Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Organisation Internationale de Normalisation
Международная организация по стандартизации

L2/10-356
WG2 N 3916

Doc Type: Working Group Document
Title: Proposal to encode two Latin letters for Jaꞑalif
(remaining characters of N3581 from 2009-03-16 which was partially accepted)
Source: Karl Pentzlin, Ilya Yevlampiev (Илья Евлампиев)
Status: Individual Contribution
Action: For consideration by JTC1/SC2/WG2 and UTC
Date: 2010-09-24

Additions for Janalif

Ь	U+A792	LATIN CAPITAL LETTER YERU → 042B cyrillic capital letter yeru → 042C cyrillic capital letter soft sign → 0184 latin capital letter tone six
ь	U+A793	LATIN SMALL LETTER YERU → 0131 latin small letter dotless i

Properties:

```
A792;LATIN CAPITAL LETTER YERU;Lu;0;L;;;;;N;;;;A793;
A793;LATIN SMALL LETTER YERU;Ll;0;L;;;;;N;;;A792;;A792
```

1. The Jaꞑalif alphabet **(fig. 3, 4; excerpt from N3581)**

In 1908–1909 the Tatar poet Säğit 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 ı. 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 Jaꞑa tatar əlifbasь/Yaña Tatar älifbası (New Tatar alphabet) society started its work at Kazan. [2]

Since 3 July 1927, Tatarstan officials have declared Jaꞑalif as the official script of the Tatar language, replacing the Yaña imlâ script. In this first variant of Jaꞑalif (acutes-Jaꞑalif), 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 Ш (she). C and Ç were realized as in Turkish and the modern Tatar Latin alphabet and later were transposed in the final version of Jaꞑalif. [1]

In 1928 Jaꞑalif was finally reformed and was in active usage for 12 years (see fig. 3, 4). This version of Jaꞑalif is the base of our proposal.

2. The Latin yeru

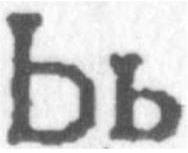


Fig. 1 - Scan from [1]

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 Jaꞑalif 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.)

Using the Cyrillic U+042C/U+044C as substitute in current citations of Jaꞑalif text (as it is in fact be done now due to the lack of an encoded Latin Ь/ь), 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, Jaꞑalif is a stable alphabet, used for several years for several languages beyond Tatar, with a definitve sorting order: the yeru is the last letter in that alphabet after Z and Ƶ (as long as the diphtong ьj 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 ь-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 Jaꞑalif 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.

2.1 The Latin yeru vs. the Latin Letter Tone Six

The Latin yeru obviously is different from the superficially similar U+0184/U+0185 LATIN CAPITAL (resp. SMALL) LETTER TONE SIX:

- The “tone six” (in both cases) always has a specific triangular appendage at the top left, which replaces the serif in the serifed fonts, and which is retained in sans-serif fonts.
- The small form of the “tone six” always has cap-height, and is distinguished from the capital form by a more round bowl (oblique at both connections to the stem). Except for the top right appendage, it usually resembles the small Latin “b”.
- The small form of the “Latin yeru” however always looks like a small capital form, retaining horizontal lines and right angles at the connection to the stem.

This is shown by the following examples in several widespread fonts (showing the Latin yeru by the related glyph of the Cyrillic soft sign first, followed by the “tone six” glyph, upper and lower case each, followed by the small Latin b:

Ьь Ƅƅ b – Arial

Ьь Ƅƅ b – Times New Roman

Ьь Ƅƅ b – Cardo

Ьь Ƅƅ b – Doulos SIL

Ьь Ƅƅ b – Gentium

Ьь Ƅƅ b – Roman Cyrillic

When the Zhuang tone letters (used for a short time in a now obsolete orthography) “two” to “six” were introduced into Unicode, according to the current encoding and the annotations in the Unicode table, two of them were unified with Cyrillic letters (shown in dark red in the following specimens), while

Ƨƨ Зз Чч Ƽƽ Ƅƅ

If, at that time, it were considered that it would be appropriate to have the same glyph for the Cyrillic soft sign and the Zhuang letter tone six, the six had been unified like the three and the four.
Thus, in the same way, it is not appropriate to unify the Latin yeru with the Latin Zhuang tone six when introducing into the Latin script. The Latin yeru is a different letter, as proposed here.

