

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

Doc Type: Working Group Document

Title: Proposal to encode Arabic characters used for Bashkir, Belarusian, Crimean Tatar, Karachay, Karakalpak, and Tatar languages

Source: See References

Authors: Ilya Yevlampiev, Karl Pentzlin, Nurlan Joomagueldinov

Status: Expert Contribution

Action: For consideration by JTC1/SC2/WG2 and UTC

Date: 28 April 2011

1. Introduction

The letters of Bashkir, Belarusian, Crimean Tatar, Karachay, Karakalpak, Tatar Arabic orthographies that are not encoded in Unicode standard are presented here.

2. Proposed Characters

Annotations below address special issues for a character, or reference to figures where such special issues are discussed. (These annotations are not intended to be retained in the character list when copied into the standard.)

Proposed codepoint	Glyph	F M I	Proposed name	Linguistic comment
061C	ﻱ	••ﻱ	ARABIC SIGN LOW ALEF • Bashkir, Tatar	<i>denotes the usage of back vowels in the word with dubious spelling in Yaña imlâ orthography</i>
08AD	ډ	ډ••	ARABIC LETTER DAL WITH THREE DOTS BELOW • Belarusian	<i>denotes the sound [dz]</i>
08AE	ڤ	ڤڤڤ	ARABIC LETTER SAD WITH THREE DOTS BELOW • Belarusian	<i>denotes the sound [ts]</i>
08AF	گ	گگگ	ARABIC LETTER GAF WITH INVERTED STROKE • Crimean Tatar, Chechen, Lak	<i>denotes the sound [j] in Crimean Tatar denotes the sound [k] in Chechen and Lak</i>
08B0	و	و••	ARABIC LETTER WAW WITH MADDAH ABOVE • Karachay-Balkar, Tatar	<i>denotes the sound [o] in Karachay orthography - Karachay denotes the sound [o:] for loanwords in Urta imlâ İdrisov's orthography - Tatar</i>
08B1	۶	۶••	ARABIC LETTER STRAIGHT WAW • Tatar	<i>denotes the sound [ʊ] in Urta imlâ Alparov's orthography</i>
08B2	ي	يبي	ARABIC LETTER FARSI YEH WITH HAMZA BELOW • Karachay-Balkar, Tatar	<i>denotes the sound [ʊ] in Karachay orthography</i>
08B3	ئ	ئئئ	ARABIC LETTER FARSI YEH WITH DAMMA ABOVE • Bashkir, Karakalpak, Tatar	<i>denotes the sounds [ʊ], [ə]</i>

3. Encoding Considerations

Decompositions

ARABIC LETTER FARSI YEH WITH DAMMA ABOVE, ARABIC LETTER FARSI YEH WITH HAMZA BELOW and ARABIC LETTER WAW WITH MADDAAH ABOVE are not decomposable.

Joining type and group for ArabicShaping.txt:

061C;LOW ALEF; U; No_Joining_Group
08AD;DAL WITH 3 DOTS BELOW; R; DAL
08AE;SAD WITH 3 DOTS BELOW; D; SAD
08AF;GAF WITH INVERTED STROKE; D; GAF
08B0;WAW WITH MADDAAH ABOVE; D; WAW
08B1;STRAIGHT WAW; R; STRAIGHT WAW
08B2;FARSI YEH WITH DAMMA ABOVE; D; FARSI YEH
08B3;FARSI YEH WITH HAMZA BELOW; D; FARSI YEH

Unicode character Properties:

061C; ARABIC SIGN LOW ALEF;Lo;0;AL;;;;;N;;;;;
08AD;ARABIC LETTER DAL WITH THREE DOTS BELOW;Lo;0;AL;;;;;N;;;;;
08AE;ARABIC LETTER SAD WITH THREE DOTS BELOW;Lo;0;AL;;;;;N;;;;;
08AF;ARABIC LETTER GAF WITH INVERTED STROKE;Lo;0;AL;;;;;N;;;;;
08B0;ARABIC LETTER WAW WITH MADDAAH ABOVE;Lo;0;AL;;;;;N;;;;;
08B1;ARABIC LETTER STRAIGHT WAW;Lo;0;AL;;;;;N;;;;;
08B2;ARABIC LETTER FARSI YEH WITH HAMZA BELOW;Lo;0;AL;;;;;N;;;;;
08B3;ARABIC LETTER FARSI YEH WITH DAMMA ABOVE;Lo;0;AL;;;;;N;;;;;

Collation:

Characters may be sorted after similar characters.

- U+061C ARABIC SIGN LOW ALEF should have secondary or tertiary sorting after (its appearance before a word should not affect sorting, it depends on the letters after (U+061c isn't a letter, but a sign before them). Only in the case when there are two words that totally coincide with each other with all letters and differs only where one has a U+061c in the beginning and the second does not should the first word be the one with U+061c (low alef) before it and the word without it be sorted afterward the second.
- U+08AD ARABIC LETTER DAL WITH THREE DOTS BELOW should sort after U+068D ARABIC LETTER DDAHAL
- U+08AE ARABIC LETTER SAD WITH THREE DOTS BELOW should sort after U+069D ARABIC LETTER SAD WITH TWO DOTS BELOW
- U+08AF ARABIC LETTER GAF WITH INVERTED STROKE should sort after U+06AF ARABIC LETTER GAF

- U+08B0 ARABIC LETTER WAW WITH MADDAH ABOVE and U+08B1 ARABIC LETTER STRAIGHT WAW should sort after U+06CB ARABIC LETTER VE
- U+08B2 ARABIC LETTER FARSI YEH WITH HAMZA BELOW and U+08B3 ARABIC LETTER FARSI YEH WITH DAMMA ABOVE should be sorted after Farsi yeh (U+06CC)

Confusability

Confusability issues would only arise if any of these characters were decomposed.

