

Unicode request for two Latin orthographic pairs derived from “I i”

Ismael Robles Haloui, independent researcher in psycholinguistics

<arbadac.arba65[at]gmail.com>

2025-09-24

1. Introduction

This document requests the addition of two casing pairs derived from the letter *I i*. If the Unicode Consortium accepts their encoding, the following four glyphs will eventually exist:

U+A7DE	ɨ	LATIN CAPITAL LETTER I WITH OBLIQUE STROKE
U+A7DF	ɨ̇	LATIN SMALL LETTER I WITH OBLIQUE STROKE
U+A7E0	Ꝛ	LATIN CAPITAL LETTER I WITH DESCENDER
U+A7E1	ꝛ	LATIN SMALL LETTER DOTLESS I WITH DESCENDER

2. I with oblique stroke

The letter ***i* with oblique stroke** (ɨ ɨ̇) is known from the following contexts:

- Abbreviation of *inter* in certain medieval Latin manuscripts (Lindsay, 1915; Cappelli, 1929)
- The practical orthography for the Tlahuica/Ocuiltec language (*pjyɛkakjó*, ISO 639-3: ocu) of Mexico, where it represents [i] (Muntze & Martínez, 2015, p. 379)

In the context of Tlahuica, due to the current lack of Unicode support, this proposed pair often appears as an ordinary *I i* with a superimposed slash (/) in the original materials. As a result, in said materials, the diacritic usually overlaps with much of the body of the basic letter *I i*, and in the lowercase form, it may even touch the dot (especially in italic text), making the letter highly challenging to read.

Consequently, the eventual UTC-approved form should take the oblique-stroke form found in Medievalist usage. Indeed, this form is attested at least once in the Tlahuica source materials (see Fig. 6).

3. (Dotless) I with descender

The letter ***i* with descender** (Ꝛ ꝛ; with dotless lowercase) was used from 1929 to 1939 in the Latinized alphabet of the Khakas language (Xakac tɨni), where it represented [ə]. Throughout

this period, the self-denomination of the language was written “Xakas t̡l̡l̡”. The inclusion of this letter in the Unicode standard will benefit Turkologists, and quite likely the general philological community as well.

This letter is unusual in that its lowercase lacked the expected dot (hence the proposed name LATIN SMALL LETTER DOTLESS I WITH DESCENDER for the lowercase).

Crucial note

The second requested pair (*I* / with descender) displays a noticeable foot between the stem and the descender in many of the source materials, likely due to typesetting issues derived from the narrowness of the letter *I* / . In those cases, the resulting display might resemble a derivative of *L* *l* rather than of *I* / . However, the UTC-approved glyph should instead look more clearly like a derivative of *I* / , as in the following sample (see also Fig. 15, which does not feature the foot either):

I l

Properties

A7DE;LATIN CAPITAL LETTER I WITH OBLIQUE STROKE;Lu;0;L;;;;;N;;;A7DF;

A7DF;LATIN SMALL LETTER I WITH OBLIQUE STROKE;Ll;0;L;;;;;N;;;A7DE;;
A7DE

A7E0;LATIN CAPITAL LETTER I WITH DESCENDER;Lu;0;L;;;;;N;;;A7E1;

A7E1;LATIN SMALL LETTER DOTLESS I WITH DESCENDER;Ll;0;L;;;;;N;;;A7E0;;
A7E0

Annotations

The following annotations are recommended:

