Addendum II to L2/21-107, Cyrillic modifier letters

Kirk Miller, kirkmiller@gmail.com   2022 January 07

This is an addendum to the set of Cyrillic modifier characters proposed in L2/21-107. Two additional characters are requested. First, a review of Kazakh dictionaries found a spacing modifier version of ұ, U+04B1 CYRILLIC SMALL LETTER STRAIGHT U WITH STROKE, a letter used only in that language. Second is a combining form of the Cyrillic "decimal *i*," the dotted letter U+0456 CYRILLIC SMALL LETTER BYELORUSSIAN-UKRAINIAN I, which is used by a number of languages.

The Cyrillic Extended-D block, created for L2/21-107, so far contains only spacing modifier letters. Their combining equivalents are found in the Cyrillic Extended-A and -B blocks, which provide most of the letters needed by East Slavic languages. However, the evidence for those characters was from medieval manuscripts and there thus remain a few gaps for modern use, such as *г*, *i*, *ə* (though double-dot *ï* is supported). Of these, the "decimal *i*" has now been found. For medieval use, *i* would not have a dot. However, Sebastian Kempgen at the University of Bamberg, who specializes in digitizing medieval Slavonic manuscripts, said that the presence or absence of the dot is not of great importance and that he would be happy for Unicode to support the modern dotted form of a combining *i* (p.c. 2021 Aug 31). Note that L2/21-107 included a *spacing* modifier *i* at U+1E04C MODIFIER LETTER CYRILLIC SMALL BYELORUSSIAN-UKRAINIAN I, but this is semantically distinct. (See Figure 6.)

Other characters (not requested)

A Bulgarian-style modifier д has been found. It resembles IPA ⟨g⟩, as seen in Figure 1. It should be encoded as U+1E034 MODIFIER LETTER CYRILLIC SMALL DE, with the graphic distinction handled by font choice or by specifying the language.

Figure 1. Bičeldej (2001: 81). Bulgarian-style superscript ⟨g⟩ (д).

To enclose combining letters in parentheses, as in Figure 2. , use U+1ABB COMBINING PARENTHESES ABOVE.

Figure 2. Pohribnyj (1986: 7). Combining superscript with parentheses.

To narrow the scope of an apostrophe or similar mark such as U+0315 COMBINING COMMA ABOVE RIGHT to a combining or modifier superscript, the CGJ character (U+034F) is needed to separate it from the base letter. For example, to add an apostrophe to both the base letter and the superscript, as in Figure 3. , the encoding sequence would be:

[*baseline letter* + *apostrophe*] + CGJ + [*superscript* + *apostrophe*].
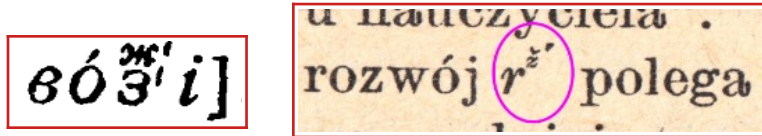

Figure 3. Pohribnyj (1986:12), combining apostrophe on Cyrillic combining superscript, and Klemensiewicz et al. (1965: 111), combining apostrophe on Latin modifier superscript.

Location

Greyed-out cells were approved by the SAH for L2/21-107 and addendum I. For the chart of the full Cyrillic Extended-D block, 1E030–1E08F, see the end of this proposal.

	...0	...1	...2	...3	...4	...5	...6	...7	...8	...9	...A	...B	...C	...D	...E	...F
Cyrillic Extended-D																
U+1E06x	ф	х	ц	ч	ш	ъ	ы	ґ	і	s	џ	ҫ	ы	¥		
U+1E07x																
U+1E08x																і̊

Characters

¥   1E06D MODIFIER LETTER CYRILLIC SMALL STRAIGHT U WITH STROKE. Figure 4–5.

і̊   1E08F COMBINING CYRILLIC SMALL LETTER BYELORUSSIAN-UKRAINIAN I. Figure 6.

Properties

```
1E06D;MODIFIER LETTER CYRILLIC SMALL STRAIGHT U WITH STROKE;
     Lm;0;L;<super> 04B1;;;;N;;;;;
1E08F;COMBINING CYRILLIC SMALL LETTER BYELORUSSIAN-UKRAINIAN I;
     Mn;230;NSM;;;;;N;;;;;
```

References

Bičeldej : К.А. Бичелдей (2001) *Звуковой строй диалектов тувинского языка.*

Z. Klemensiewicz, T. Lehr-Spławiński & S. Urbańczyk (1965) *Gramatyka historyczna języka polskiego.* Warsaw.

Pohribnyj : Уклав М.І. Погрібний (1986) *Орфоепічний словник.* Kiev.

Uäli et al. : Н. Уәли, Қ. Күдеринова & А. Фазылжанова (2004) *Қазақ тілінің орфоэпиялық сөздігі.* Арыс, Almaty.