4. Usage and Encoding Rationale

Additions for Bashkir, Karakalpak, Tatar

The proposed letters were used in 1910s-1920s – in 1920s in Yaña imlâ (New Orthography) orthography and several experimental transition orthographies of 1910s colloquially named *Urta imlâ* (Middle Orthography) [1][2][4][5].

For the Tatar the majority of examples are stored here:

<http://fotki.yandex.ru/users/ievlampiev/album/64267>

<http://fotki.yandex.ru/users/ievlampiev/album/67014/> (also Bashkir)

<http://fotki.yandex.ru/users/ievlampiev/album/116321/>

See fig. 1 for Tatar alphabet, fig. 2 for Karakalpak alphabet.

Explanation of the Bashkir-Tatar Low Alef (based on ref [1][2]): Low Alef (ا; Tatar: qalınlıq bilgäse, калыңлык билгәсе, i.e. the hard sign) is a special character marking, that a word (or at least its first part) have back vowels. It was placed in the beginning of the words only.

Low alef isn't connected to other letters and occurs in initial and standalone positions only.

It works this way: the Tatar language has 5 pairs of vowels, one of the pair mates is a diphthong: ı - e [ы- э], i - ıy (ı) [и - ый], o-ö [о - ө], u-ü [у - ү], a-ä [а - ә]. In the Yaña imlâ alphabet, all of them used one vowel sign for each pair, except the last, i.e. Farsi yeh-damma (proposed code U+08B3), Farsi yeh U+06CC (ع), waw-damma U+06C7 (و), waw U+0648 (و), and paired alef U+0627 (ا) and Arabic ae U+06D5 (ا). Possibly, the last pair was signed by different characters due to many Arabic words that don't obey the vowel harmony law. Low alef was used in places where dubious reading might occur, for example, tor (stay) and tör (variety) without it. It was placed before the word that consisted of back vowels, i.e. in “tor” (تور) but not for “tör” (تور). See fig 3.

In some cases it was omitted: for words containing “q” (ق) or “ğ” (غ) - they usually have back vowels, or words, containing “a” (ا), as their other vowels are back due to vowel harmony law.

Rationale for encoding low alef as a separate character can be derived from 1) its form; 2) its usage.

- 1) Unlike the subscript alef (U+0656) low alef isn't placed below base characters, but is a standalone alef placed as subscript index. Moreover, subscript alef and low alef coexisted in the Tatar writing, but in the different orthographies (see fig. 4 & 5). Low alef appears in the last Arabic orthography Yaña imlâ, and the subscript alef appears in the transition Urta imlâ experimental orthographies, developed and used by many Tatar scholars until the Yaña imlâ was adopted. Subsequently they both should occur in the same font where they will have different glyph view. Also, subscript alef couldn't be used instead of low alef, as in any ordinary font it

will be shown as haraka – for historical usage it is better for reader to see the “box” than incorrect view.

- 2) Low alef has the similar behavior to high hamza (U+0674) used for Arabic in Central Asian languages. It occurs only in beginning of the word and doesn't interact with other letters of the word. Low alef is not haraka, and not a letter, it has another sense, similar to Central Asian high hamza. See Fig. 3, 9.

Rationale for encoding Farsi yeh-damma as a single code instead of using any proposed code-sequences:

- 1) Main: In its initial and medial form this character has no dots (nuqta) below, making a significant difference in rendering. However, in the majority of standard fonts any proposed sequence will have dots below, as there is no yeh-based letters without dots in the initial form (excluding yeh-hamza U+0626). However, for historical usage it will be better to see no character than wrong. See fig. 9.
- 2) View details 1: Also, the absence of dots is a sufficient for this letter – instead of them another mark (damma or comma) is used. Only in Naskh based fonts it is a typical damma, however for some popular fonts used that time it has a comma, and is similar to nun with comma instead of dot. See fig. 4.
- 3) View details 2: The damma or comma in this character doesn't obey the standard behavior of harakas for the Tatar language. Tatar language used harakas for the Arabic loanwords in *İske* and *Urta imlâs*. In *Urta imlâ* harakas coexisted with Farsi yeh-damma. The printing standard of 1910s made high harakas be placed at the same high for all characters and relatively old computer fonts with such behavior (like Times New Roman) are perfect to display a Tatar text except the proposed letter. Even modern fonts, where harakas are shown at the height depending on the base letter's height, place the damma at the same height for the final, medial, standalone and initial yeh or any yeh-based. In the Tatar yeh-damma character the damma behaves in the same way as the dots of yeh or hamza in yeh-hamza (U+0626), i.e. it visibly changes its placement depending on position. See fig. 9.
- 4) Possibility to represent printing technical level of 1910s: as in the Tatar printing harakas and yeh-damma were placed in the different lines, it should be a technical possibility to place any haraka above or below the Farsi yeh-damma. It also should be a possibility to place damma above any letter (even selected to be a basement for code sequence to represent Farsi yeh-damma) with damma rendered at higher position than in yeh-damma.
- 5) Usage 1: the dotless damma-yeh is a separate Tatar letter, used in *Yaña* and *Urta imlâ*. It's damma is a modifier that makes sense of short vowel comparing to sound signed by Farsi yeh, as it actually does in short Uighur U (U+06C7) and long waw (U+0648). In *Urta imlâ* it may coexist with harakas above and below. Farsi damma-yeh is not decomposable.
- 6) Usage 2: Farsi yeh-damma behaved in the similar way, first, as yeh-hamza (which has no dots below) U+0626 and, second, as Uighur U (U+06C7), all three are used in Bashkir and Tatar alphabet as letters. In the last two letters damma and hamza have a sense of modifier, not haraka. Those characters perfectly fit the usage of *Urta* and *Yaña imlâ* in the Tatar language – they are opposed to sequences with harakas, that was also used in the Tatar language in the same orthography for actual harakas in Oriental loanwords and names, as their modifiers behave not like haraka. See fig 5, 6.

