L2/26-176

Unicode request for Latin dotless i with descender

Ismael Robles Haloui, independent researcher in psycholinguistics

<arbadac.arba65[at]gmail.com>

2026-06-28

The letter ***i* with descender** (Ɨ ɨ; with dotless lowercase), used from 1929 to 1939 in the Latinized alphabet of the Khakas language, where it represented [ɘ] and was one of the most common letters.

In modern times, the inclusion of this letter in the Unicode standard will benefit Turkologists, philologists, and linguistics scholars when citing and/or digitizing Latinized-era Khakas texts, and sometimes illustrating the old orthgoraphy. (Of note, Khakas-language literature and writing saw an unprecedented period of flourishing throughout the Latinized era.)

In many of the source materials, the letter displays a noticeable base separating the stem and the descender, as if it were derived from capital *L*. This was likely due to typesetting issues derived from the narrowness of the letter *I ı*. Indeed, the script form of Ɨ ɨ lacks the base outright (see Figure 2).

Now, of course, the UTC-approved glyph should instead look more clearly like a derivative of *I ı*, as in the following sample (see also Figures 5 and 6, which do not feature the base either):

Ɨ ɨ

Collation

The proposed collation order is derived from Samrin (1934), cited in Figure 2:

a << A < в << B < c << C < ç << Ç < d << D < ə (e) << Ə (E) < f << F < g << G < ŋ << Ŋ < i << I < j << J < k << K < l << L < m << M < n << N < ŋ << Ŋ < o << O < ɵ << Ɵ < p << P < r << R < s << S < ş << Ş < t << T < u << U < v << V < x << X < y << Y < z << Z < ƶ << Ƶ < ь << Ь < ɨ << Ɨ

Notes:

1. Unusually, ꞅ was sorted after ь, instead of after *i*. This was likely for phonological reasons; in Khakas, the front variant of /ɯ/ is /ə/, *not* /i/.
2. *Ə ə* (1929–1935) and *E e* (1935–1939) were two ways to write the same vowel, /e/.

Properties

```
A7FE;LATIN CAPITAL LETTER I WITH DESCENDER;Lu;0;L;;;;;N;;;;A7FF;
```

```
A7FF;LATIN SMALL LETTER DOTLESS I WITH DESCENDER;Ll;0;L;;;;;N;;;A7FE;;
A7FE
```

Annotations

The following annotations are recommended:

```
@ Additions for Khakas
```

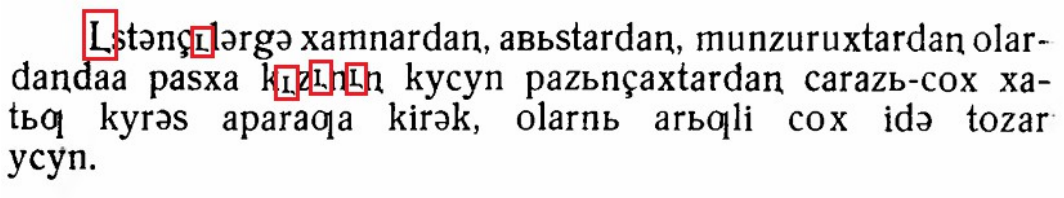
```
A7FE  LATIN CAPITAL LETTER I WITH DESCENDER
```

```
A7FF  LATIN SMALL LETTER DOTLESS I WITH DESCENDER
```

