Additions for Janalif

\[ \text{U+A790 LATIN CAPITAL LETTER N WITH DESCENDER} \rightarrow 04A2 cyrillic capital letter n with descender \]

\[ \text{U+A791 LATIN SMALL LETTER N WITH DESCENDER} \rightarrow 042B cyrillic capital letter yeru \]

\[ \text{U+A792 LATIN CAPITAL LETTER YERU} \rightarrow 0131 latin small letter dotless i \]

\[ \text{U+A793 LATIN SMALL LETTER YERU} \rightarrow 042C cyrillic capital letter soft sign \]

\[ \text{U+A793 LATIN SMALL LETTER YERU} \rightarrow 0184 latin capital letter tone six \]

Properties:

A790;LATIN CAPITAL LETTER N WITH DESCENDER;Lu;0;L;;;;;N;;;;A791;
A791;LATIN SMALL LETTER N WITH DESCENDER;Ll;0;L;;;;;N;;;A790;;A790
A792;LATIN CAPITAL LETTER YERU;Lu;0;L;;;;;N;;;;;;;;;;;;;;;;A793;
A793;LATIN SMALL LETTER YERU;Ll;0;L;;;;;N;;;A792;;A792

1. The Jañalif alphabet (fig. 3, 4)

In 1908–1909 the Tatar poet Sä_git Rämiev started to use the Latin alphabet in his own works. He offered the use of digraphs: ea for ä, eu for ü, eo for ö and ei for i. But Arabists turned down his project. In the early 1920s Azerbaijanis invented their own Latin alphabet, but Tatarstan scholars set a little store to this project, preferring to reform the İske imlâ (en.wikipedia.org/wiki/iske_imla). The simplified İske imlâ, known as Yaña imlâ (en.wikipedia.org/wiki/yana_imla) was used from 1920–1927. [1]

But Latinization was adopted by the Soviet officials and the special Central Committee for a New Alphabet was established in Moscow. The first project of the Tatar-Bashkir Latin alphabet was published in Eşce (The Worker) gazette in 1924. The pronunciation of the alphabet was similar to English, unlike the following. Specific Bashkir sounds were written with digraphs. However, this alphabet was declined. [1]

In 1926 the Congress of Turkologists in Baku recommended to switch all Turkic languages to the Latin alphabet. Since April of 1926 the Jañ tatar elifbası/Yaña Tatar alifbasi (New Tatar alphabet) society started its work at Kazan. [2]

Since 3 July 1927, Tatarstan officials have declared Jañalif as the official script of the Tatar language, replacing the Yaña imlâ script. In this first variant of Jañalif (acutes-Jañalif), there weren't separate letters for K and Q (realized as K) and for G and Ğ (realized as G), V and W (realized as W). Ş (sh) looked like the Cyrillic letter LL (she). C and Ç were realized as in Turkish and the modern Tatar Latin alphabet and later were transposed in the final version of Jañalif. [1]
In 1928 Janařif was finally reformed and was in active usage for 12 years (see fig. 3, 4). This version of Janařif is the base of our proposal. Some sources claim this alphabet having 34 letters, but the last was a digraph bj, used for the corresponding Tatar diphthong. [1] Another source states that the 34th letter was an apostrophe. They also give another sorting of the alphabet. (ө after A, ӓ after E) [2]

In 1939 Cyrilization of USSR was initiated. As was said, alphabet was switched to Cyrillic "by labor's request."

There are also several projects of Cyrilization. Ilminski's alphabet was already forgotten and it couldn't be used, due to its religious origin. As early as 1938 professor M. Fazlílin introduced an adaptation of the Russian alphabet for the Tatar language, without any additional characters. Specific Tatar letters should be signed with the digraphs, consisting of similar Russian letters and the letters ю and ә. [1]