Additions for Crimean Tatar, Chechen and Lak

The proposed letter was used in Crimean Tatar to designate [j] (fig. 14) in some orthographies and for Chechen and Lak to designate [kʼ] (fig 15, 16) in 1920s.

For Crimean Tatar usage examples see <http://fotki.yandex.ru/users/ievlampiev/album/67179/> .

Additions for Tatar and Karachay

The proposed letters were used in 1920-1926 orthography of Karachay alphabet. See fig. 11, 12.

Rationale for encoding Farsi yeh-hamza below as a single code instead of using any proposed code-sequences:

- 1) In its initial and medial form this character has no dots below. However, in the majority of standard fonts any proposed sequence will have dots below, as there is no yeh-based letters without dots in the initial form (excluding yeh-hamza U+0626). However, for historical usage it will be better to see no character than wrong. Also, in its final form hamza crosses the final yeh base. Hamza placement doesn't depend on base's form.
- 2) Here hamza has no sense of hamza, but it is a modifier only used to sign another vowel than signed by yeh letter in Karachay language. Yeh-hamza below is not decomposable.

Rationale for encoding waw-maddah above as a single code instead of using code-sequences U+0648 U+0653 (ّ).

- 1) Karachay has no harakas. Maddah is used as modifier to denote new letter. Also, maddah is placed rather low, in the majority of standard fonts alternative sequence results in high maddah above waw. However, for historical usage it will be better to see no character than wrong.
- 2) In Tatar this waw-maddah also considered to be a single letter, like in Karachay it is not decomposable. See fig 13. For Tatar, where the usage of haraka and modifiers coexisted, it is noteworthy to preserve a possibility to represent printing technical level of 1910s: as in the Tatar printing harakas and waw-maddah were placed in the different lines, it should be a technical possibility to place any haraka above or below the waw-maddah. It also should be a possibility to place maddah above any letter (even waw) with maddah rendered at higher position than in Tatar waw-maddah.

Additions for Belarusian

The proposed letters were used in handwriting in the 16th-19th centuries for writing Belarusian and Polish among the Belarusian Tatars. In the 20th century it was used in printing among scholars studying this alphabet [3]. See fig 7, 10.

Examples of texts are found at:

<http://fotki.yandex.ru/users/ievlampiev/album/64978>

5. Acknowledgements

Many thanks to Tatarstan National Library workers for access for copying some texts.
Special thanks to Lorna Priest for valuable help in composing this proposal.

6. References

- [1] See <http://fotki.yandex.ru/users/ievlampiev/album/116321/> (2nd and 3rd lines form title an <http://fotki.yandex.ru/users/ievlampiev/view/403389?page=0> Русь мактяблярэ учюн татар тлэ дярслегу, Казань, 1925)
- [2] Bashkir grammar (р .III; р. 9) found on: <http://fotki.yandex.ru/users/ievlampiev/album/67014/>
- [3] Антонович А.К. Белорусские тексты, писанные арабским письмом, и их графико-орфографическая система. Вильнюс, 1968. <http://fotki.yandex.ru/users/ievlampiev/album/64978/>
- [4] Курбатов, Хәлиф Рәхим улы. Татар теленең алфавит һәм орфография тарихы / Х.Р. Курбатов.—Казан: Татарстан китап нәшр., 1960.—131 б.; 23 .—
<URL:http://z3950.ksu.ru/knigi/X_Ourbatov_Tatar_teleneng_alfavit_ham_orfografia_tarixi.djvu>.
- [5] Әхмәров К. З. Башкорт ызыуы тарихенән = Из истории башкирской письменности: (башкорт әз әби теленең алфавите һәм орфографияһы тарихе) / проф. К. З. Әхмәров; яуаплы ред. Ж. Ғ. Кйекбаев.—Өфө: Башкортостан китап нәшр. , 1972.—132, [2] б.; 21.—Тит. л. парал. рус.— Библиогр.: с. 129-133 и в подстроч. примеч. <http://fotki.yandex.ru/users/ievlampiev/album/67013/>

7. Examples and Figures

Fig. 1. Tatar alphabet sorting from [1]

ئا، (ا)، ئه، (ه)، (ا)، ب، پ، ت، ج، چ، ح، د، ر، ز، ژ،
س، ش، ع، ف، ق، ك، گ، گ، ل، م، ن، ئو، (و)، ئو، (و)،
ؤ، ه، د، (د)، ئي، (ي).

Fig. 2. Karakalpak alphabet from Karakalpak primer [Nurlan Joomagueldinov's scans collection with no titles]