A7DE LATIN CAPITAL LETTER I WITH OBLIQUE STROKE

A7DF LATIN SMALL LETTER I WITH OBLIQUE STROKE

* Medievalist usage, Tlahuica

A7E0 LATIN CAPITAL LETTER I WITH DESCENDER

A7E1 LATIN SMALL LETTER DOTLESS I WITH DESCENDER

* Khakassian

References

- Baskakov, N. A., & Inkižekova-Grekul, A. I. (1953). *Chakasskij jazyk* (p. 368). Retrieved from https://altaica.ru/LIBRARY/BASKAKOV/khakas_grammar.pdf
- Cappelli, A. (1929). *Dizionario Di Abbreviature* (p. 28). Retrieved from <https://archive.org/details/CappelliDizionarioDiAbbreviature/page/168/mode/2up>
- Dirección Lingüística (2023, January 5). *Relato tlahuica "Lindi benzopy tumi"* [video]. YouTube. <https://www.youtube.com/watch?v=T2eUsWc6Sgs>
- Jušmanov, N. V. (1941). *Opredelitelj jazykov* (p. 368). Akademija Nauk SSSR. Retrieved from <https://www.booksite.ru/fulltext/173314/text.pdf>
- Kipling, R. (1937). *Rikki-Tikki-Tavi* (T. Baltъzakov, trans.). Ogiz Хъзыл-Сар Кragiz-нъг Хакассиядашъ Filialъ. Retrieved from <https://nbdrx.ru/pdf/bx0000498.pdf> (Original work published 1894)
- Lindsay, W. M. (1915). *Notae latinae: an account of abbreviation in Latin mss. of the early minuscule period (c. 700-850)* (p. 496). Cambridge University Press.
- Martínez Ortega, A. P. (2012). *Clases verbales, transitividad y valencia verbal en el pɣyɛkakjó, tlahuica de San Juan Atzingo* (p. 2) [thesis]. El Colegio de México. Retrieved from <https://repositorio.colmex.mx/concern/theses/hh63sw15j?locale=es>
- Muntzel, M. C., & Martínez, A. (2015). El alfabeto práctico pɣyɛkakjo. *Estudios De Cultura Otopame*, 9(01), 375–387. <https://revistas.unam.mx/index.php/eco/article/view/52051>
- Muntzel, M. C., & González Nestor, N. (n.d.). Archivo de lenguas - tlahuica. Retrieved from https://www.academia.edu/38563507/Archivo_de_lenguas_tlahuica
- Samrin, K. K. (1934). *Bukvar*.
- Unidentified author (1935). *Irgi de* (N. L. Kokov, trans.). Kraj OGIZ-тъг Хакасташъ Filialъ.

Figures

<i>id est</i>	i ÷ or ·i·	id̄t id̄st id̄s	id̄	id̄
<i>ideo</i>	—	id̄o	—	—
<i>Iesus</i>	ihs, iħc	iħs	iħs	iħs
<i>igitur</i>	iḡ ig' g'	—	—	—
<i>inter</i>	†	—	—	—

Fig. 1. Capital I (=inter) in Lindsay (1915).

I	I
J. (i) Iacobus, - Iesus (abbr. giur.) XIV	† (i) id est XIII
L. (i) Istituzioni (abbr. giur.)	† (i) id est XV
J. (i) Iurisperi-tus XV m.	† (i) id est XIII
J. (i) Item XIV m.	† (i) inest XIII
† (i) in XII	† (i) inter VIII
† (i) institutus VIII	† (i) unica, - illa, - prima XIV
† (i) ibi XV	† (i) Iacobus XIV
† (i) idest XIV	† (i) unius XIV p.

Fig. 2. Cappelli (1929).

A don Quirino Barreto (†)

Por su amor, su compromiso y su entrega a la lengua y la cultura tlahuica, por su hospitalidad y su generosidad en cada clase, en cada narración, en cada oración, en cada palabra.

Litutijmbia mulajñi ndixokikj napjə naktjo jo ndixokikj muty† napjə pjyɛkakjo.
Tondomuteñu'u metpøky por ndyetso temikitundo'umtikj.

Fig. 3. Tlahuica-language acknowledgments in Martínez Ortega (2012).

- (1) A, E, I, O, U, Ø, †B, K, CH, D, F, G, J, L, M, N, Ñ, P, R, S, T, W, X, Y, TS, DS, ?, BY, DY, LY, PY, TY, KY, XY.

Fig. 4. Muntzel & Martínez (2015, p. 376), with capital †.

Alfabeto práctico	AFI	Palabra en tlahuica	Traducción al español
u	[u]	<i>nku</i>	‘lagartija’
ē	[ə]	<i>ndyē</i>	‘flor’
ī	[i]	<i>ñirī</i>	‘iglesia’
ø	[ʌ]	<i>nmøtoo</i>	‘tejón’
ū	[ʊ]	<i>nimūly</i>	‘zorrito’

El tlahuica, en su inventario fonético, tiene nueve vocales, cada una de las cuales presenta a su vez alargamiento (Martínez 2010: 4). Es decir, en total son 18 vocales. A continuación, se presenta un listado de los sonidos vocálicos hallados sin tomar en cuenta las vocales con alargamiento (Martínez 2010:4). Se incluye la representación de los sonidos vocálicos por medio del alfabeto *pjyēkakjō*:

Cuadro 2. Descripción articulatoria de las vocales del tlahuica

AFI	Alfabeto <i>pjyēkakjō</i>	Descripción articulatoria
[i]	< i >	cerrado, anterior, no labializado
[i]	< ī >	cerrado, central, no labializado
[u]	< u >	cerrado, posterior, labializado

Fig. 5. Description of Tlahuica vowels in Muntzel & Martínez (2015, p. 378), featuring lowercase *ī* alongside three other vowel letters with stroke: *ē* [ə], *ø* [ʌ], and *ū* [ʊ]. Note that the diacritic on *ī* and *ū* appears as a slash (/) superimposed on the main letters, due to the lack of font support. (Of these two letters, *ū* *ū*, also found in the genealogically related Mazahua, was added as part of Unicode 11.0 in 2018, i.e. 3 years after the original Tlahuica work was released.)

(7) Ejemplos con <ī>

<ntyī>	‘hilo de lana’
<tindy>	‘blando’
<tañixty>	‘se quema’

Fig. 6. Muntzel & Martínez (2015, p. 381). In this variant, which aligns with the Medievalist glyph, the stroke appears noticeably shorter, making the letter much more legible. Thus, this variant is proposed as the future UTC-approved glyph.



Fig. 7. Dirección Lingüística (2023). Here, ï is evidently a stand-in for the unavailable ɨ.

2.1. Relato acerca del Chupamirto

(Recopilado el 11 de agosto de 2004 en pjyɛkakjo por doña Natividad González Nestor)

"Lindi mbendo tumi"

1. ndimɔjnɔ tɔɔ ɲejosuerte kakjo.
2. waɪ mityeby t'aki mitye pɔki t'api.
3. ɲechu ndem t t ndimaha ndityeby.
4. pɔki t'apy techituhɔ.
5. tatyebɪ pɔki t'apy jo tetutewily
6. jo mula tut'umty litsutsily mbla nlindi
7. bebarril jo ɲukiljɲu michinyɔ kitlewily
8. lit'apy xa ɲututsitikj
9. tujɔ pitukjo'ijnɔjo p'aki
10. michinyɔ lewily t'apy

Fig. 8. A transcription of a Tlahuica folk tale about the hummingbird, in Muntzel & González-Nestor (n.d., p. 41).

M 'o kikjondoklo
 [ʔo kikhondoklo]
 /ʔo ki-kho- nto- klo/
 cómo 2S- MAN-RESP- estar
 ¿cómo está?

ya mbo mas ndok'íhik'
 [ya mbo mas ndokʔičikʔ]
 /ya mpo mas nto- kʔiči -kʔ/
 si ya más RESP- sentir -2S
 si ya se siente más mejor

N wetunatsi tu'ijñikj
 [wetúnáɕi tuʔihnikh]
 /we- tú- náɕi tu- ʔihní -kh/
 CLAS- RESP- señora 3S.PAS- decir -1S
 La señora me dijo

ndimø notutu'íhik'
 [ndimʌ notutuʔihikʔ]
 nti- mʌ no- tutu- ʔili- kʔ/
 3S- decir NEG- 3S- lastimar -2S
 dice que no te opere/ te lastime

Fig. 9. Muntzel & González-Nestor (n.d., p. 50).

G: ¡Ah! kimikistlo Bet'í chindzø t'akua
 [a kimikistloβet'i čindʒʌ t'akwa]
 /a kimi-ki-stlo βet'i čindʒʌ t'akwa/
 INTERJ PAS.IMP-2S-estar arriba Loma hoy no
 ¿Fuieste a la Loma hoy, no?

Fig. 10. Muntzel & González-Nestor (n.d., p. 53).



Fig. 11. Samrin (1934, p. 53). ᐃ ᐃ (print and script forms), alongside a text.