3. Use of Cyrillic-like Letterforms in Newly Designed Latin Alphabets

The concern was raised that encoding the Latin yeru would open the door to “duplicate the whole Cyrillic alphabet as Latin letters”.

In fact, this is not the case.
In the late 1920s and the 1930s (in some cases even earlier, the Latin alphabet was introduced for many minority languages in the Soviet Union. A considerable part of these were extensions of the Janalif, thus containing the Latin yeru to denote the sound of the Cyrillic yeru or related vowels.

While it was “natural” that the designers of the extra letters were influenced by the Cyrillic letters they know, they did copy the exact letterforms (especially doing so for both cases) only in rare instances.
In preparation of a proposal to encode such letters (which will not reach a presentable state before 2011), we until now found no evidence. Alphabets of that era apparently were designed avoiding using of Cyrillic-like letterforms, while containing the Latin yeru whether they are derived from Janalif or not (see fig. 8, 10, 11, 12).

If examples for use of true Cyrillic letterforms are found which are used only in single alphabets for a small community, in fact it can be considered to treat them like the Zhuang tones three and five, i.e. unifying them with the resembling Cyrillic letters.

However, the Latin yeru is a letter used throughout the Soviet Union during the lifetime of Janalif, by small minorities as well as by large ethnic groups like the Tatars.
Thus, it is appropriate to recognize its use as a Latin letters by giving it a separate code point, like the Latin P and the Cyrillic P have separate code points within their scripts.

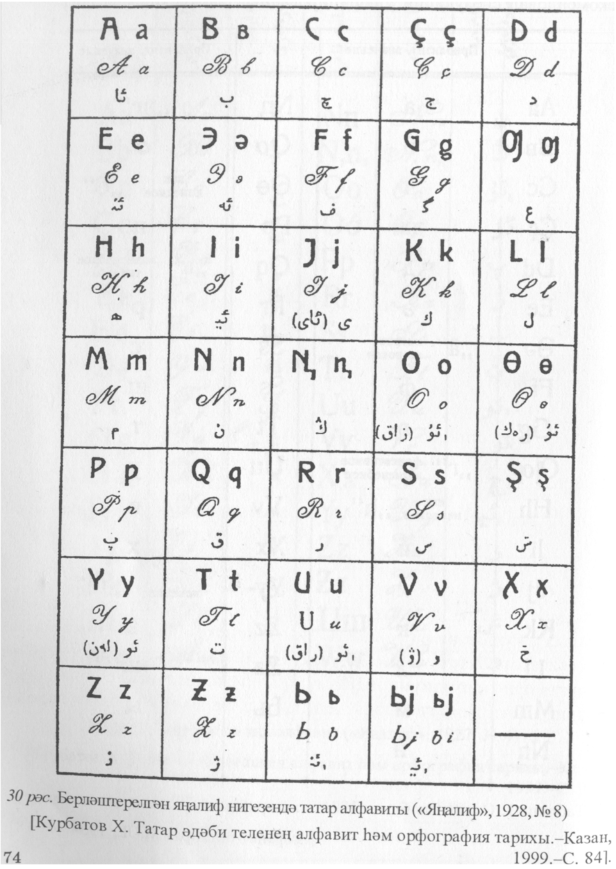
4. References:

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

[2] "Яңалиф". Tatar Encyclopedia. (2002). Kazan: Tatarstan Republic Academy of Sciences Institution of the Tatar Encyclopaedia.

[3] Michael Everson et al., "Proposal to encode additional Cyrillic characters in the BMP of the UCS" (2007-03-21). Unicode document L2/07-003R; SC2/WG2 document N3194R.