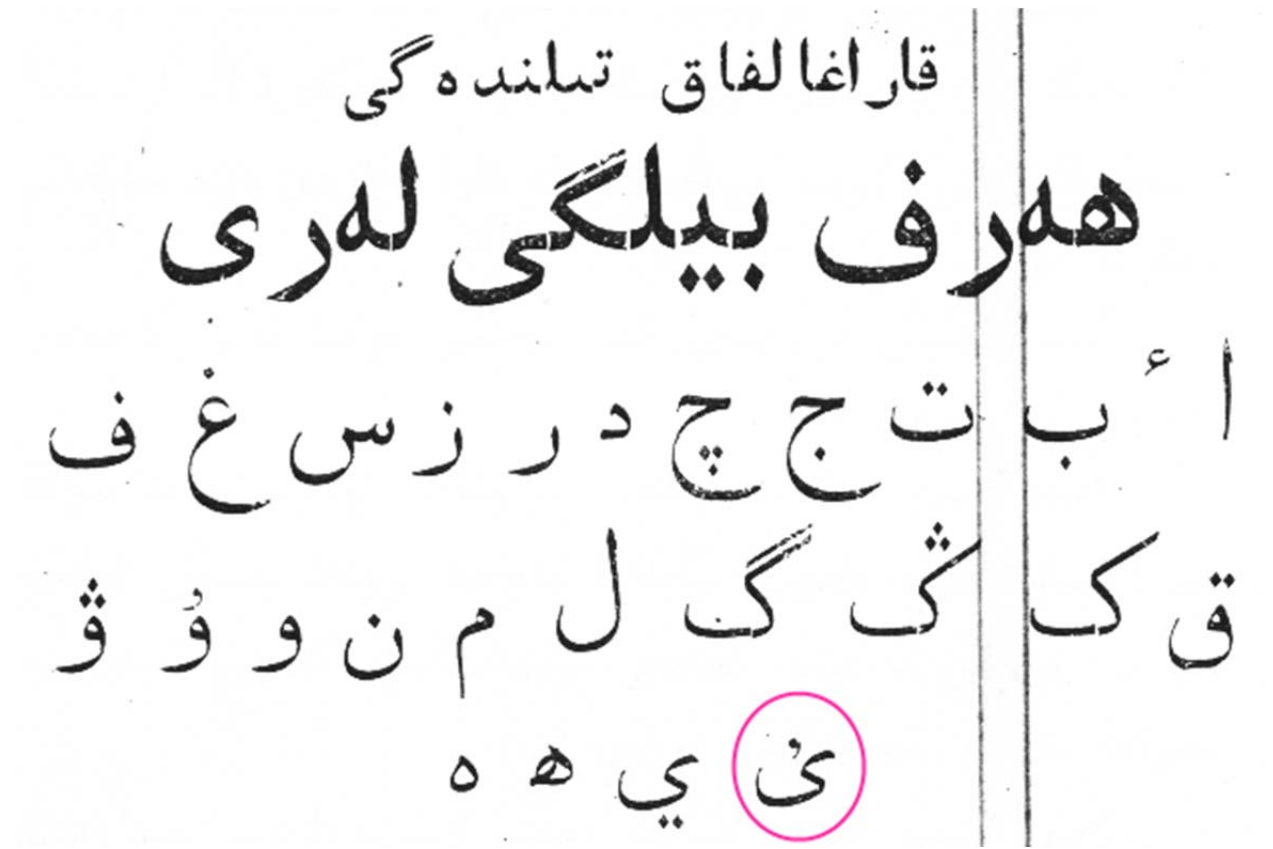


Fig. 3. Usage of words with low alef and without it [1]

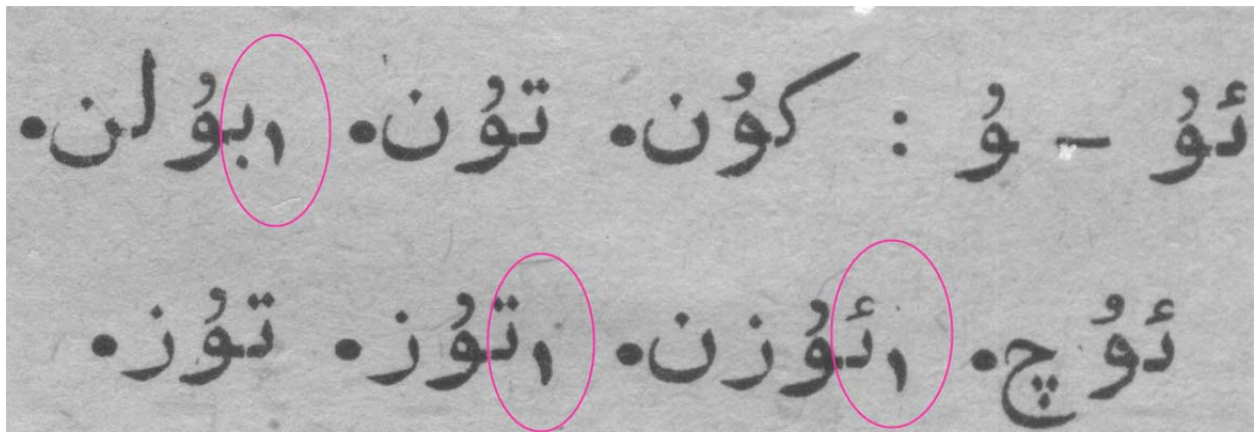


Fig. 4. Non-Naskh style of Farsi yeh-damma [1].

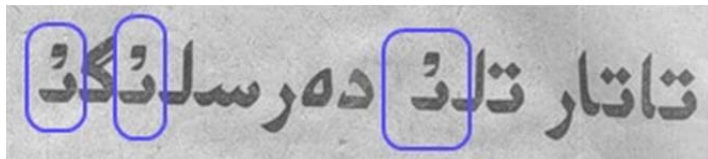


Fig. 5. Usage of Farsi yeh-damma with subscript alef [5].

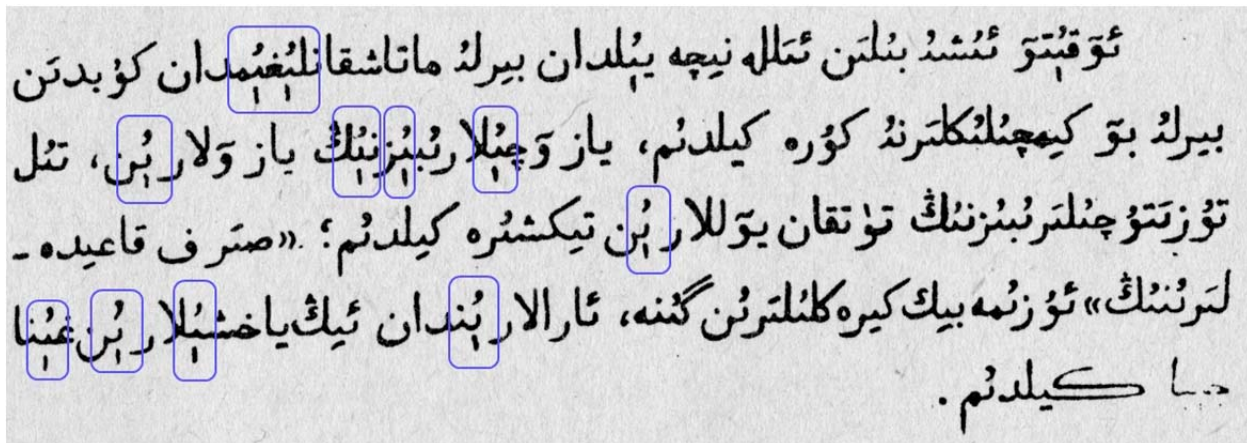


Fig. 6. Simultaneous usage of harakas (pink frames) and modifier dammas over yeh (blue frames). [See book title on <http://fotki.yandex.ru/users/ievlampiev/view/141416?page=6>]

