALL SIGMA;Lm;0;L;<super> 03C3;;;;;N;;;;;

1DFB5;MODIFIER SMALL LETTER CHI WITH STROKE;Lm;0;L;<super> **1DF54**;;;;;N;;;;;

1DFB6;MODIFIER LETTER SMALL C WITH BAR;Lm;0;L;<super> A793;;;;;N;;;;;

1DFB7;MODIFIER LETTER SMALL L WITH HALF STROKE;Lm;0;L;<super> **1DF8C**;;;;;N;;;;;

1DFB8;MODIFIER LETTER SMALL BARRED ENG;Lm;0;L;<super> **1DF4F**;;;;;N;;;;;

1DFB9;MODIFIER SMALL LETTER R WITH STROKE;Lm;0;L;<super> 024D;;;;;N;;;;;

1DFBA;MODIFIER LETTER SMALL SCRIPT R;Lm;0;L;<super> AB4B;;;;;N;;;;;

1DFBB;MODIFIER LETTER SMALL REVERSED SCRIPT R;Lm;0;L;<super> **1DF8B**;;;;;N;;;;;

1DFBC;MODIFIER SMALL LETTER STIRRUP R;Lm;0;L;<super> AB45;;;;;N;;;;;

1DFBD;LATIN SUBSCRIPT SMALL LETTER C WITH BAR;Lm;0;L;<sub> A793;;;;;N;;;;;

1DFBE;LATIN SUBSCRIPT SMALL LETTER D WITH STROKE;Lm;0;L;<sub> 0111;;;;;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;;;;;

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

Annotations

Besides notes on look-alike characters, subscript *c* and *s* are attested with a combining cedilla; indeed, modifier *c* was adopted precisely so it could take a cedilla.

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

02E2 MODIFIER LETTER SMALL S

- supports 0327 combining cedilla

1AF1 COMBINING INVERTED LAZY S

→ 1DD1 COMBINING UR ABOVE

→ 1AB2 COMBINING INFINITY

1AF2 COMBINING HOOK AND ACUTE

→ 1FCE GREEK PSILI AND OXIA

1D9C MODIFIER LETTER SMALL C

- supports 0327 combining cedilla

209B LATIN SUBSCRIPT SMALL LETTER S

- supports 0327 combining cedilla

1DFC1 LATIN SUBSCRIPT SMALL LETTER C

- supports 0327 combining cedilla

Charts

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	◌̿̎	◌̿̏	◌̿̐	◌̿̑	

Latin Extended-E
AB30 **AB60**

	AB3	AB4	AB5	AB6
0	Ǽ	Ǿ	ǡ	ǣ
1	Ǽ	Ǿ	ǡ	ǣ
2	Ǽ	Ǿ	ǡ	ǣ
3	Ǽ	Ǿ	ǡ	ǣ
4	Ǽ	Ǿ	ǡ	ǣ
5	f	h	χ	Ω
6	g	R	x	dz
7	†	r	χ	ts
8	‡	r	χ	‡
9	‡	‡	χ	^
A	na	‡	y	†
B	na	‡	‡	†
C	na	‡	h	‡
D	v	‡	†	‡
E	ø	u	†	U
F	ø	u	u	u