5. Examples

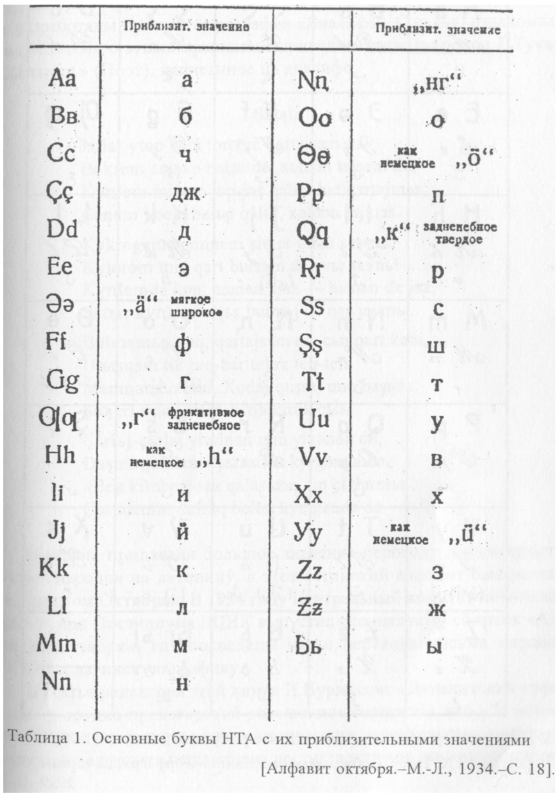


A a A a ئا	B в B в ب	C c C c چ	Ç ç Ç ç ج	D d D d د
E e E e ئە	Ə ə Ə ə ئه	F f F f ف	G g G g گ	Ƣ ƣ غ
H h H h ه	I i I i ئی	J j J j ی (ئای)	K k K k ك	L l L l ل
M m M m م	N n N n ن	Ꞑ ꞑ ڭ	O o O o ئو، (زاق)	Ɵ ɵ Ɵ ɵ ئۆ (رەك)
P p P p پ	Q q Q q ق	R r R r ر	S s S s س	Ş ş ش
Y y Y y ئو (لەن)	T t T t ت	U u U u ئو، (راق)	V v V v و (ۋ)	X x X x خ
Z z Z z ز	Ƶ ƶ Ƶ ƶ ژ	Ь ь Ь ь ئ،	Ьj ьj Ьj ьi ئی،	—

30 рәс. Берләштерелгән яңалиф нигезендә татар алфавиты («Яңалиф», 1928, № 8)
[Курбатов Х. Татар әдәби теленең алфавит һәм орфография тарихы.–Казан, 1999.–С. 84].

74

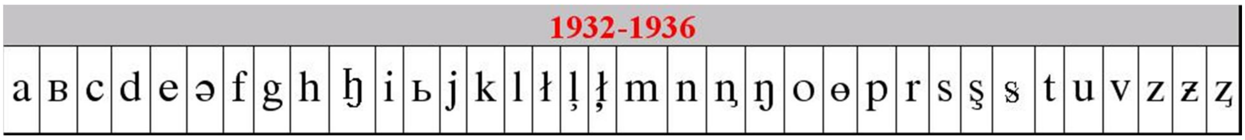
Fig. 3: Table of Jaŋalif, from [1]



	Приблизит. значение		Приблизит. значение
Aa	а	Ꞑꞑ	„нг“
Bв	б	Oo	о
Cc	ч	Ɵɵ	как немецкое „ö“
Çç	дж	Pp	п
Dd	д	Qq	„к“ заднеисбное твердое
Ee	э	Rr	р
Əə	„ä“ мягкое широкое	Ss	с
Ff	ф	Şş	ш
Gg	г	Tt	т
Ƣƣ	„г“ фрикативное задненебное	Uu	у
Hh	как немецкое „h“	Vv	в
Ii	и	Xx	х
Jj	й	Yy	как немецкое „ü“
Kk	к	Zz	з
Ll	л	Ƶƶ	ж
Mm	м	Ьь	ы
Nn	н		