In 1939 Qorbangaliev and Ramazanov offered their own projects that planned to use additional Cyrillic characters. Letters О, Э, Я, Й were inherited from Janařif, but Ж and Х were invented by analogy with Ж and Ц. Г and К should be used to designate Щ and Ч. By this project "гадат" ("custom") was spelled as "ғәдәт", "гар" ("snow") as "қәр". In Ramazanov's project W (Janařif V) was marked by В before the vowel, and У, Я in the end of syllable. Janařif: vaq - вәк; тау - тәү; дек - деә. In 5 May 1939 this project was established as official by the Supreme Soviet of TASSR. Surprisingly, "Tatar society disagreed to this project" and during 1940 July conference Cyrillic alphabet was finally standardized. 10 January 1941 this project was passed. According to this version, "гадат" was spelled as "раәт", "гар" as "қәр". The principles were following: if ra/ro/ry/g/ka/kо/ky/kү/ is followed by "soft syllable", containing "ә, е, е, и, у" or soft sign "ю", they are spelled as әй/ёй/ёй/же/же/же, in other cases as әй/ёй/ёй/же/же/же, ra/ro/ry/ka/ke/ky/ke are spelled as әй/ёй/ёй/же/же/же, in 5 May 1939 this project was passed. According to this version, "гадат" was spelled as "раәт", "гар" as "қәр". The principles were following: if ra/ro/ry/g/ka/kо/ky/kү/ is followed by "soft syllable", containing "ә, е, е, и, у" or soft sign "ю", they are spelled as әй/ёй/ёй/же/ же, in other cases as әй/ёй/ёй/же/ же/ же, ra/ro/ry/ka/ke/ky/ke are spelled as әй/ёй/ёй/ же/ же/ же, in other cases as әй/ёй/ёй/ же/ же/ же. Similar practice were applied for е, ю, я, that could be spelled as ye, yә, yә and as yи, yи, yи. Examples: канәгать - qәңәгәт (satisfied); ел - yил (year); әым - әым (charm). So, in Tatar Cyrillic soft sign hasn't sense of iotation, as in Russian, but a sense of vowel harmony. Unlike modern Russian, some words can end with ю, to sign a "hard g" after the "soft vowel", as in бәлигъ - балигъ (of the full legal age). [1]

All Russian words are written as in Russian and should be pronounced with Russian pronunciation. In the 1990s some wanted to restore Janařif, or Janařif+W, as being corresponding to modern Tatar phonetics. But technical problems, such as font problems and the disuse of Uniform Turkic alphabet among other peoples forced to use "Turkish-based alphabet". In 2000 that alphabet was adopted by the Tatarstan government, but in 2002 it was abolished by the Russian Federation. [1]

2. The N with descender

Fig. 2 - Scan from [1]

The descender of the proposed letters U+790/U+791 LATIN CAPITAL (resp. SMALL) LETTER N WITH DESCENDER look like the descenders of e.g. U+2677/U+2678 LATIN CAPITAL (resp. SMALL) LETTER H WITH DESCENDER.

Therefore, the names proposed here were selected according to this example. In current citations of Janařif texts, these letters are usually replaced by U+014A/U+014B LATIN CAPITAL (resp. SMALL) LETTER ENG, as these letters have a superficial but recognizable similarity to the correct Janařif letter, and as they are usually attributed to the same sound.

Also, the letter's usage was considered in 2000 Tatar Latin alphabet. Only some Tatar fonts use this glyph at the position of N.

Nevertheless, their form is distinctive and clearly different from the eng, which is also distinctive (even for the upper case eng of which all glyph variants concurn in the form of their lower right appendage). The lower right appendage of the n with descender is always straight and placed right of the right n stem, while the lower right appendage of the eng is always a prolongation of the right n stem and bound

Proposal to encode four Latin letters for Janařif — 2009-03-16
inwards.
Thus, the n with descender is no glyph variant of the eng.
If it were so, the letters U+0220/U+019E LATIN CAPITAL (resp. SMALL) LETTER N WITH LONG LEG
also had to be regarded as glyph variant of the eng, as they in fact are more similar (the lower right
appendage being straight but a prolongation of the right n stem).

Additional, the N with descender was used in parallel to the eng in the Latin alphabet used to the Khanty
language about 1931-1936 (fig. 5).
Thus, it is a separate letter from eng in any case.