Latin Extended-G

1DF00

1DFFF

	1DF0	1DF1	1DF2	1DF3	1DF4	1DF5	1DF6	1DF7	1DF8	1DF9	1DFA	1DFB	1DFC	1DFD	1DFE	1DFF
0	ƒ	ɣ	ɔ̥	ɔ̥	Ȧ	ɹ	ω	ɿ	Ȧ	ɹ			b	ɣ	ɔ̥	ɿ
1	ɹ	ɿ	ɔ̥	ɔ̥	Ȧ	ɹ	ω	ɿ	Ȧ	ɹ		ɹ	c	ɹ	ɔ̥	ɿ
2	ɔ̥	ɔ̥	ɔ̥	ɔ̥	Ȧ	ɹ	ω	ɿ	Ȧ	ɹ		ɹ	d	ɔ̥	ɔ̥	ɿ
3	ɔ̥	ɔ̥	ɔ̥	ɔ̥	Ȧ	ɹ	ω	ɿ	Ȧ	ɹ		ɹ	f	ɔ̥	ɔ̥	ɿ
4	ɔ̥	ɔ̥	ɔ̥	ɔ̥	Ȧ	ɹ	ω	ɿ	Ȧ	ɹ		ɹ	g	ɔ̥	ɔ̥	ɿ
5	ɔ̥	ɔ̥	ɔ̥	ɔ̥	Ȧ	ɹ	ω	ɿ	Ȧ	ɹ		ɹ	h	ɔ̥	ɔ̥	ɿ
6	ɔ̥	ɔ̥	ɔ̥	ɔ̥	Ȧ	ɹ	ω	ɿ	Ȧ	ɹ		ɹ	i	ɔ̥	ɔ̥	ɿ
7	ɔ̥	ɔ̥	ɔ̥	ɔ̥	Ȧ	ɹ	ω	ɿ	Ȧ	ɹ		ɹ	j	ɔ̥	ɔ̥	ɿ
8	ɔ̥	ɔ̥	ɔ̥	ɔ̥	Ȧ	ɹ	ω	ɿ	Ȧ	ɹ		ɹ	k	ɔ̥	ɔ̥	ɿ
9	ɔ̥	ɔ̥	ɔ̥	ɔ̥	Ȧ	ɹ	ω	ɿ	Ȧ	ɹ		ɹ	l	ɔ̥	ɔ̥	ɿ
A	ɔ̥	ɔ̥	ɔ̥	ɔ̥	Ȧ	ɹ	ω	ɿ	Ȧ	ɹ		ɹ	m	ɔ̥	ɔ̥	ɿ
B	ɔ̥	ɔ̥	ɔ̥	ɔ̥	Ȧ	ɹ	ω	ɿ	Ȧ	ɹ		ɹ	n	ɔ̥	ɔ̥	ɿ
C	ɔ̥	ɔ̥	ɔ̥	ɔ̥	Ȧ	ɹ	ω	ɿ	Ȧ	ɹ		ɹ	o	ɔ̥	ɔ̥	ɿ
D	ɔ̥	ɔ̥	ɔ̥	ɔ̥	Ȧ	ɹ	ω	ɿ	Ȧ	ɹ		ɹ	p	ɔ̥	ɔ̥	ɿ
E	ɔ̥	ɔ̥	ɔ̥	ɔ̥	Ȧ	ɹ	ω	ɿ	Ȧ	ɹ		ɹ	q	ɔ̥	ɔ̥	ɿ
F	ɔ̥	ɔ̥	ɔ̥	ɔ̥	Ȧ	ɹ	ω	ɿ	Ȧ	ɹ		ɹ	r	ɔ̥	ɔ̥	ɿ

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: Hans GOEBL *et al.* (1998) *Atlant linguistich di ladin dolomitich y di dialec vejins*, 1^a pert, 7 vols. Wiesbaden: Reichert. (2012) *Atlant linguistich dl ladin dolomitich y di dialec vejins*, 2^a pert, 7 vols. 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://ucl.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 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/>
- CLEX: González SALGADO & José ANTONIO. *Cartografía lingüística de Extremadura*. <https://geolectos.com/>

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

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.

SUF: Sabine KRÄMER-NEUBERT & Norbert Richard WOLF (2005–2009) *Sprachatlas von Unterfranken*. Heidelberg: Universitätsverlag Winter. <https://regionalsprache.de/suf.aspx>

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 Dialecti 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 Apprendre* 35: 47–60. <https://anr-appi.univ-lille.fr/>

Silviu-Ioan BEJINARIU & Florin-Teodor OLARIU (2017) Romanian Linguistic Atlases in Digital Format: A New Approach. *Philologica Jassyensia* 13 (1): 13–23.

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.cld.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/>
- Joao SARAMAGO (2006) O Atlas linguístico-etnográfico de Portugal e da Galiza (ALEPG). *Estudis Romànics* 28: 281–298.
- 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 (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: 58–144, 206–265.
—— (1984) *Middle Chinese: A Study in Historical Phonology*. Vancouver: UBC Press.
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>

Figures²⁴

ã, õ, ẽ, etc., voyelles à nasalisation complète.
ã, õ, ẽ, etc., voyelles à nasalisation incomplète

Figure 1. ALJA 1: introduction, explaining the difference between <õ> and <ẽ>.