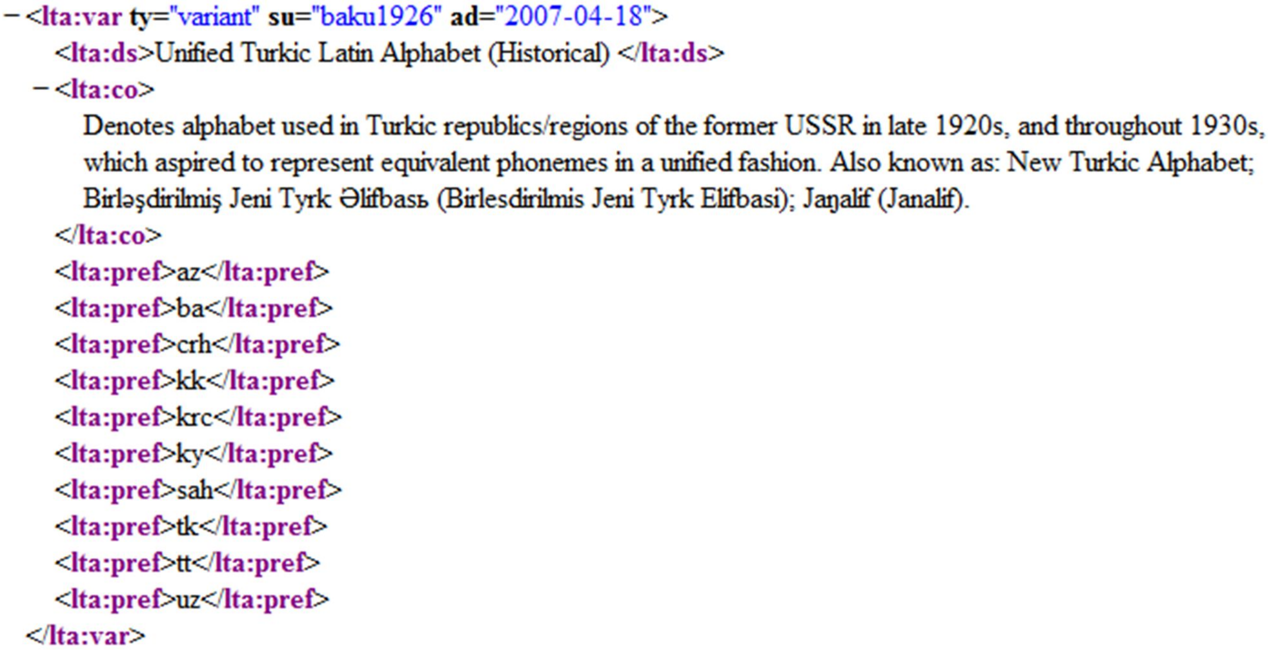
Таблица 1. Основные буквы НТА с их приблизительными значениями
[Алфавит октября.–М.-Л., 1934.–С. 18].

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



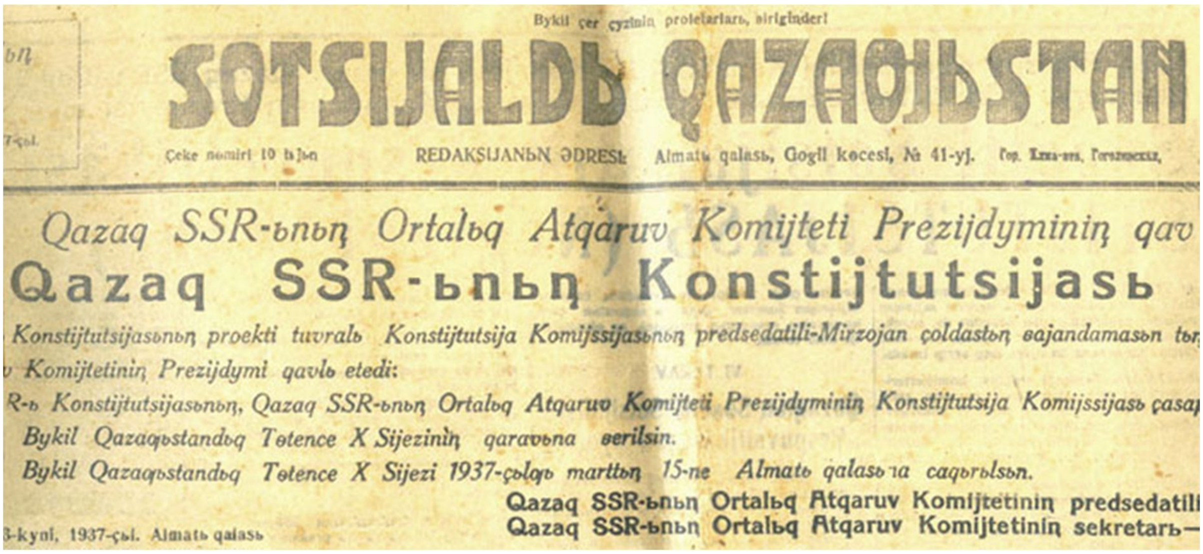
1932-1936																																		
a	в	c	d	e	ə	f	g	h	ꜧ	i	ь	j	k	l	ł	ļ	ł̦	m	n	ꞑ	ŋ	o	ө	p	r	s	ş	ꞩ	t	u	v	z	ƶ	ⱬ