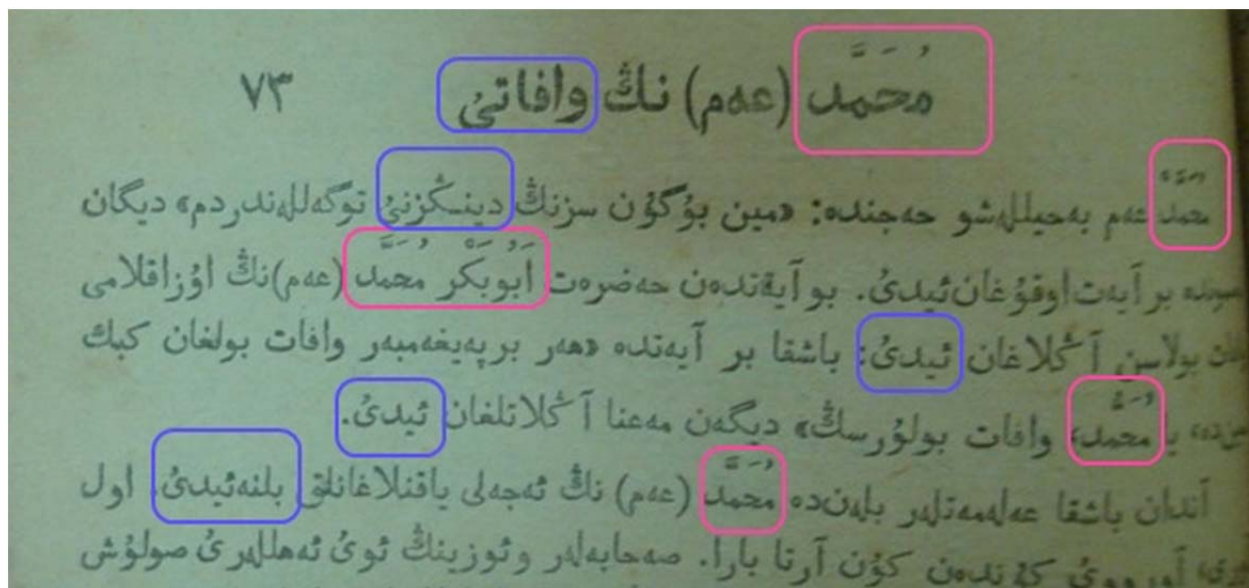


Fig. 7. Usage of Belarusian Arabic. Proposed letters are in frames. [3]

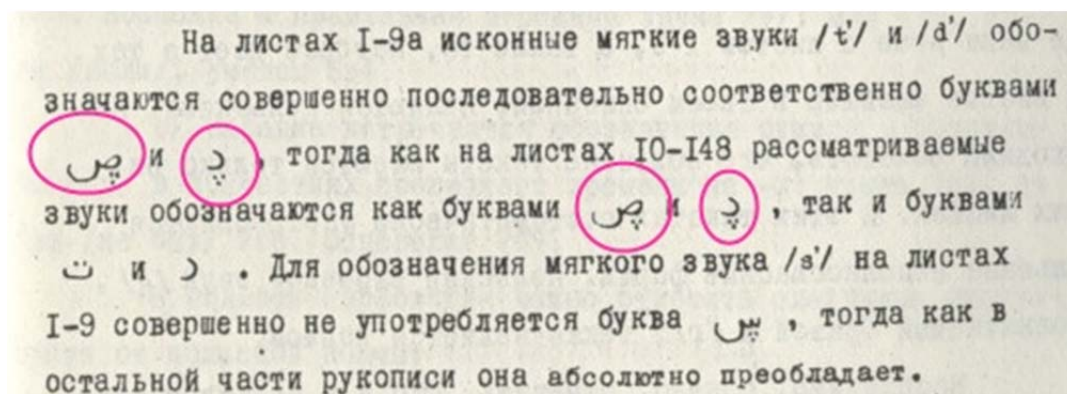


Fig. 9. Usage of low alef and Farsi yeh-damma in the Bashkir language [2]