Figure 2. Combining ẽ versus combining õ. ① ALJA map 73 points 82 and 71, ② ALLy map 56 points 56 and 48, ③ ALPic map 82 points 15 and 14 (closed allograph of ẽ), ④ TPh column 177 points 6 and 1, ⑤ AIS map 842 points 121 and 941, ⑥ ALN map 1458 point 2 (reversed allograph of ẽ found in vol. 5), ⑦ ALN map 823 point 1, ⑧ ALAL map 9 points 30 and 5, ⑨ ALF map 9 points 694 and 691.



Figure 3. Combining ẽ. ① AIS map 889 point 544, ② ALD 2 map 10 point 81, ③ VIVALDI «l'angelo» (Ceppomorelli), ④ ALEIC map 2 point 4.

²⁴In addition to the aforementioned atlases, we cite some German (SBS, SUF), Italian (ALEIC), and Romanian atlases (ALR-sn, NALR-B, NALR-MB, NALR-O) that use closely related phonetic alphabets. In some cases we also provide supporting examples from other linguistic publications.

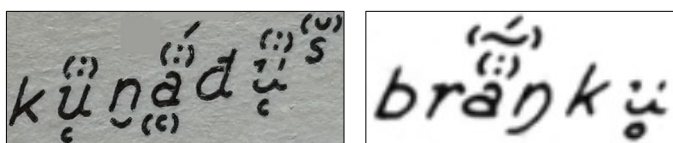


Figure 4. Combining colon above, ¨. [ALPI map 61](#) point 279 and [map 25](#) point 241. The colon indicates a single feature that can be attenuated with parentheses but cannot be subdivided the way ʷ is separated from ʷ in the example at right.

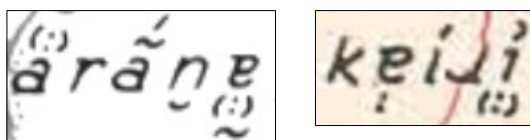


Figure 5. Combining colon below with and without parentheses, ̣ and ̣̣. [ALPI map 16](#) point 143 and [map 31](#) point 272.

ô	mixta, vocal central semicerrada labializada.
ö	vocal anterior media labializada abierta, con mixta (̣) como en port. <i>casa</i> ,

Figure 6. ALPI, introduction. <ô> and <ö> indicated central and front equivalents of <o>. *<ô> with a single dot is undefined. Similarly, <̣> is a reduced central vowel, and <̣̣> is not defined.

Baseline letters

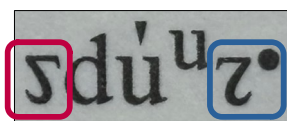


Figure 7. Reversed z. ALI map 100 point 933, contrasting reversed <̣> (red) and Unicode-supported <z> (blue). See also Figure 8.

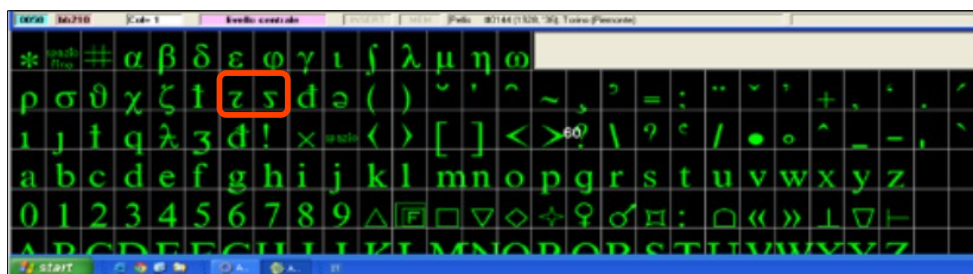


Figure 8. Cerruti *et al.* (2025: 41). Contrast between <z> and <̣> on an ALI keyboard.

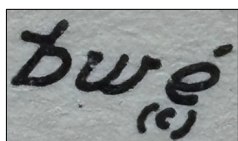


Figure 9. B with half stroke, b. ALPI [map 28 point 354](#).

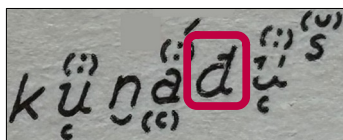


Figure 10. D with half stroke, d. ALPI [map 61 point 279](#).