References

- Aəşin, Ơl., & Samrin, K. (1932). *Kolxonьx ycyn*. OGIZ. https://nbdrx.ru/pdf/bx0000224.pdf
- Ager, S. (2025, July 23). *Khakas language, alphabet and pronunciation*. Omniglot. Retrieved October 9, 2025 from https://www.omniglot.com/writing/khakas.htm
- Gladkov, F. (1935). *Ciit Xorвь* (N. Kokov, trans.). OGIZ. https://nbdrx.ru/pdf/bx0000503.pdf
- Jušmanov, N. V. (1941). *Opredelitelj jazykov*. Akademija Nauk SSSR. https://www.booksite.ru/fulltext/173314/text.pdf
- KNAB / Place Names Database (2012, September 25). *Hakassi / Khakass / Хакас Hakas*. Eesti Keele Instituut. https://arhiiv.eki.ee/knab/lat/kblkjh.pdf
- Samrin, K. K. (1934). *Bukvar*. OGIZ.
- Vukotic, O. (1936). *Toƣьnçьçax (direktiva): pɨr pөlɨktɨg p'jesa (M. Kol'tsoftьŋ fel'jeton temazьna)* (Je. Nacinova, trans.). OGIZ. https://nbdrx.ru/pdf/bx0000500.pdf

Figures



Lştənçılərgə xamnardaŋ, aвьstardaŋ, munzuruxtardaŋ olardaŋdaa pasxa kılzılnıŋ kycyn pazьnçaxtardaŋ carazь-cox xatьqı kyrəs aparaqa kirək, olarnь arьqli cox idə tozar ycyn.

Fig. 1. Aəşin & Samrin (1932, p. 63).

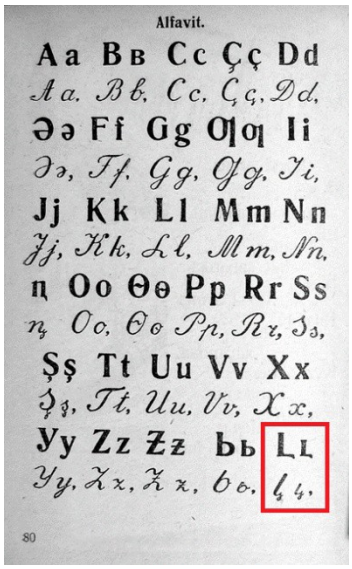
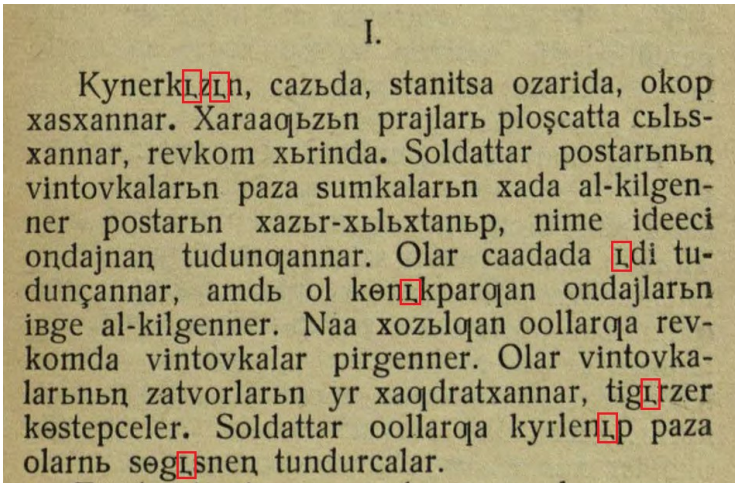


Fig. 2. Samrin (1934, p. 80). Khakas alphabet (print and script forms), with Ị ị as the last letter.



I.

Kynerkılzın, cazьda, stanitsa ozarida, okop xasxannar. Xaraaqьzьn prajlarь ploşcatta cьlьsxannar, revkom xьrinda. Soldattar postarьnьŋ vintovkalarьn paza sumkalarьn xada al-kilgenner postarьn xazьr-xьlьxtanьp, nime ideeci oŋdajnaŋ tudunqannar. Olar caadada ıdi tudunçannar, amdь ol kənlkparqan oŋdajlarьn ibge al-kilgenner. Naa xozьlqan oollarqa revkomda vintovkalar pirgenner. Olar vintovkalarьnьŋ zatvorlarьn yr xaqdratxannar, tigırzer kəstepceler. Soldattar oollarqa kyrlenıp paza olarnь səglsneŋ tundurcalar.

Fig. 3. Gladkov (1935, p. 3).