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



```
– <lta:var ty="variant" su="baku1926" ad="2007-04-18">
    <lta:ds>Unified Turkic Latin Alphabet (Historical) </lta:ds>
  – <lta:co>
      Denotes alphabet used in Turkic republics/regions of the former USSR in late 1920s, and throughout 1930s,
      which aspired to represent equivalent phonemes in a unified fashion. Also known as: New Turkic Alphabet;
      Birləşdirilmiş Jeni Tyrk Əlifbasь (Birlesdirilmis Jeni Tyrk Elifbasi); Jaŋalif (Janalif).
    </lta:co>
    <lta:pref>az</lta:pref>
    <lta:pref>ba</lta:pref>
    <lta:pref>crh</lta:pref>
    <lta:pref>kk</lta:pref>
    <lta:pref>krc</lta:pref>
    <lta:pref>ky</lta:pref>
    <lta:pref>sah</lta:pref>
    <lta:pref>tk</lta:pref>
    <lta:pref>tt</lta:pref>
    <lta:pref>uz</lta:pref>
  </lta:var>
```

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



Bykil çer cyzinin proletarları, birigıŋder!

SOTSIJALDЬ QAZAQЬSTAN

Çeke nomiri 10 tıjьn REDAKSIJANЬN ƏDRESI: Almatь qalasь, Gogil kocesi, № 41-yj.

Qazaq SSR-ьnьꞑ Ortalьq Atqaruv Komijteti Prezijdymininꞑ qav

Qazaq SSR-ьnьꞑ Konstijtutsijasь

Konstijtutsijasьnьꞑ proekti tuvralь Konstijtutsija Komijssijasьnьꞑ predsedatili-Mirzojan çoldastьꞑ ʙajandamasьn tьꞑ

Komijtetininꞑ Prezijdymi qavlь etedi:

R-ь Konstijtutsijasьnьꞑ, Qazaq SSR-ьnьꞑ Ortalьq Atqaruv Komijteti Prezijdymininꞑ Konstijtutsija Komijssijasь çasa

Bykil Qazaqьstandьq Tөtence X Sijezininꞑ qaravьna ʙerilsin.

Bykil Qazaqьstandьq Tөtence X Sijezi 1937-çьlqь martьnꞑ 15-ne Almatь qalasьna caqьrьlsьn.

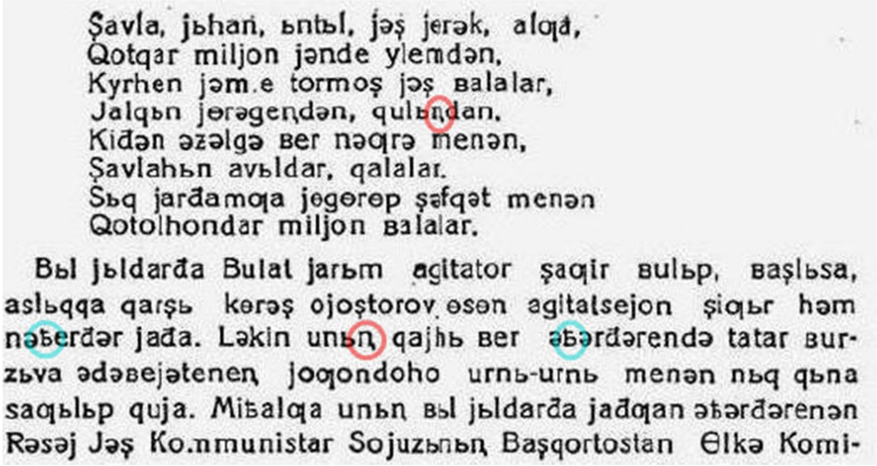
Qazaq SSR-ьnьꞑ Ortalьq Atqaruv Komijtetininꞑ predsedatili

Qazaq SSR-ьnьꞑ Ortalьq Atqaruv Komijtetininꞑ sekretarь—

3-kyni, 1937-çьl. Almatь qalasь

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 .