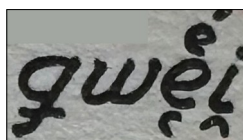
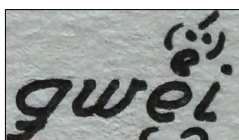


Figure 11. G with half stroke, g, with normal g for comparison. ALPI [map 28 point 402](#).

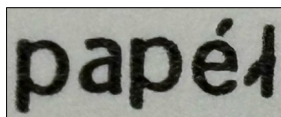
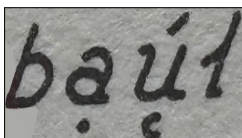


Figure 12. L with half stroke, l. ① ALPI [map 24 point 615](#), ② ALMEx map 81 point 189.

Modifier Greek letters

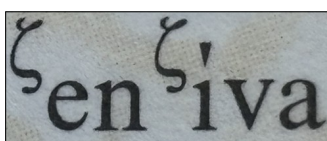


Figure 13. Modifier ζ. ALI map 32 point 349.

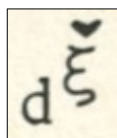


Figure 14. Modifier ξ. Genre (1978: 71). <ζ> and <ξ> are allographic in the atlases; they have the same phonetic value and no atlas contrasts them. However, the SEW recommends that they be encoded separately.

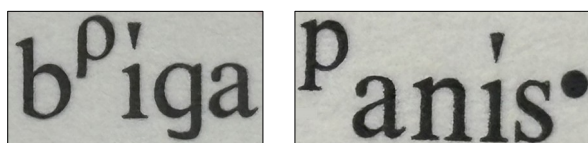


Figure 15. Modifier ρ . ALI map II point 94, with modifier p from point 1001 (caption) on the same map for comparison.

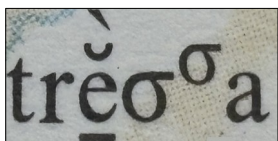


Figure 16. Modifier σ . ALI map 14 point 446.

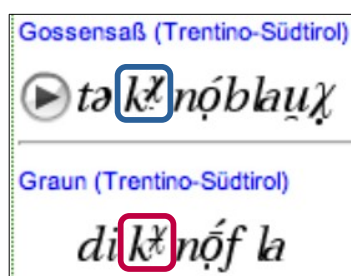


Figure 17. Modifier χ . VIVALDI «l'aglio». Modifier χ (red) contrasts with modifier χ (blue).

Modifier Latin letters

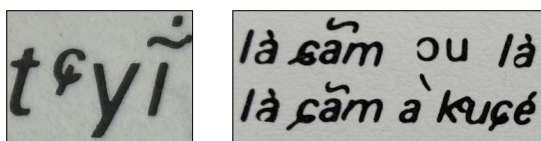


Figure 18. Modifier ϵ . ① ALPic [map 188 point 20](#). The stroke is set at an angle; at right are several tokens from ② ALRé map 252 (caption), showing similar variation in bar positioning in baseline ϵ .

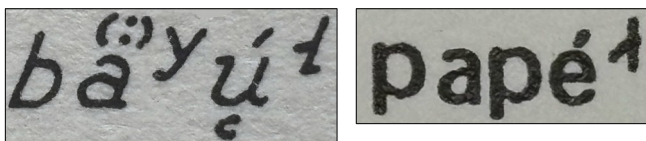


Figure 19. Modifier l . ① ALPI [map 24 point 135](#), ② ALMEx map 81 point 183.

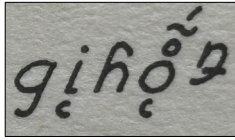


Figure 20. Modifier η . ALPI [map 11 point 519](#).



Figure 21. Combining η . ALD 2 [map 154 point 177](#).

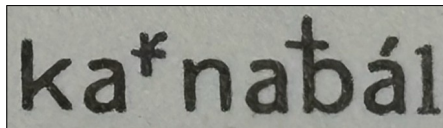


Figure 22. Modifier τ . ALMEx map 56 point 185.

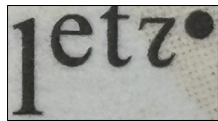


Figure 23. Modifier τ . ALI map 1 point 943.



Figure 24. Modifier reversed τ . ALI map 1 point 46 – the same map as for modifier τ above.

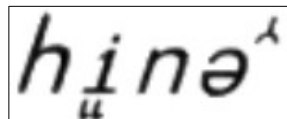


Figure 25. Modifier λ . SUF map 75.