Figures

Modifier ұ (ᵞ)

19.	у+дауысты	уақ	ᵞу	➡ ᵞуақ
20.	у+дауыссыз	улы	ұу	ұулұ
32.	р+жуан	рас	ыр	ырас
33.	р+жіңішке	рет	�	ᵓрет
34.	р+еріндік жуан	ру	ᵞр	➡ ᵞрұу
35.	р+еріндік жіңішке	рұстем	ᵞр	ᵞрұстөм

Figure 4. Uäli (2004: 11). Contrast between modifier [ᵞ] and other epenthetic vowels, including straight u without a stroke, [ᵞ], in the bottom row.

жыл ᵞуан, жыл ᵞуаңға

мақ ᵞрұм, мақ ᵞрұм ͡ ғалұу

Figure 5. Uäli (2004: 309, 482, 486, 706). Random examples of [ᵞ] in pronunciations of dictionary entries.

ᵞлоӄсұу, ᵞлоӄсұ, ᵞлоӄсұйдұ

ұсұнұу, ұсұмбау, ұ
ұсᵞнұушұ
ұсᵞнұлұу
ұсᵞнұм
ұсᵞнұс
ұсᵞнұсұу, ұсᵞнұсса

Combining i (◌ᶦ)

[мᶦи], [нᶦива]; [мᵘи], [нᵘива].

[контрас ͫ ний] [на вóзᶾᶦі], ⁻[ч̌ᶦіно], [ц̌ᶦінá]

Figure 6. Pohribnyj (1986: 9). Combining *i* vs combining *u*. These are semantically distinct from spacing superscripts. Cf. the spacing superscript *m* at bottom left, in explanatory notes on p. 5, which indicates a transitional or epenthetic sound, vs combining superscript ж ш ч from p. 12 at bottom right, which indicate shades of sound or conflation of sounds. Thus the ⟨и̇⟩ illustrated here is a sound intermediate between [и] and [i] (i.e., a monophthong), whereas spacing ⟨иᶦ⟩ would be [и] with an off-glide [i] (i.e., a diphthong).

Full code block chart

	1E03	1E04	1E05	1E06	1E07	1E08
0	а 1E030	т 1E040	ӏ 1E050	ф 1E060		
1	δ 1E031	у 1E041	а 1E051	х 1E061		
2	в 1E032	ф 1E042	δ 1E052	ц 1E062		
3	г 1E033	х 1E043	в 1E053	ч 1E063		
4	д 1E034	ц 1E044	г 1E054	ш 1E064		
5	е 1E035	ч 1E045	д 1E055	ъ 1E065		
6	ж 1E036	ш 1E046	е 1E056	ы 1E066		
7	з 1E037	ы 1E047	ж 1E057	ґ 1E067		
8	и 1E038	э 1E048	з 1E058	і 1E068		
9	к 1E039	ю 1E049	и 1E059	ѕ 1E069		
A	л 1E03A	ѕ 1E04A	к 1E05A	ц 1E06A		
B	м 1E03B	ə 1E04B	л 1E05B	ç 1E06B		
C	о 1E03C	і 1E04C	о 1E05C	ы 1E06C		
D	п 1E03D	ј 1E04D	п 1E05D	¥ 1E06D		
E	р 1E03E	ө 1E04E	с 1E05E			
F	с 1E03F	ү 1E04F	у 1E05F			i◌̈ 1E08F

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 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:** *Cyrillic modifier letters*
2. Requester's name: *Kirk Miller*
3. Requester type (Member body/Liaison/Individual contribution): *individual*
4. Submission date: *2022 January 07*
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: *Cyrillic 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   **x**   B.1-Specialized (small collection) ___   B.2-Specialized (large collection) ___
   C-Major extinct ___   D-Attested extinct ___   E-Minor extinct ___
   F-Archaic Hieroglyphic or Ideographic ___   G-Obscure or questionable usage symbols
4. Is a repertoire including character names provided? *yes*
   a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document? *yes*
   b. Are the character shapes attached in a legible form suitable for review? *yes*
5. Fonts related:
   a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?
   Kirk Miller
   b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):
   SIL (Gentium release)
6. References:
   a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided? *yes*
   b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached? *yes*
7. Special encoding issues:
   Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)? *no*

8. Additional Information:

Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at 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? _ _ _ *no* _ _ _

    If YES explain _

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? _ _ _ _ _ _ _ *Sebastian Kempgen, U Bamberg* _ _ _ _ _ _ _ _

        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) _ _ *phonetic* _ _

    Reference: _

5. Are the proposed characters in current use by the user community? _ _ *yes* _ _ _

    If YES, where? Reference: _ _ _ _ _ _ _ *See references* _ _ _ _ _ _ _ _ _ _ _ _ _

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)? _ _ _ *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: _