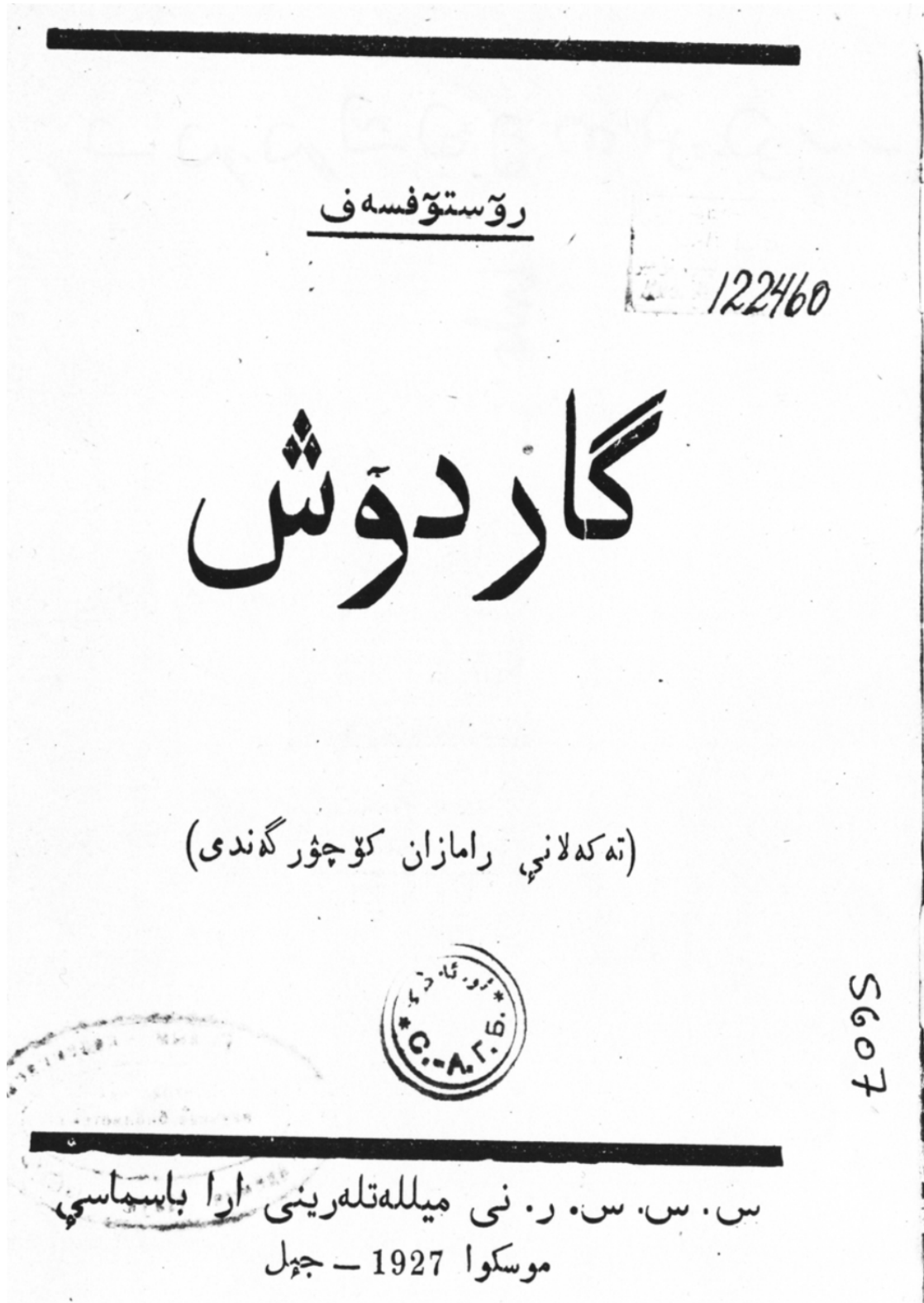
بىگە بىيل عىلمى مەرکەز تارافىنان باشقۇرت تىلنىڭ سارف،
 نەحوەن ياذب ئۇلگرتو تاشىرلعاينى. ھەزرگە سارف ئۇلۇشنى ياذب
 تەمام ئىتىدك. باشقۇرت تىلنى سارف ياذو بىرنىسى تەزرىبە بولۇوى،
 ئول تورالا بر تۇرلئە قوللانما، ماتىرىيال بولماوى ھەم باشقا سەبەبتەر
 تىلەكتى تولى كۆپى ئوتەوگە يول بىرمەنى.
 مەكتەبتەر ئۇسۇن تەقدىم ئىتىلگەن سارف كىتابى فائىتىگە

ھۈيلەم يەكى ھوز تۇركۇمدەرى بولھا ئىبتەشتەر ئىسنىنەن برەوھى
 بولماھا ئىكنىسى برەوھى شول ھوزدى ئاگلاب، قالعان ئىبتەشتەرلئە
 تۇشۇندۇر ب ھۈيلەب بىرە ئالا. بر كىشى ياكىد ئەذرلەگەن دەرىستەن

Fig. 10. Usage of proposed letters in Belarusian text [Кітабы — унікальная зьява ў беларускай мове/ В. І. Несьцяровіч http://www.pravapis.org/art_kitabl.asp].

تىرئفوق ایدى كوز نوبى كوز مذب • خروشا اوتىم ایت دجى ورمىدى •
 صمنا بلو اىچ تىرئفوق نوبى • قوز بىل صاباى نىرحال
 يهور يفاز بىل ائردى ظلمات • سانس كىر تىرئفوق قىامت •
 دوزچ باشو صمنا بلو • وىسر وىران
 — Адрывак лягенды «Мэрадж» з «Аль-Кітабу».

Fig. 11. Usage of Karachay letters [see title on the image].



گار دۆش

اني بيتديرمه كليك، جېيماقلىق،
ساقلاماقلىق، ساتماقلىق.

خالقني مؤلكؤنده گاردۆشنو بهرگهن خايري،
گاردۆشنو قايدان چيققاني. قالغان خانسلادان
گاردۆشنو بيتيميني نه باشخاليجي باردي؟ گاردۆشنو
بيتيمي بلا اني تاميرلاري.

بو ساغاتدا گاردۆشچا خار قايدادا جايلغان باشخا
خانس تابارغا بهك قييندي. بو ساغاتدا گاردۆش بيتمه گهن
گاردۆش سالينماغان دؤنياني باشيندا جهرجو قودوده گهن ادام
انتينا ۋاللوك تويۇلدۇ. گاردۆشنو بېلاي جايلغانينا سەير-
سەينيرگه كهرهك تويۇلدۇ، نه ۋچۇن دەسەك، - ادامغا جاراغان
خانسلاني باريسيندанда گاردۆشدان خايرلي زات چوقدو.
سووق، جېلي دەب بۆلماين گاردۆش قېيلا جانيندا ئەم
جېلي جهرله دەن باشلاب كؤن تيممه گهن جانيندا ئەم سووق
جهرلەنى باريسيندادا بيتەدى. سېتاب قاراساك گاردۆش
بيتمه گهن جهر چوقدو:

Fig. 12. Table of Karachay alphabets [Литературная Энциклопедия, Т. 5 / Ком. Акад.; Секция лит., искусства и яз.; Ред. коллегия: Лебедев-Полянский П. И., Маца И. Л., Нусинов И. М., Скрыпник И. А. Фриче В. М.; Отв. ред. Луначарский А. В.; Отв. секретарь Бескин О. М. —[б.м.]: Изд-во Ком. Акад., 1931. — [IV], 784 стб.: ил. — Карачаевская литература, автор Ислам Карачайлы (<http://feb-web.ru/feb/litenc/encyclp/le5/le5-1171.htm?cmd=2&istext=1>)].