Subscript Latin letters

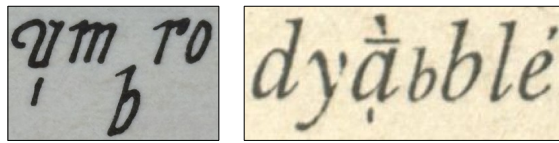


Figure 26. Subscript b. ① ALAL map 10 point 20, ② ALF [map 403 point 699](#). The first generation of atlases set ‘subscript’ letters in a smaller type size, as at right, but the modern convention is to use true subscripts, as at left, and that is how we suggest digitizing the data.¹⁰

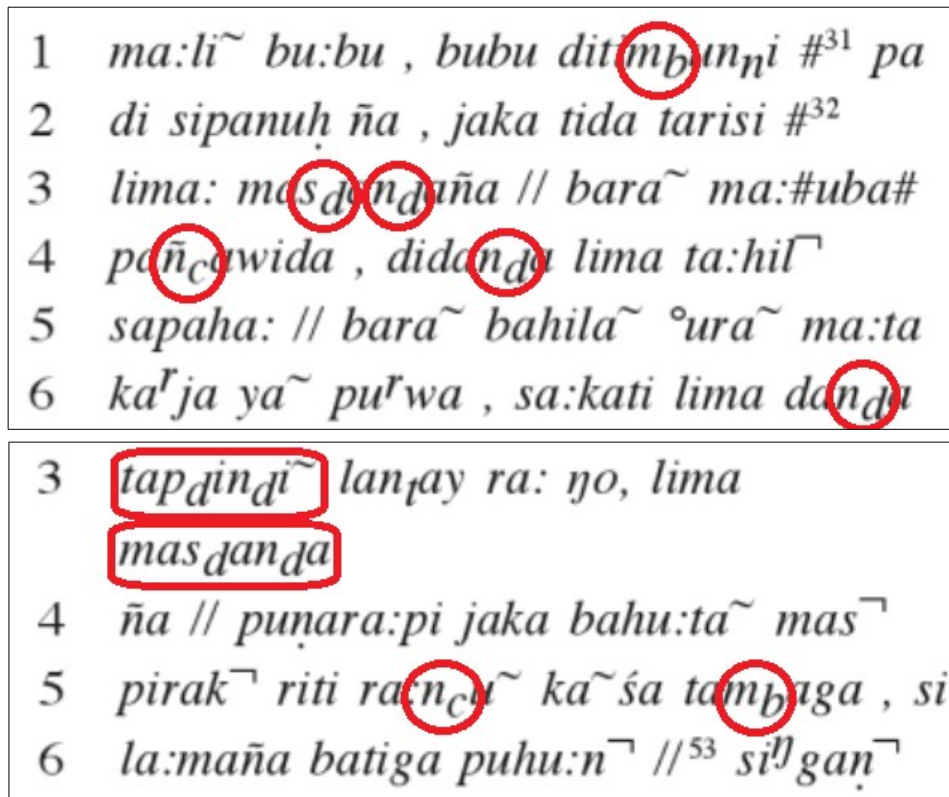


Figure 27. Subscript b, c and d. Kozok (2015: 69, 71). Conjuncts transliterated as subscripts. They commonly occur after nasals, but note <s_d, p_d>. Subscripting in the diplomatic transliteration is not predictable from the normalized transliteration. For example, in Kozok (2015: P/5/6), normalized *geggah rabutti rampassi* is diplomatic <gaggah rabut_iirampas_i>, with neither the second g in *gg* nor the p in *mp* subscript.

[п̣̌'] / [p̣̌'],

Figure 28. Subscript *b*. Kalnyn' & Maslennikova (1981: 396). Equivalent Cyrillic and Latin transcriptions. Cyrillic is already supported. Figure 42 shows that the superscript minus sign does not need to appear directly over the subscript, and that this can be digitized <п̣̌'/p̣̌'>.

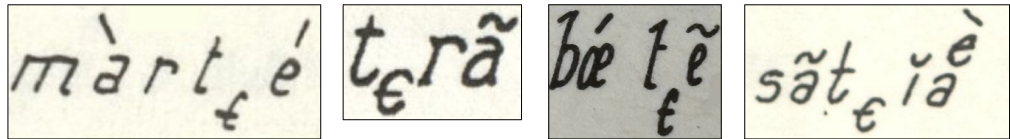


Figure 29. Subscript *e*. ① ALLy [map 914 point 31](#), ② ALN map 823 point 3, ③ ALAL map 3 point 53, ④ ALCe map 1344 point 30.