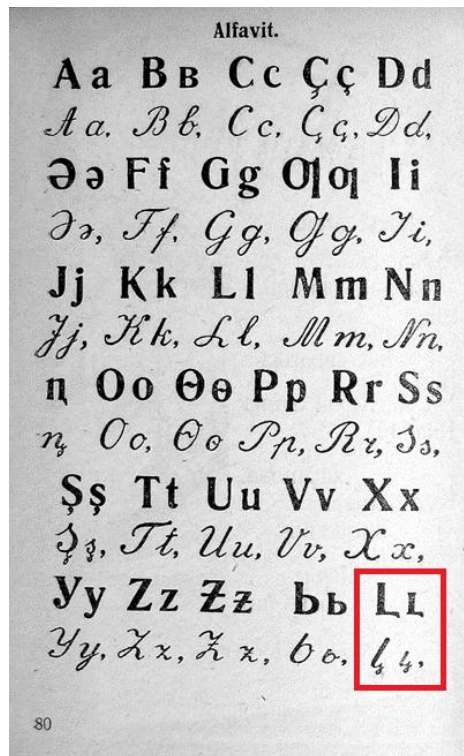


Fig. 12. Khakas Latin alphabet, in Samrin (1934, p. 80), with I i (last letter). The cursive form displays a much shorter foot than in print form.



Fig. 13. Capital I (highlighted in yellow) in a cropped section of the cover of "Irglde"; from an unidentified author (1935).

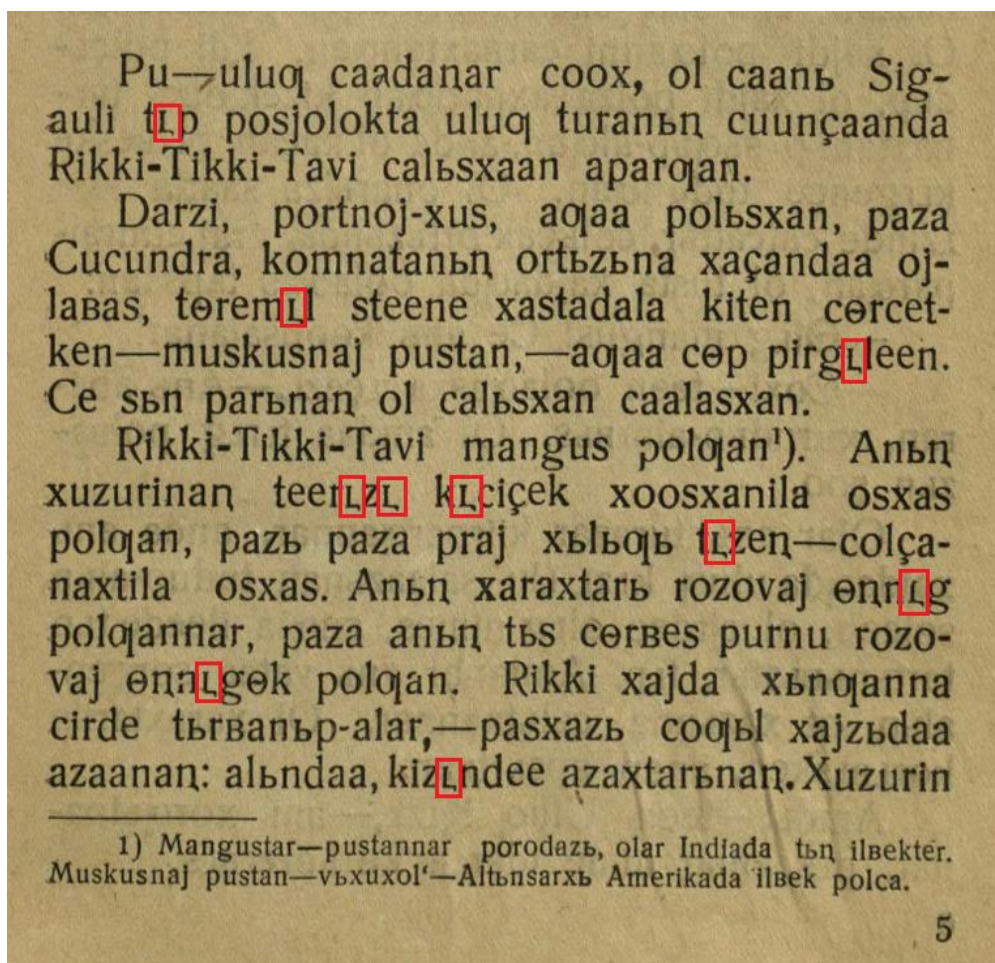


Fig. 14. Multiple instances of lowercase ɿ, in Kipling (1894/1937, p. 5).