На араб-ской основе	НТА	На рус-ской основе	Примерное произношение
ا	а	а	а
ب	в	б	б
پ	р	п	п
ت	т	т	т (с придыханием)
ڤ	ç	дч	в карач.—англ. дж; в балк.—дз
س	с	ч	в карач.—г; балк.—ц
خ	х	х	х
د	д	д	д
ر	р	р	р
ز	з	з	з
ژ	z	ж	жь
س	с	с	с
ش	ş	ш	шь
ڭ	г	г	глубокое (проточное) г
ف	ф	ф	ф
ق	к	к	глубокое (заднее) к
ك	к	к	к
گ	г	г	г
ڭ	нг	нг	н (заднеязычное), нем. ng — lang
ل	л	л	л, ль
م	м	м	м
ن	н	н	н
ه	h	г	нем. h
و	и	у	у
و	у	ю	нем. ü
و	о	о	о
و	ө	ё	нем. ö
و	е	е	е
ي	и	и	и
ي	ј	я	я
ي	ь	ы	звук типа ы

Fig. 13. Usage of waw-maddah in Tatar: Map of agricultural and meteorology stations in Tatarstan, 1922
<http://karta.turizmkazan.ru/tatarstan/raritet/tatarstan-starinnaya-karta-tatrskiy-yazik-arabskiy-shrift>

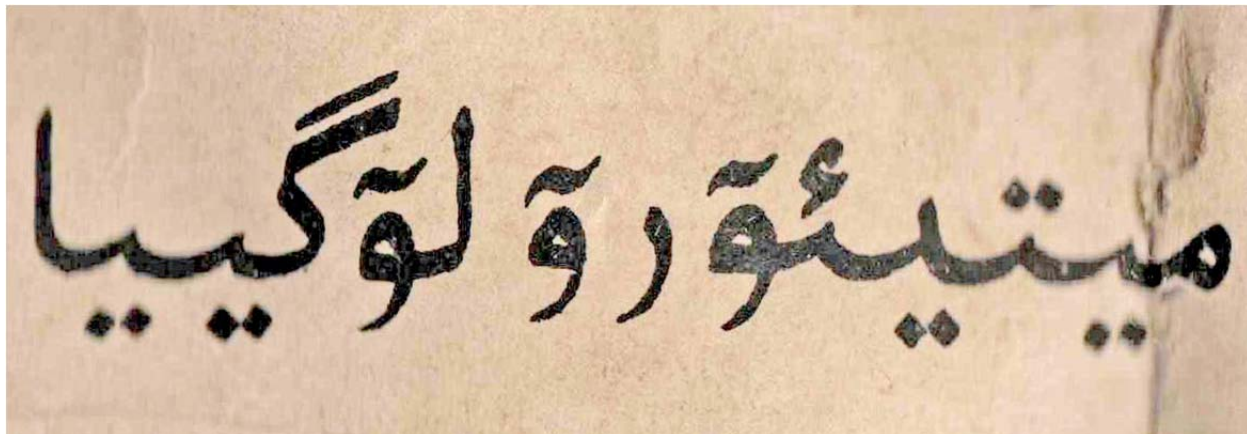


Fig. 14. Page of a Crimean Tatar primer for reading Qur'an [title is here
<http://fotki.yandex.ru/users/ievlampiev/view/158092?page=0>]