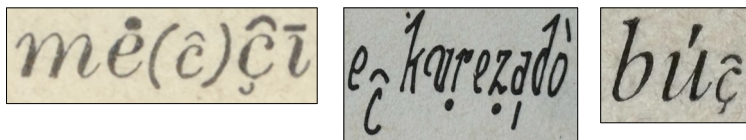


Figure 30. Subscript *c* with turned breve and cedilla. ① TPh [column 381 line 13](#), ② ALAL map 42 point 27, ③ ELPA questionnaire 38. As is the case with superscript <ç>, a subscript <ç> should be encoded as subscript *c* plus a combining cedilla.

The following consonants are found in my list of words :

h. k, x, q. y. n ; t ; s, c, **t_ç** ts, tc. m, p. l, l'.

qā'aqāas	qat _ç 'ěqāas	my sister's husband.
taxō'olxul'	tatsxō'olxul'	axe. t_ç = dento-alveolar t.
ět_ç 'ě'xaa	ōot!'ō'qxaa	house.

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

-ēs, my. -ěts, thy. -qēs, his. -tcuks, her.

-t_çux, our. -stětc, your. -tcāas, their.

For instance : taxō'lxl'ēs—my axe. hē'něětcuks—her father.

hē'nēt_çux—our father.

Figure 31. Subscript *c* with cedilla. Boas (1892: 37 fn, 38, 39, 41).

kw'e[c']en', ko[c']oŋ - šmu[t_c]ek, s'w'a[t_c]eŋko > tn, tž,
 tž, t' || tn', f kw'e[t_c]n'u, smu[t_c]ny, smu[t_c]n'eŋ,
 ko[t_c]ŋy, f ko[t_c]la; [z']en', gar[z']oŋak, bru[z']i -

Figure 32. Subscript c. Kalnyn' & Maslennikova (1981: 360). The alignment of the apostrophe needs to be handled by the font; U+0315 should work.

h_p - S₁K
 ç cɪzɪr 'sizzling'

Figure 33. Subscript c with cedilla. Winnick (1972: 163). This should be encoded as subscript c plus a combining cedilla.

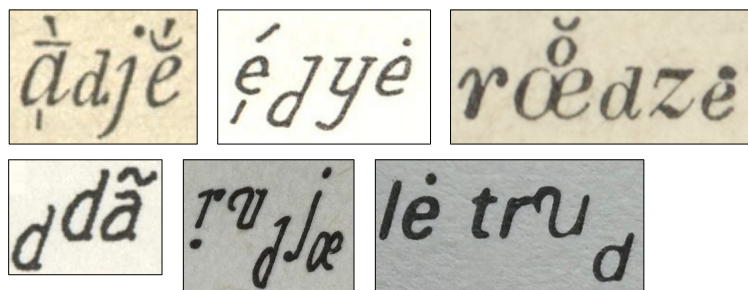


Figure 34. Subscript d. ① ALF [map 9 point 794](#), ② ALJA map 135 point 41, ③ TPh [column 248 line 17](#), ④ ALN map 783 point 21, ⑤ ALAL map 8 point 33, ⑥ ALRé map 8 point 37.

меняемость [t]¹ || [t_d] [d]¹ || [t_d] (t, d — символы перечислен-
 ных 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]) ин-
 терпретируется словоформа, содержащая перед РП глухой

Figure 35. Subscript d. Kalnyn' (1973: 327). Illustrations specific to Ukrainian are presented in Cyrillic, <т>.

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

Figure 36. 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*>.

resulta casi muda. Para representar esta variante empleamos el signo *ɖ*: *libertad*-*libɛrtáɖ*, *huésped*-*wéspəɖ*, *bondad*-*bonɖáɖ*, *virtud*-*birtúɖ*, *venid*-*beníɖ*, *esperad*-*esperáɖ*, *traed*-*traéɖ*, *callad*-*kaláɖ*. En pronunciación culta esta *ɖ* puede convertirse en *ɗ* y aun, a veces, en *ɗ̣*.