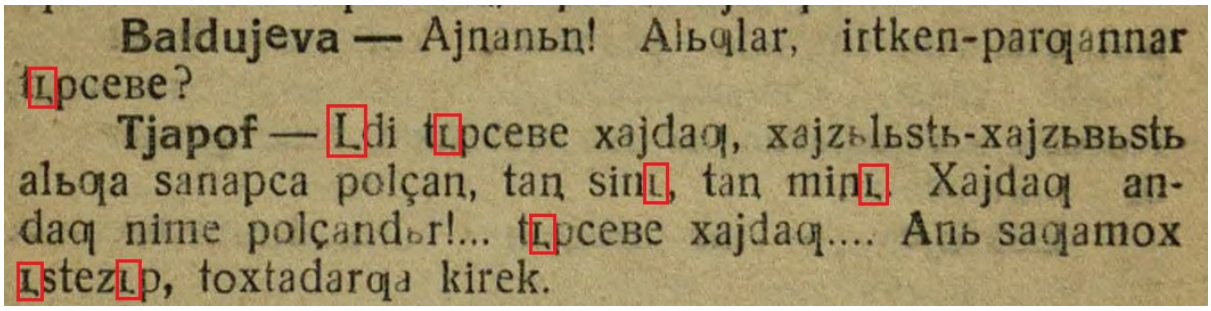

Baldujeva — Ajnanьn! Alьɋlar, irtken-parɋannar tɭpceвe?
Tjapof — ɭdi tɭpceвe xajdaɋ, xajzьlьstь-xajzьвьstь alьɋa sanapca polçan, taŋ sinɭ, taŋ minɭ. Xajdaɋ andaɋ nime polçandьr!... tɭpceвe xajdaɋ.... Anь saɋamox ɭstezɭp, toxtadarɋa kirek.

Fig. 4. Vukotic (1936, p. 9).

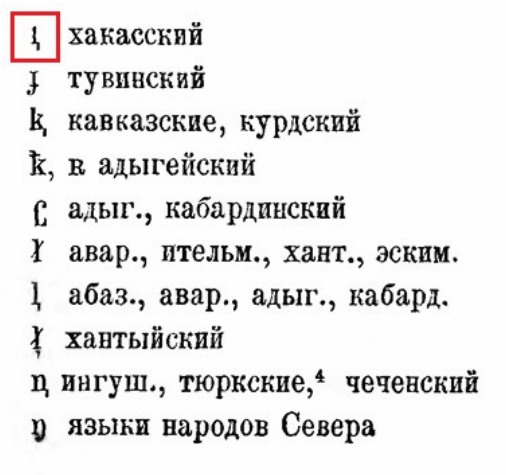

ɩ хакасский
ɟ тувинский
ƙ кавказские, курдский
ƙ, в адыгейский
ɾ адыг., кабардинский
ɣ авар., ительм., хант., эским.
ɭ абаз., авар., адыг., кабард.
ɣ хантыйский
ɳ ингуш., тюркские,[4] чеченский
ŋ языки народов Севера

Fig. 5. Jušmanov (1941, p. 368), featuring lowercase ɩ. This form, lacking the base, is proposed for Unicode. See also Figure 6.

Ladina tähestik 1929–1939				**Roman alphabet** 1929–1939			
A a		G g		L l		P p	
B в		Ɋ ɋ	= ğ	M m		R r	
C c	= č	I i		N n		S s	
D d		Ɪ ɪ	= ï	ɳ ɳ	= ŋ	Ş ş	= š
Ə ə	= e	J j		O o		T t	
F f		K k		Ɵ ɵ	= ö	U u	

Fig. 6. KNAB (2012). This form likewise lacks the base.

