

Unicode request to accommodate diacritics below turned *i*

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

2025 November 15

This addendum to our proposal for Romance dialectology symbols covers two problematic characters: a dotless turned *i* and its modifier equivalent, which may be needed if a soft-dotted property cannot be applied to the existing characters U+1D09 and U+1D4E. Turned *i* is used with diacritics below in the RFE alphabet, where turned *a e i* indicate centralized vowels.

In Unicode properties issue #493 *Soft-dotted property for ! U+1D09 and ! U+1D4E*, we suggested that the **Soft_Dotted** property should be applied to U+1D09 LATIN SMALL LETTER TURNED I (!) and U+1D4E MODIFIER LETTER SMALL TURNED I (!), to handle combining diacritics below that suppress the tittle. However, the comments on GitHub noted that doing so would require an implementation to know whether the tittle was above or below in order to determine whether a particular combining mark should suppress it. A better approach would therefore be to create a new **Soft_Dotted_Below** property that would apply specifically to these two characters. Unfortunately, in the opinions of the commentators, this would make algorithms such as context-specific casing overly complex. (See [Table 3-17](#) in the Core Spec.) Without a property to handle the replacement of the tittle, dedicated dotless characters with non-canonical equivalences would be needed.

If this were done, a dotless ! followed by a combining dot below would not be canonically equivalent to !. The UTC might consider whether the suggested **Soft_Dotted_Below** property would be a better solution. If it is decided that separate dotless characters would be preferable, they are requested here.

Characters

- ! 1DF97 LATIN SMALL LETTER TURNED DOTLESS I. Figure 1.
- ! 1DFAA MODIFIER LETTER SMALL TURNED DOTLESS I. Figure 2.

Properties

1DF97;LATIN SMALL LETTER TURNED DOTLESS I;LI;0;L;;;;N;;;;
 1DFAA;MODIFIER LETTER SMALL TURNED DOTLESS I;Lm;0;L;<super> 1DF97;;;;N;;;;

DoNotEmit data

A dotless ! followed by a combining dot below is not canonically equivalent to !.

1DF97 0323; 1D09; Precomposed_Form # LATIN SMALL LETTER TURNED DOTLESS I, COMBINING DOT BELOW; LATIN SMALL LETTER TURNED I
 1DFAA 0323; 1D4E; Precomposed_Form # MODIFIER LETTER SMALL TURNED DOTLESS I,
 COMBINING DOT BELOW; MODIFIER LETTER SMALL TURNED I

Annotations

1DF97;LATIN SMALL LETTER TURNED DOTLESS I

- stands in for 1D09 when that needs to be soft-dotted

1DFAA;MODIFIER LETTER SMALL TURNED DOTLESS I

- stands in for 1D4E when that needs to be soft-dotted

References

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>

Charts

Suggested characters are highlighted. Greyed-out characters are requested in the main proposal.

Latin Extended-G

1DF00

1DFFF

	1DF00	1DF1	1DF2	1DF3	1DF4	1DF5	1DF6	1DF7	1DF8	1DF9	1DFA	1DFB	1DFC	1DFD	1DFE	1DFF
0	ſſ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ
1	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ
2	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ
3	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ
4	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ
5	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ
6	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ
7	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ
8	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ
9	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ
A	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ
B	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ
C	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ
D	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ
E	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ
F	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ	ꝑ

Figures

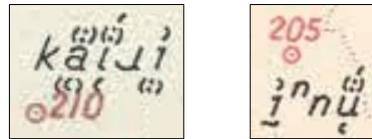


Figure 1. Turned *i*. (1) ALPI [map 31 point 210](#), (2) ALPI [map 66 point 205](#). In (1), ignoring the extra acute accent, the third letter is *i*. The last letter is this same *i* rotated, with the open mark above and the colon below; the rotation can also be seen in the placement of the terminal. The soft-dotted behaviour of the two letters is the same. (2) is similar, but with a tilde replacing the tittle in a rotated *i*.

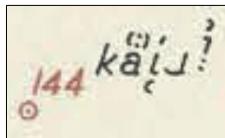


Figure 2. Modifier turned *i*. ALPI [map 31 point 144](#). The diacritic below is a unitary colon, not a second dot: a single dot diacritic does not occur below turned vowels.

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646¹

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>Accommodating diacritics below turned I</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 15</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>Latin Extended-G</i>	
2. Number of characters in proposal:	<i>2</i>	
3. Proposed category (select one from below – see section 2.2 of P&P document):		
A-Contemporary <input checked="" type="checkbox"/>	B.1-Specialized (small collection) <input type="checkbox"/>	B.2-Specialized (large collection) <input type="checkbox"/>
C-Major extinct <input type="checkbox"/>	D-Attested extinct <input type="checkbox"/>	E-Minor extinct <input type="checkbox"/>
F-Archaic Hieroglyphic or Ideographic <input type="checkbox"/>	G-Obscure or questionable usage symbols <input type="checkbox"/>	
4. Is a repertoire including character names provided?		
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.

¹ 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?	<input type="checkbox"/> <i>no</i>
If YES explain	<input type="text"/>
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.)?	<input type="checkbox"/> <i>yes</i>
If YES, with whom?	<input type="text"/> <i>Authors are members of the user community.</i>
If YES, available relevant documents:	<input type="text"/>
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?	<input type="checkbox"/>
Reference:	<input type="text"/>
4. The context of use for the proposed characters (type of use; common or rare)	<input type="checkbox"/> <i>transcription</i>
Reference:	<input type="text"/>
5. Are the proposed characters in current use by the user community?	<input type="checkbox"/> <i>yes</i>
If YES, where? Reference:	<input type="text"/>
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?	<input type="checkbox"/> <i>no</i>
If YES, is a rationale provided?	<input type="checkbox"/>
If YES, reference:	<input type="text"/>
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	<input type="checkbox"/> <i>no</i>
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?	<input type="checkbox"/> <i>no</i>
If YES, is a rationale for its inclusion provided?	<input type="checkbox"/>
If YES, reference:	<input type="text"/>
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?	<input type="checkbox"/> <i>no</i>
If YES, is a rationale for its inclusion provided?	<input type="checkbox"/>
If YES, reference:	<input type="text"/>
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?	<input type="checkbox"/> <i>no</i>
If YES, is a rationale for its inclusion provided?	<input type="checkbox"/>
If YES, reference:	<input type="text"/>
11. Does the proposal include use of combining characters and/or use of composite sequences?	<input type="checkbox"/> <i>no</i>
If YES, is a rationale for such use provided?	<input type="checkbox"/>
If YES, reference:	<input type="text"/>
Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?	<input type="checkbox"/> <i>no</i>
If YES, reference:	<input type="text"/>
12. Does the proposal contain characters with any special properties such as control function or similar semantics?	<input type="checkbox"/> <i>no</i>
If YES, describe in detail (include attachment if necessary)	<input type="text"/>
13. Does the proposal contain any Ideographic compatibility characters?	<input type="checkbox"/> <i>no</i>
If YES, are the equivalent corresponding unified ideographic characters identified?	<input type="checkbox"/>
If YES, reference:	<input type="text"/>