«Andaría *usted*¹ cerca de la *verdaɖ*¹, si todas esas cosas me entusiasmaran a ratos, o en los libros, o vistas desde mi casa, muy *arrellanado* en el sillón; pero *usted*¹ sabe muy bien que no hay faena de labranza ni entretenimiento *honrado* aquí, en

Figure 37. Subscript *ɗ*. Navarro Tomás (2004: 79–80). At bottom of the top clip it says, 'In educated pronunciation, this *ɖ* can become *ɗ* and even sometimes *ɗ̣*.'

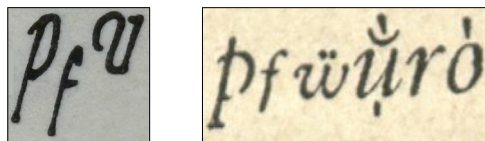


Figure 38. Subscript *f*. ① ALAL map 317 point 11, ② ALF [map 1683 point 805](#).

$h_{p^i} - h_{K_f K_r}$	<i>püfür</i>	'breezy'
$h_{p^i} - h_{K_f K_r}$	<i>küfür</i>	'coolness'

Figure 39. Subscript *f* in <^hK_fLK_r>. Winnick (1972: 163).

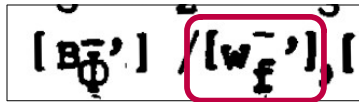


Figure 40. Subscript *f*. Kalnyn' & Maslennikova (1981: 396). Equivalent Cyrillic and Latin transcriptions. Figure 42 shows that the superscript minus sign does not need to appear directly over the subscript, so this example can be digitized <в̣̣'/ẉ̣'>.

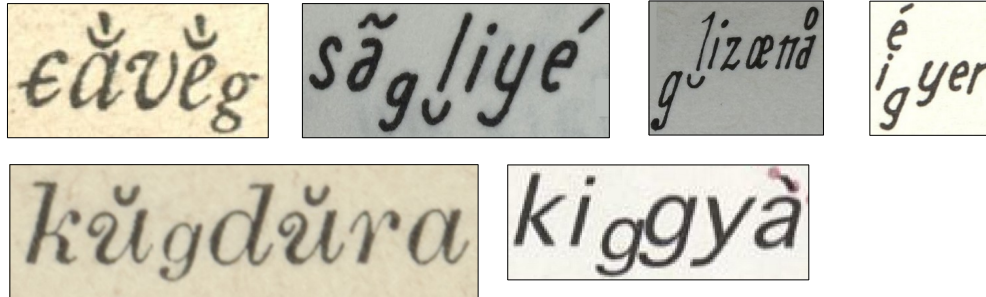


Figure 41. Subscript *g*. ① ALF [map 1200 point 988](#), ③ ALO map 396 point 38, ④ ALAL map 74 point 17, ⑤ ALCe map 230 point 30, ⑥ TPh [column 145 line 31](#), ⑦ ALN map 1410 point 30. This is a regular *g* set in italics, not a script *g*.

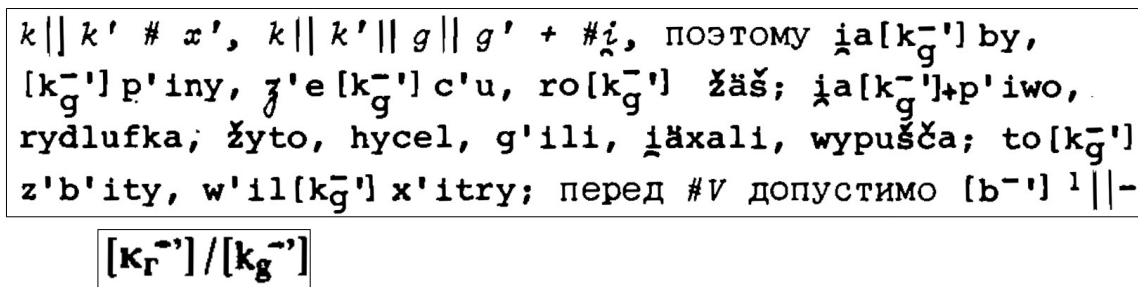


Figure 42. Subscript *g*. Kalnyn' & Maslennikova (1981: 350, 372). The second example shows (a) that there is no distinction between print <g> and script <g>, and (b) that the C' superscript notation can be shifted to the right of the C_g subscript notation, and does not need to appear directly above it – such kerning would appear to be merely a space-saving measure.