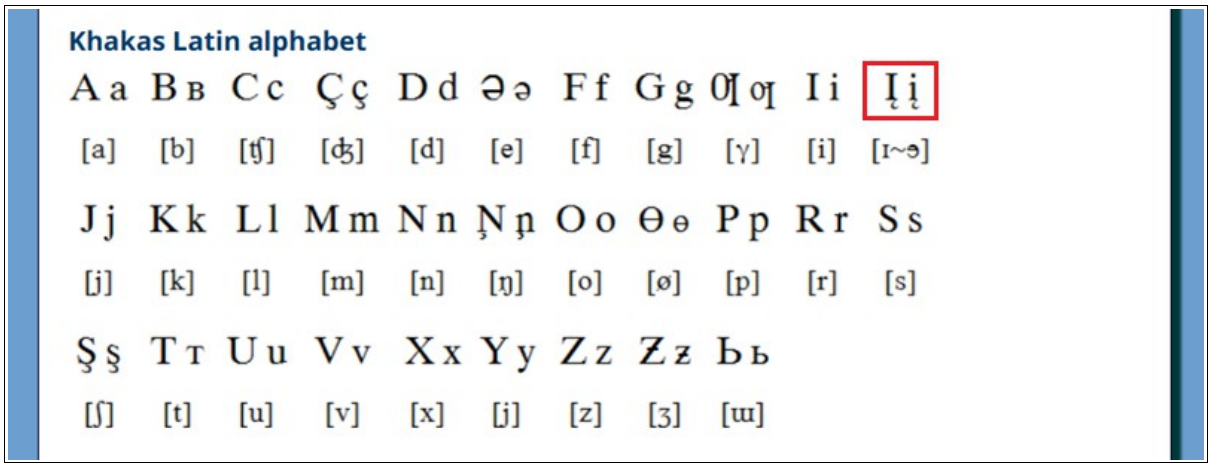


Fig. 7. Cropped section from Ager (2025). Note that Ɨ ɨ is often emulated as Į į in digital environments. Nevertheless, lowercase į is fully dotted, unlike the dotless form (ɨ) found throughout earlier examples.

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646[1]
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:** *Proposal to encode two Latin orthographic pairs derived from "I i"*
2. Requester's name: *Ismael Robles*
3. Requester type (Member body/Liaison/Individual contribution): *Individual contribution*
4. Submission date: *2026-06\-28*
5. Requester's reference (if applicable):
6. Choose one of the following:
   This is a complete proposal: *Yes*
   (or) More information will be provided later:

B. Technical – General

1. Choose one of the following:
   a. This proposal is for a new script (set of characters): *No*
      Proposed name of script:
   b. The proposal is for addition of character(s) to an existing block: *Yes*
      Name of the existing block: *Latin Extended-D*
2. Number of characters in proposal: *2*
3. Proposed category (select one from below - see section 2.2 of P&P document):
   A-Contemporary A | 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?
   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? *Ismael Robles Haloui*
   b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.): *Google (Noto Sans)*
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)? *Yes*

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.

This letter, I with descender, is the last letter of Latinized Khakas, as confirmed in Samrin (1934; see Figure 2), and Unicode

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)

should sort it accordingly.

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before? *Yes*
   If YES explain *See proposal*
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.)? *No*
   If YES, with whom?
   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? *Turkicists, philologists, linguists*
   Reference: *See proposal*
4. The context of use for the proposed characters (type of use; common or rare) *Formerly common*
   Reference: *See proposal*
5. Are the proposed characters in current use by the user community? *Yes*
   If YES, where? Reference: *See proposal*
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? *Yes*
   If YES, is a rationale provided? *Yes*
   If YES, reference: *To keep them with associated characters (e.g. Ƣ ƣ, Ɲ ɲ, Ɵ ɵ)*
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? *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?
    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:

UnicodeData.txt

```
14364 A7FE;LATIN EPIGRAPHIC LETTER I LONGA;Lo;0;L;;;;;N;;;;;
1     A7FE;LATIN CAPITAL LETTER I WITH DESCENDER;Lu;0;L;;;;;N;;;;A7FF;
14365 A7FF;LATIN EPIGRAPHIC LETTER ARCHAIC M;Lo;0;L;;;;;N;;;;;
2     A7FF;LATIN SMALL LETTER DOTLESS I WITH DESCENDER;Ll;0;L;;;;;N;;;A7FE;;
      A7FE
```

+2 -2