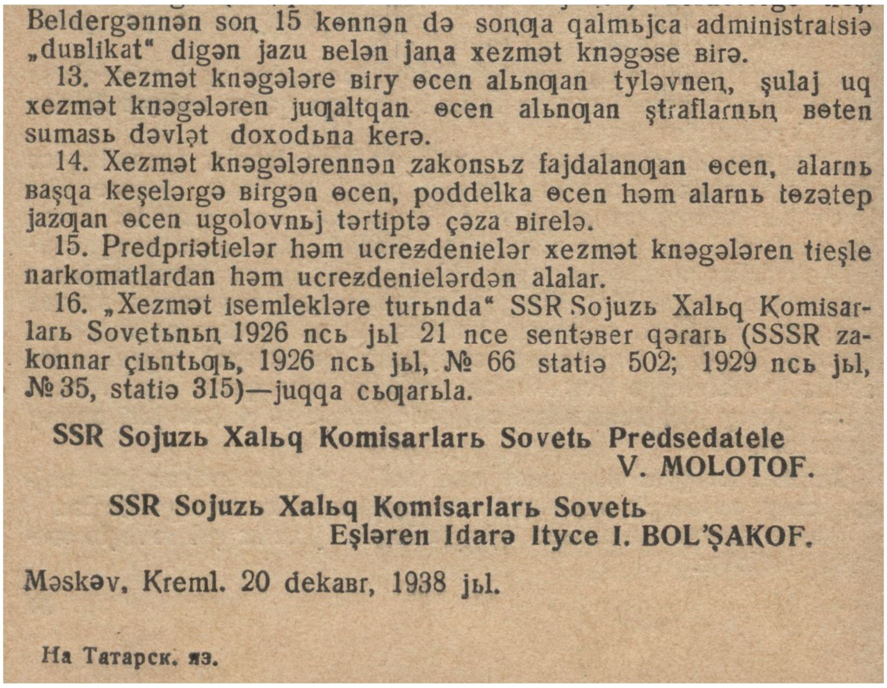
The descender of the lower case n with descender shows a drop-like form here in the headline font, showing that the letter has developed some glyph variants during the time of its use.



Şavla, jьhan, ʙntьl, jəş jerək, alƣđ,
Qotqar miljon jənde ylemdən,
Kyrhen jəm.e tormoş jəş ʙalalar,
Jalqьn jөrəgenꞑdən, qulьꞑdan.
Kiđən əƶəlgə ʙer nəƣrə menən,
Şavlahьn avьldar, qalalar.
Şьq jarđamƣa jөgөrөp şəfqət menən
Qotolhondar miljon ʙalalar.

Bьl jьldarđa Bulat jarьm agitator şaƣir ʙulьp, ʙaşlьsa, aslьqqa qarşь kөrəş ojoştorov ɵsən agitatsejon şiƣьr həm nəƀerđər jađa. Ləkin unьꞑ qajhь ʙer əƀərđərendə tatar ʙurzьva ədəʙejətenenꞑ joƣondoho urnь-urnь menən nьq qьna saƣьlьp quja. Miƀalƣa unьꞑ ʙьl jьldarđa jađƣan əƀərđərenən Rəsəj Jəş Ko.nmunistar Sojuzьnьꞑ Başqortostan Ɵlkə Komi-

Fig. 8: Example from a Bashkir text of the Jaŋalif 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.)
Retrieved 2008-10-28 from
http://ru.wikipedia.org/wiki/Википедия:Проект:Внесение_символов_алфавитов_народов_России_в_Юникод
Picture reference: http://ru.wikipedia.org/wiki/Изображение:Bashqortalifba.jpg



Beldergənnən soꞑ 15 kөnnən də soꞑƣa qalmьjca administratsiə „duʙlikat“ digən jazu ʙelən jaꞑa xezmət knəgəse ʙirə.

13. Xezmət knəgələre ʙiry ɵcen alьnƣan tyləvneꞑ, şulaj uq xezmət knəgələren juƣaltqan ɵcen alьnƣan ştraflarnьꞑ ʙөten sumasь dəvlət doxodьna kerə.

14. Xezmət knəgələrennən zakonsьƶ fajdalanƣan ɵcen, alarnь ʙaşqa keşelərgə ʙirgən ɵcen, poddelka ɵcen həm alarnь tөzətep jazƣan ɵcen ugolovnьj tərtiptə çəza ʙirelə.