3. The Latin yeru

Fig. 1 - Scan from [1]

While the proposed U+A792 "LATIN CAPITAL LETTER YERU" (with its lower case counterpart U+A793
"LATIN SMALL LETTER YERU") looks like the Cyrillic letters U+042C/U+044C CYRILLIC CAPITAL
(resp. SMALL) LETTER SOFT SIGN, it is by no ways a soft sign and never used as such in Jaŋalif
context.

In fact, it is a Latin equivalent to U+042B/U+044B CYRILLIC CAPITAL (resp. SMALL) YERU.
Thus, it is an "i" variant by function, equivalent to the Turkish/Azerbaijani dotless i.
(The proposed naming does not prevent anybody from using the character as soft sign in nonstandard
Cyrillic transcriptions or transliterations, as anybody is free to use any letters in any way.)

The letter is obviously different from the superficially similar U+0184/U+0185 LATIN CAPITAL (resp.
SMALL) LETTER TONE SIX, where the vertical stem is terminated at the top by a distinctive slanted
appendage, and where both capital and small form have cap-height and are distinguished by the lateral
extension of the bowl.

Using the Cyrillic U+042C/U+044C as substitute in current citations of Jaŋalif text (as it is in fact be done
now due to the lack of an encoded Latin b/ᵦ), is as undesirable as having to use U+0420/U+0440
CYRILLIC CAPITAL (resp. SMALL) LETTER ER to denote the "p" in Latin text, as a substitute for a
(hypothetically) not encoded U+0050/U+0070 LATIN CAPITAL (resp. small) LETTER P.

There also some points shall be noted which are similar to the situation of the Kurdish W/w [3], which
was encoded at last (U+051C/051D). As pointed out above, Janalif is a stable alphabet, used for several
years for several languages beyond Tatar, with a definitive sorting order: the yeru is the last letter in that
alphabet after Z and Ŷ (as long as the diphong sʲ is not considered). Since Tatar, over its history, is
written in the Latin as well as in the Cyrillic alphabet, a multilingual wordlist cannot sort Kurdish correctly
because the y-looking letter (beyond its complete different function) cannot be in two places at the same
time. (Sorting here means ordinary plain-text sorting, for instance of files in a directory.) Expecting
Janalif users to have recourse to special language-and-script tagging software for these two letters
alone is simply not a credible defense for the retention of the unification of two letters with complete
different function.

4. References:
[1] (Russian) М.З. Закиев. Тюрко-татарское письмо. История, состояние, перспективы. Москва,
"Инсан", 2005

of the Tatar Encyclopaedia.

[3] Michael Everson et al., "Proposal to encode additional Cyrillic characters in the BMP of the UCS"
5. Examples

Fig. 3: Table of Jaŋalif, from [1]
Fig. 4: Another table of Jaŋalif, from [2]

Fig. 5: Table of the Latin alphabet used 1932-1936 for the Khanty language, showing the n with descender and the eng side by side as different letters. Retrieved 2008-10-31 from http://upload.wikimedia.org/wikipedia/commons/9/9c/Hanti_latin_alphabet.jpg

Fig. 6: Entry in http://www.w3.org/2008/05/ita/lsr.xml (as of 2009-03-16). It shows the Latin yeru in a registry entry (Əlfibasi with transliteration Elibasi, using the ь as well as the ё as substitutes for the correct Jaŋalif characters, as such a database is by nature confined to already encoded Unicode characters).

Fig. 7: Title page from a Kazhak newspaper from about 1937, showing all proposed letters. Retrieved 2008-10-25 from http://en.wikipedia.org/wiki/Image:Sotsijaldy_qazaqstan.jpg
The descender of the lower case n with descender shows a drop-like form here in the headline font, showing that the letter has developed some glyph variants during the time of its use.

Fig. 8: Example from a Bashkir text of the Jañalif era. While there are a lot of easy to find Latin yerus, some n with descender are encircled in red. (The letters encircled in cyan are special Bashkir Latin letters which are unencoded yet but not subject of this proposal.)