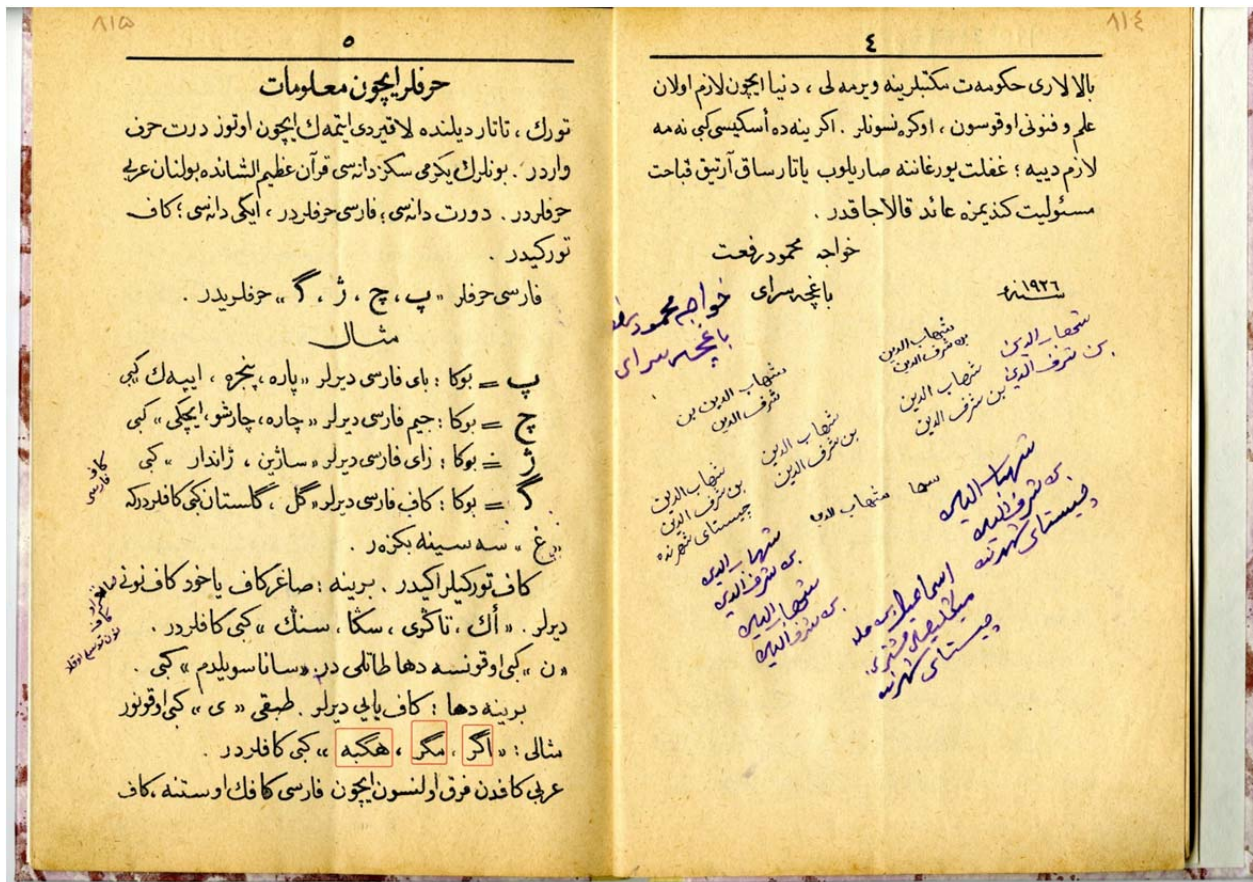


Fig. 15. Title of Chechen primer and its p. 42 [title is on the image,
<http://fotki.yandex.ru/users/ievlampiev/album/66437/>]

۴۲۶

<p>گور گور گومور گیر گیرا گهل گهیدی</p> <div style="display: flex; justify-content: space-around;">   </div> <p>گودول حاجگی</p> <p>دآدآ دآوردآ باخو وائی گاور گهل گهیدی یو ، عائشة نانا یوقغو گودولایخ خی دالماش ، ما گه بنار دالمایخی ، دآورد گیرانا ماسا ده دو؟ وور ده دو ، بهتیق ماسا گیرا دو؟ دی گهیرانا شیخ</p>	  <p>گازا بوچیگ دیگ</p> <p>عایشات - عائشة ! شون یاخان یاخانآ گازلا - بوچیگنا یهئانا ، بوچیگینا یائان همیانا لوی گازلا اووژلا ، شول لوق امدانا ، هوندا لوسور - الله امدادی شول مالا .</p> <p>بوچیگ بو ---</p> <p>گ گ گ گ</p> <p>گا کاخ گانت گادیی</p>
---	---

Fig. 16. Lak Arabic alphabet table [Adyge alphabet table, retrieved from Wikipedia 2010-05-08.

Accompanying information:

Lak arabic alphabet.JPG

English: Lak arabic alphabet from 1925 book

Source: Букварь на лакском языке. Primer auf Lak. Буйнакск, 1925 Buinaksk, 1925]

ا	ب	ت	پ	ج	ح	خ	چ
ج	د	ز	ر	س	ش	پس	ر
ط	ع	غ	ف	ق	ک	گ	گ
ک	ل	م	ن	و	ه	ی	او
ای	ث	ص	ض	ذ	ظ	ث	خ

(۱) وای قعور اگو حارف لاکو مازراو قانایستار عاراب
رایستارا چیچین داقا .
(۲) هارپا عاین جالاستا موقو (۴) بیشایستاکوما مثلا : خوزو -
قعورو هارپا بوقواستا موقول یا لکو (۵) دیشایستار مثلا : کارو - کارو .

**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://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 Arabic characters used for Bashkir, Belarusian, Crimean Tatar, Karachay, Karakalpak, and Tatar languages*
2. Requester's name: *Ilya Yevlampiev, Karl Pentzlin, Nurlan Joomagueldinov*
3. Requester type (Member body/Liaison/Individual contribution): *Expert Contribution*
4. Submission date: *28 April 2011*
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: *No*

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: *Arabic and Arabic Extended-A*
2. Number of characters in proposal: *8*
3. Proposed category (select one from below - see section 2.2 of P&P document):
- | | | | | | |
|---------------------------------------|--------------------------|---|-------------------------------------|------------------------------------|--------------------------|
| A-Contemporary | <input type="checkbox"/> | B.1-Specialized (small collection) | <input checked="" type="checkbox"/> | B.2-Specialized (large collection) | <input type="checkbox"/> |
| C-Major extinct | <input type="checkbox"/> | D-Attested extinct | <input type="checkbox"/> | E-Minor extinct | <input type="checkbox"/> |
| F-Archaic Hieroglyphic or Ideographic | <input type="checkbox"/> | 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? *Lorna Priest, SIL International*
- b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):
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*
Sorting and linguistic representations are discussed in the proposal

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

¹ Form number: N3702-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)

inclusion in the Unicode Standard.

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before? If YES explain	<i>No</i>
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:	<i>Yes</i> <i>Linguists, librarians</i> <i>See examples in proposal</i>
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Reference:	<i>once used by 3 million people</i>
4. The context of use for the proposed characters (type of use; common or rare) Reference:	<i>Historically common</i>
5. Are the proposed characters in current use by the user community? If YES, where? Reference:	<i>Historical use</i>
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:	<i>Yes</i> <i>Should be placed with similar characters</i>
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	
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:	<i>No</i>
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:	<i>No</i> <i>Discussion in proposal</i>
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:	<i>No</i>
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:	<i>No</i>
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)	<i>No</i>
13. Does the proposal contain any Ideographic compatibility character(s)? If YES, is the equivalent corresponding unified ideographic character(s) identified? If YES, reference:	<i>No</i>