ɿ	хакаский
ɿ	тувинский
ɿ	кавказские, курдский
ɿ	в адыгейский
ɿ	адыг., кабардинский
ɿ	авар., ительм., хант., эским.
ɿ	абаз., авар., адыг., кабард.
ɿ	хантыйский
ɿ	ингуш., тюркские, ⁴ чеченский
ɿ	языки народов Севера

Fig. 15. A section from Jušmanov (1941), featuring lowercase ɿ. This form, lacking the foot, is proposed for Unicode.

Таблица алфавитов

Современный алфавит 1939 г.	Латинизированный алфавит 1929–1939 гг.	Первый хакасский алфавит 1924–1929 гг. на основе миссионерского алфавита 1893–1899 гг.	Академический алфавит транскрипции 1885–1907 гг.	Названия букв	Значение знаков и соответствие их русским звукам
А а	А а	А а	А а	а	а
Б б	В в	Б б	Б б	бэ	б
В в	У у	В в	— —	вэ	в
Г г	Г г	Г г	Г г	гэ	г
— Ғ (1946)	— ОҒ	— h (1927)	— Ҫ	Ҫы	см. § 4
Д д	Д д	Д д	Д д	дэ	д
Е е	Э э, Е е*	Е е	Е е	е	е
Ж ж	З з	Ж ж (1924)	Ж ж	жэ	ж
З з	З з	З з	З з	зэ	з
И и	И и	И и	И и	и	и
І і	І і	І і	І і	и с точкой	см. § 4
Й й	Ј ј	Й й	— ј	й	й
К к	К к	К к	К к	ка	к
Л л	Л л	Л л	Л л, І	эл	л
М м	М м	М м	М м	эм	м
Н н	Н н	Н н	Н н	эн	н
— нь	— ң	— н (1924)	— ң	эң	см. § 4
О о	О о	О о	О о	о	о
Ө ө	Ө ө	Ө ө	Ө ө	ө	см. § 4
П п	Р р	П п	П п	пэ	п
Р р	Р р	Р р	Р р	эр	р
С с	С с	С с	С с	эс	с
Т т	Т т	Т т	Т т	тэ	т

* До 1936 г. вместо буквы Е е в хакасском алфавите тот же звук обозначался буквой Э э.

Fig. 16. Baskakov & Inkižekova-Grekul (1953). Note that нь was used instead of modern ң from 1939 until 1962.

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 <http://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 <http://std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html>.

See also <http://std.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html> for latest Roadmaps.

A. Administrative

1. Title:	<i>Proposal to encode two Latin orthographic pairs derived from "I i"</i>		
2. Requester's name:	<i>Ismael Robles</i>		
3. Requester type (Member body/Liaison/Individual contribution):	<i>Individual contribution</i>		
4. Submission date:	<i>2025-8-1</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-D</i>		
2. Number of characters in proposal:			<i>4</i>
3. Proposed category (select one from below - see section 2.2 of P&P document):			
A-Contemporary	<i>A</i>	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?			
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>Ismael Robles Haloui</i>		
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):	<i>Google (Noto Sans)</i>		
6. References:			
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?			
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 <http://www.unicode.org> for such information on other scripts. Also see Unicode Character Database (<http://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? If YES explain	No
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.)? If YES, with whom? If YES, available relevant documents:	No
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Reference:	Medievalists, Mexicanists, Turkicists
4. The context of use for the proposed characters (type of use; common or rare) Reference:	Orthographical
5. Are the proposed characters in current use by the user community? If YES, where? Reference:	Yes
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? If YES, is a rationale provided? If YES, reference:	Yes Yes To keep them with associated characters
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	Yes
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? If YES, is a rationale for its inclusion provided? If YES, reference:	No
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? If YES, is a rationale for its inclusion provided? If YES, reference:	No
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? If YES, is a rationale for its inclusion provided? If YES, reference:	No
11. Does the proposal include use of combining characters and/or use of composite sequences? 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? If YES, reference:	No
12. Does the proposal contain characters with any special properties such as control function or similar semantics? If YES, describe in detail (include attachment if necessary)	No
13. Does the proposal contain any Ideographic compatibility characters? If YES, are the equivalent corresponding unified ideographic characters identified? If YES, reference:	No