15. Predpriətielər həm ucreƶdenielər xezmət knəgələren tieşle narkomatlardan həm ucreƶdenielərdən alalar.

16. „Xezmət isemlekləre turьnda“ SSR Sojuzь Xalьq Komisarlarь Sovetьnьꞑ 1926 ncь jьl 21 nce sentəʙer qərarь (SSSR zakonnar çiьntьƣь, 1926 ncь jьl, № 66 statiə 502; 1929 ncь jьl, № 35, statiə 315)—juqqa cьƣarьla.

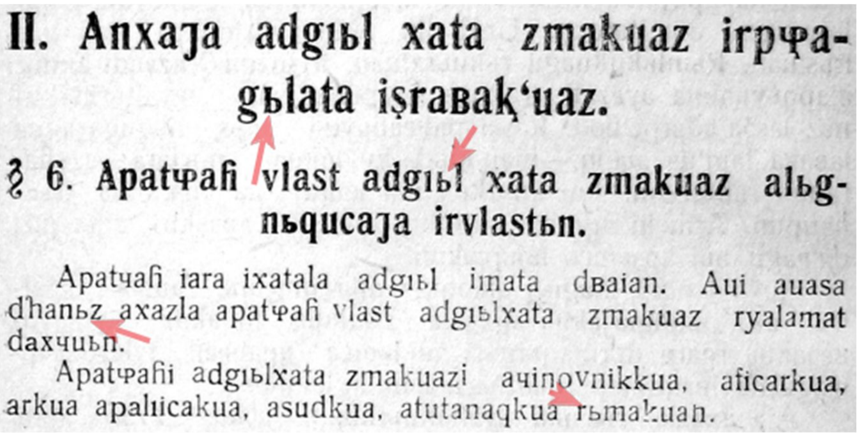
SSR Sojuzь Xalьq Komisarlarь Sovetь Predsedatele
V. MOLOTOF.

SSR Sojuzь Xalьq Komisarlarь Sovetь
Eşləren Idarə Ityce I. BOL'ŞAKOF.

Məskəv, Kreml. 20 dekaʙr, 1938 jьl.

На Татарск. яз.

Fig. 9: Scan from the workbook (Трудовая книжка - Xezmət knəgəse) 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 Jaŋalif to make the b dissimilar from the Latin yeru.



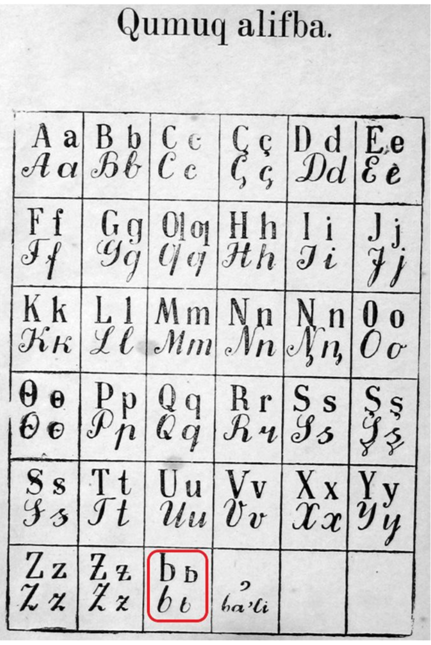
II. Anxaȷa adgıƅl xata zmakuaz irpꝕa-gƅlata işraвaķ'uaz.

§ 6. Apatꝕafi vlast adgıƅl xata zmakuaz alƅg-nƅqucaȷa irvlastƅn.

Apatчafi iara ixatala adgıƅl imata dвaian. Aui auasa d'hanƅz axazla apatꝕafi vlast adgıƅlxata zmakuaz ryalamat daxчuƅn.

Apatꝕafii adgıƅlxata zmakuazi aчinovnikkua, aficarkua, arkua apalıicakua, asudkua, atutanaqkua rƅmakuan.

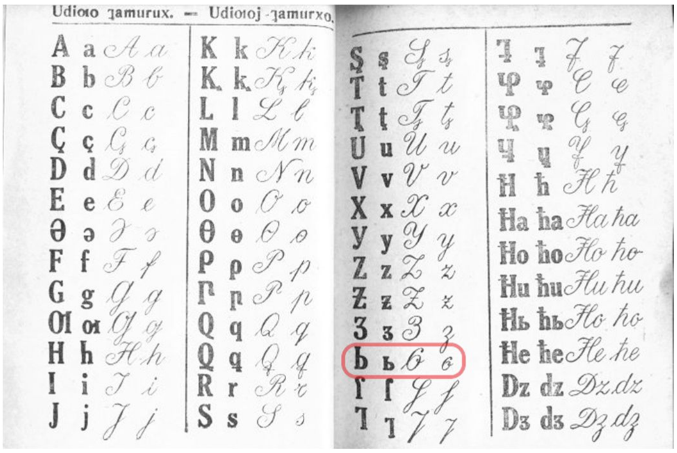
Fig. 10: An Abaza example, showing the Latin yeru besides some other yet unencoded letters.