Fig. 9: Scan from the workbook (Трудовая книжка - Xezmëtknegëse) from В.П. Емельянов, the grand-grandfather of one of the authors of this proposal (I.Ye.), about 1938. This example shows many Latin yerus and some n with descender (e.g. the last letter of the second word of the first line). — By the way, this example also shows the use of U+0299 LATIN SMALL CAPITAL LETTER B as lower case counterpart for U+0042 LATIN CAPITAL LETTER B (see e.g. the first word in the second line), as it came into use for Jañalif to make the b dissimilar from the Latin yeu.
# Proposal Summary Form to Accompany Submissions

**For Additions to the Repertoire of ISO/IEC 10646**

Please fill all the sections A, B and C below.

Please ensure you are using the latest form from [http://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html](http://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html) for guidelines and details before filling this form.


---

## A. Administrative

<table>
<thead>
<tr>
<th>1. Title:</th>
<th>Proposal to encode four Latin letters for Janalif</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. Requester's name:</td>
<td>Karl Pentzlin, Ilya Yevlampiev</td>
</tr>
<tr>
<td>3. Requester type (Member body/Liaison/Individual contribution):</td>
<td>Individual Contribution</td>
</tr>
<tr>
<td>4. Submission date:</td>
<td>2008-11-03, revised 2009-03-16</td>
</tr>
<tr>
<td>5. Requester's reference (if applicable):</td>
<td></td>
</tr>
<tr>
<td>6. Choose one of the following:</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>This is a complete proposal:</td>
</tr>
</tbody>
</table>

---

## B. Technical – General

<table>
<thead>
<tr>
<th>1. Choose one of the following:</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. This proposal is for a new script (set of characters):</td>
</tr>
<tr>
<td>Proposed name of script:</td>
</tr>
<tr>
<td>b. The proposal is for addition of character(s) to an existing block:</td>
</tr>
<tr>
<td>Name of the existing block:</td>
</tr>
<tr>
<td>2. Number of characters in proposal:</td>
</tr>
<tr>
<td>3. Proposed category (select one from below - see section 2.2 of P&amp;P document):</td>
</tr>
<tr>
<td>A-Contemporary</td>
</tr>
<tr>
<td>C-Major extinct</td>
</tr>
<tr>
<td>F-Archaic Hieroglyphic or Ideographic</td>
</tr>
<tr>
<td>4. Is a repertoire including character names provided?</td>
</tr>
<tr>
<td>a. If YES, are the names in accordance with the “character naming guidelines” in Annex L of P&amp;P document?</td>
</tr>
<tr>
<td>b. Are the character shapes attached in a legible form suitable for review?</td>
</tr>
<tr>
<td>5. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?</td>
</tr>
<tr>
<td>If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:</td>
</tr>
<tr>
<td>6. References:</td>
</tr>
<tr>
<td>a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?</td>
</tr>
<tr>
<td>b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?</td>
</tr>
<tr>
<td>7. Special encoding issues:</td>
</tr>
<tr>
<td>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)?</td>
</tr>
</tbody>
</table>

---

## 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](http://www.unicode.org) for such information on other scripts. Also see [http://www.unicode.org/Public/UNIDATA/UCD.html](http://www.unicode.org/Public/UNIDATA/UCD.html) and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

---

## Technical - Justification

1. Has this proposal for addition of character(s) been submitted before?  
   
   | 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?  
   
   | One of the authors (I.Ye.) is himself a member of the user community |

3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?  
   Reference:  
   
   | see text |

4. The context of use for the proposed characters (type of use; common or rare)  
   Reference:  
   
   | common within their context (see text) |

5. Are the proposed characters in current use by the user community?  
   If YES, where?  
   Reference:  
   
   | see text |

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:  
   
   | Keeping in line with other Latin 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 an existing character?  
    If YES, is a rationale for its inclusion provided?  
    If YES, reference:  
    
    | Yes | See text (in short: resembles a Cyrillic character in form but not in function) |

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:  
    
    | 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)  
    If YES, reference:  
    
    | No |

13. Does the proposal contain any Ideographic compatibility character(s)?  
    If YES, is the equivalent corresponding unified ideographic character(s) identified?  
    If YES, reference:  
    
    | No |