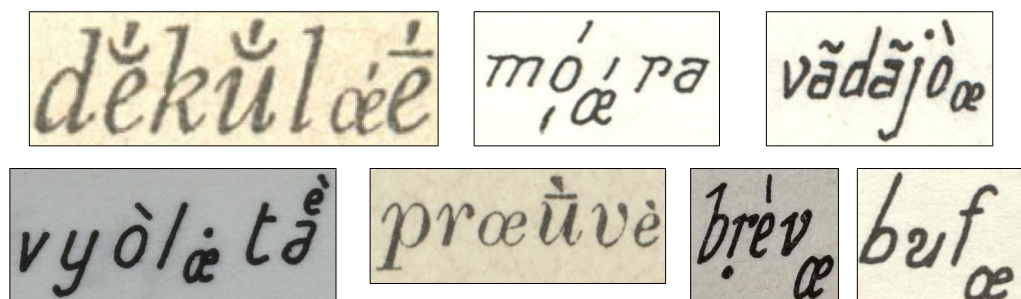


Figure 43. Subscript æ. ① ALF [map 1138 point 275](#), ② ALJA map 440 point 49, ③ ALLy [map 205 point 2](#), ④ ALO map 362 point 84, ⑤ TPh [column 318 line 5](#), ⑥ ALAL map 3 point 74, ⑦ ALCe map 18 point 65.

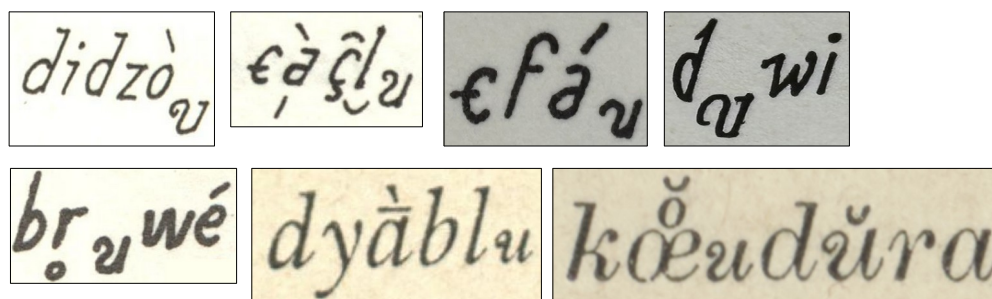


Figure 44. Subscript u. ① ALJA map 92 point 81, ② ALLy [map 216 point 8](#), ③ ALO map 576 point 58, ④ ALAL map 118 point 70, ⑤ ALCe map 13 point 42, ⑥ ALF [map 403 point 691](#), ⑦ TPh [column 145 point 25](#).

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646²⁵

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:	<i>Romance dialectology symbols</i>
2. Requester's name:	<i>Nicolas Datua and Kirk Miller</i>
3. Requester type (Member body/Liaison/Individual contribution):	<i>individual</i>
4. Submission date:	<i>2025 November 13</i>
5. Requester's reference (if applicable):	
6. Choose one of the following:	
This is a complete proposal:	<i>yes</i>
(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):	<i>no</i>
Proposed name of script:	
b. The proposal is for addition of character(s) to an existing block:	<i>yes</i>
Name of the existing block:	<i>Combining Diacritical Marks Extended, Latin Extended-E, Latin Extended-G</i>
2. Number of characters in proposal:	<i>30</i>
3. Proposed category (select one from below - see section 2.2 of P&P document):	
A-Contemporary ☒	B.1-Specialized (small collection) ☐
C-Major extinct ☐	B.2-Specialized (large collection) ☐
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?	<i>yes</i>
a. If YES, are the names in accordance with the “character naming guidelines” in Annex L of P&P document?	<i>yes</i>
b. Are the character shapes attached in a legible form suitable for review?	<i>yes</i>
5. Fonts related:	
a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?	<i>Kirk Miller</i>
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):	<i>SIL (Gentium Release)</i>
6. References:	
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?	<i>yes</i>
b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?	<i>yes</i>
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)?	<i>no</i>

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.

²⁵ 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 <i>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</i>	
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? <i>Authors are members of the user community.</i>	
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:	