Qumuq alifba.

A a A a	B b B b	C c C c	Ç ç Ç ç	D d D d	E e E e
F f F f	G g G g	Ƣ ƣ Ƣ ƣ	H h H h	I i I i	J j J j
K k K k	L l L l	M m M m	N n N n	Ꞑ ꞑ Ꞑ ꞑ	O o O o
Ɵ ɵ Ɵ ɵ	P p P p	Q q Q q	R r R r	S s S s	Ş ş Ş ş
Ꞩ ꞩ Ꞩ ꞩ	T t T t	U u U u	V v V v	X x X x	Y y Y y
Z z Z z	Ƶ ƶ Ƶ ƶ	Ь ь Ь ь	ʔ ba'li		

Fig. 11: A page from a Kumyk primer of the 1930s, showing an alphabet which contains the Latin yeru besides a yet unencoded "s with short diagonal stroke".



Udiƣo ȷamurux. — Udiƣoj ȷamurxo.

A a A a
B b B b
C c C c
Ç ç Ç ç
D d D d
E e E e
Ə ə Ə ə
F f F f
G g G g
Ƣ ƣ Ƣ ƣ
H h H h
I i I i
J j J j

K k K k
Ķ ķ Ķ ķ
L l L l
M m M m
N n N n
O o O o
Ɵ ɵ Ɵ ɵ
P p P p
Ƥ ƥ Ƥ ƥ
Q q Q q
Ꝗ ꝗ Ꝗ ꝗ
R r R r
S s S s

Ş ş Ş ş
T t T t
Ţ ţ Ţ ţ
U u U u
V v V v
X x X x
Y y Y y
Z z Z z
Ƶ ƶ Ƶ ƶ
Ʒ ʒ Ʒ ʒ
Ь ь Ь ь
ʃ ʃ ʃ ʃ
Ꞁ ꞁ Ꞁ ꞁ

Ꞁ̵ ꞁ̵ Ꞁ̵ ꞁ̵
Ҹ ҹ Ҹ ҹ
Ҹ̧ ҹ̧ Ҹ̧ ҹ̧
Ч ч Ч ч
Ħ ħ Ħ ħ
Ħa ħa Ħa ħa
Ħo ħo Ħo ħo
Ħu ħu Ħu ħu
Ħь ħь Ħь ħь
Ħe ħe Ħe ħe
Dz dz Dz dz
Dʒ dʒ Dʒ dʒ

Fig. 12: A page from an Udi primer of the 1930s, showing an alphabet which contains the Latin yeru besides some yet unencoded characters.

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://www.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://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html.
See also http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html for latest *Roadmaps*.

A. Administrative

1. **Title:** ***Proposal to encode two Latin letters for Jaŋalif***
2. Requester's name: *Karl Pentzlin, Ilya Yevlampiev*
3. Requester type (Member body/Liaison/Individual contribution): *Individual Contribution*
4. Submission date: *2010-09-24*
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 ____ B.1-Specialized (small collection) *X* 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. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for
   publishing the standard? *Karl Pentzlin*
   If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools
   used: *http://www.pentzlin.com/proposalfont.zip (more information in the info.txt file included in that archive)*
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 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.

[1] Form number: N3152-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)

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? *One of the authors (I. Ye.) is himself a member of the user community*

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? *see text*

Reference: *see text*

4. The context of use for the proposed characters (type of use; common or rare) *common*

Reference: *common within their context (see text)*

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

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? *Yes*

If YES, is a rationale provided? *Yes*

If YES, reference: *Keeping in line with other Latin (especially Janalif) 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? *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 an existing character? *Yes*

If YES, is a rationale for its inclusion provided? *Yes*

If YES, reference: *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? *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? *n/a*

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 character(s)? *No*

If YES, is the equivalent corresponding unified ideographic character(s) identified?

If YES, reference: