

To: UTC

Title: Request for glyph changes and annotations for Kazakh, Kyrgyz, and Uyghur

From: Lorna Priest Evans (SIL International)

Date: 7 December 2020

## Introduction

Kazakh, Kyrgyz, and Uyghur have a number of issues related to display of certain characters in the Arabic script. Some of them can be handled by adding annotations and some of them by changing the design of some of the glyphs. Most of these characters are not used by any other languages. This document is a request for added annotations and/or glyph changes to these characters.

The characters to be discussed are U+0626, U+0641, U+0674..U+0678, U+06C5, and U+06C7.

## Kazakh

The Kazakh [kaz] language is spoken by over 13 million people in Kazakhstan and in at least 11 other countries. Kazakh can be written with the Arabic, Cyrillic, and Latin script (script use is generally based on the country of use) [1].

The CLDR [2] lists these Exemplar Characters for Kazakh in Arabic script:

◌ُ	◌ُ	ئ	ا	ب	پ	ت	ج	چ	ح	د	ر	ز	س	ش	ع	ف
064F	0654	0626	0627	0628	067E	062A	062C	0686	062D	062F	0631	0632	0633	0634	0639	0641
ق	ك	ڭ	گ	ل	م	ن	ه	ه	و	و	و	و	و	و	ى	ي
0642	0643	06AD	06AF	0644	0645	0646	06BE	06D5	0648	0676	06C6	06C7	0677	06CB	0649	064A

**Note about exemplars:** Kazakh does not actually use U+064F, U+0654, nor U+0626 (colored red). Kazakh *does* use U+0674, U+0675, and U+0678. In practice, however, people do not use U+0675..U+0678; instead they use U+0674 to prefix the word. As indicated in the CLDR, Kazakh does use U+0641 (ف), however, fonts for Kazakh use a shape similar to U+06A7 (ف).

## Kyrgyz

The Kyrgyz [kir] language is spoken by 4-5 million people in Kyrgystan, Afghanistan, China, Kazakhstan, Russia, Tajikistan, Turkey, and Uzbekistan. Kyrgyz can be written with the Arabic, Cyrillic, and Latin script (script use is generally based on the country of use) [1].

The CLDR [2] lists these Exemplar Characters for Kyrgyz in Arabic script (with status of unconfirmed):

ئ	ا	ب	پ	ت	ج	چ	ح	خ	د	ر	ز	س	ش	ع	
0626	0627	0628	067E	062A	062C	0686	062D	062E	062F	0631	0632	0633	0634	0639	
ق	ف	ك	ڭ	گ	ل	م	ن	ه	و	و	و	و	و	ى	ي
0642	06A7	0643	06AD	06AF	0644	0645	0646	06D5	0648	06C5	06C7	06C9	06CB	0649	064A

**Note about exemplars:** As indicated in Figures 1a and 1b, U+062E is not actually used in Kyrgyz. Also, in practice, people do not use U+06A7, they use the codepoint U+0641. However, their fonts use a shape similar to U+06A7.

## Uyghur

The Uyghur [uig] language is spoken by over 10 million people in China and in at least 10 other countries. Uyghur can be written with the Arabic, Cyrillic, and Latin script (script use is generally based on the country of use) [1].

The CLDR [2] lists these Exemplar Characters for Uyghur in Arabic script:

ئا	ئە	ئو	ئۇ	ئۆ	ئې	ئى									
0626 0627	0626 0647	0626 0648	0626 06C7	0626 06C6	0626 06C8	0626 06D0	0626 0649								
ا	ە	ب	پ	ت	ج	چ	خ	د	ر	ز	ژ	س	ش	غ	ف
0627	06D5	0628	067E	062A	062C	0686	062E	062F	0631	0632	0698	0633	0634	063A	0641
ق	ك	گ	ڭ	ل	م	ن	ھ	و	ۇ	ۆ	ۈ	ۋ	ې	ى	ي
0642	0643	06AF	06AD	0644	0645	0646	06BE	0648	06C7	06C6	06C8	06CB	06D0	0649	064A

**Note about exemplars:** As indicated in the CLDR, Uyghur does use U+0641 (ف), however, fonts for Uyghur use a shape similar to U+06A7 (ف).

## U+0626 ARABIC LETTER YEH WITH HAMZA ABOVE

See Figures 1a, 1b, 1c, 2, 3, 12, 13, 16

### Visual difference

In the standard display of YEH WITH HAMZA ABOVE, the *hamza* sits above the bowl of the YEH (isolate and final) like this: ئىئىئى

Kyrgyz language users prefer the *hamza* to be above the top of the stem (isolate and final), like this: ئىئىئى

### Historical implementations

The Kyrgyz have tended to use Uyghur fonts because they often contain all the required characters need for Kyrgyz. In the early 2000's, the fonts used in Xinjiang were made by various local companies including Uighursoft. Some of these fonts included the needed vowels for Kyrgyz and Kazakh, but since Uyghur only uses the *hamza* in word initial or medial positions, U+0626 did not ligate to the final or isolate forms:

ئىئىئى دۇ

This was not seen as a huge problem because the final and isolate forms of U+0626 are extremely rare in Kyrgyz. In cases where it does occur, a workaround was to add a non-phonetic ALEF MAKSURA (U+0649) to make sure words like “наби” showed up as:

ئابى and not ئابى

### Current usage

When the official orthographic dictionary was produced for the Kyrgyz language in Xinjiang (Figures 1a, 1b, 1c), however, such an approach was not acceptable. Both in the chart and in the dictionary, words ending in the YEH WITH HAMZA ABOVE (U+0626) are shown with all four forms. The *hamza* is positioned above the top of the stem.

## Desired outcome

An annotation should be added to U+0626.

0626 عى ARABIC LETTER YEH WITH HAMZA ABOVE

≡ 064A عي 0654 عْ

- in Kyrgyz the hamza is consistently positioned to the top right in isolate and final forms

There are compatibility characters (U+FE89 and U+FE8A). However, since the design of the default is not changing the compatibility characters do not need modifying.

Additionally, the core spec should be updated to illustrate what this would look like. In Chapter 9, under “Other Letters for Extended Arabic”, a paragraph should be added to describe this character variant. Suggested text:

U+0626 ARABIC LETTER YEH WITH HAMZA ABOVE normally has the *hamza* positioned over the bowl of the glyph in isolate and final forms. For the Kyrgyz language the *hamza* is positioned at the top right of the glyph in isolate and final forms.

Table 9.x

Default Position	Kyrgyz Language
	

## U+0641 ARABIC LETTER FEH

See Figures 1a, 1b, 6, 8, 11, 12, 13, 14, 16

The issues regarding this character will be addressed in another document.

## U+06C5 ARABIC LETTER KIRGHIZ OE

KIRGHIZ OE was added specifically for the Kyrgyz language. It is unknown if there are any other languages using this character. No other languages list U+06C5 in exemplars in CLDR.

See Figures 1a, 1b, 2, 3, 4, 12, 13, 15, 16

## Visual difference

There are two shapes that have been used for U+06C5. The one in the Unicode charts uses a bar and is similar to the middle glyph in the graphic below. There is also a glyph with a loop instead of a bar which is more common in current usage (left glyph). The below graphic demonstrates the two styles of KIRGHIZ OE (left two glyphs) alongside a standard WAW (on the right).



## Historical implementations and current usage

Originally the left glyph was used in handwriting or sometimes in stylistic fonts, while the center glyph was used in printing from 1982 until 1994 (and at least one example of something printed in 1998 uses the middle glyph: it depended on the specific publisher and typesetter).

Printing has tended to use the left glyph since 1995 (after the introduction of electronic typesetting).

Kyrgyz and Uyghur are both Turkic languages [1]. Many more fonts support Uyghur, so about 155 fonts supporting the **Uyghur** language were examined. Most of these were produced in country. Less than half have the Kyrgyz characters (especially KIRGHIZ OE which is not used for Uyghur, Kazakh or any other language that we know of) which is unsurprising, since most people are only concerned about writing their own language.

Of those fonts which included KIRGHIZ OE, about **half have the bar** and **half have the loop** (none of those produced outside the country have the loop shape e.g. Arial, Courier New, Geeza Pro, Times New Roman etc.).

## Desired outcome

We believe the glyph shape in the Unicode charts for U+06C5 should be changed to a loop rather than a bar.

An annotation should be added:

06C5     ء    ARABIC LETTER KIRGHIZ OE  
• a barred form also occurs

The compatibility characters should also have their glyphs changed (U+FBE0 and U+FBE1).

Additionally, Arabic Shaping for U+06C5 should be changed from WAW WITH BAR to WAW WITH LOOP.

## U+0677 ARABIC LETTER U WITH HAMZA ABOVE and U+06C7 ARABIC LETTER U

See Figures 1a, 1b, 1c, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16

## Visual difference

The Unicode charts display both U+0677 and U+06C7 with a standard *damma* shape: ؤ ؤ

Kazakh, Kyrgyz, and Uyghur require the *damma* (it should actually not even be associated with a *damma* character) to appear more like a comma where the loop is filled in and the design is more upright. ؤ ؤ

Microsoft Sans Serif and Microsoft Uighur do use the appropriate shape for U+06C7 (but not for U+0677): ؤ ؤ

## Desired outcome

We believe the glyph shape for U+0677 and U+06C7 should be changed to be a filled in comma form (not a *damma*).

Additionally, an annotation should be added for U+06C7:

06C7     ؛    ARABIC LETTER U  
• Azerbaijani, Kazakh, Kyrgyz, Uyghur

The compatibility characters should also have their glyphs changed (U+FBD7, U+FBD8, U+FBDD) to use the comma form.

In order to disassociate these glyphs from the *damma*, the Arabic Shaping names for U+0677 and U+06C7 should change from “WITH DAMMA ABOVE” to “WITH COMMA ABOVE.”

## U+0674 ARABIC LETTER HIGH HAMZA and associated characters (U+0675..U+0678)

See Figures 6, 7, 8, 9, 12, 15, 16

## Visual difference

U+0674 and associated characters U+0675..U+0678 are often designed with the *high hamza* a bit too high (as they are in the Unicode charts). It is also important to remember the high hamza takes up horizontal space. For U+0675..U+0678, the high hamza should not sit on top of any part of the main character; it should always be to the right of it and perhaps a little lower than the top of an ALEF. This is clearly visible in figures 6, 7, 8, 9, 12, 15, 16.

This is how one popular font positions the high hamza at all different heights:

أؤؤى

These characters should all have the high hamza at the exact same vertical position:

أؤؤى

U+0675..U+0678 are only used in isolate or initial positions, but their properties allow for medial or final positions as well.

In practice, people usually use U+0674 followed by the base character.

## Properties

### Compatibility Decompositions

The compatibility decompositions for these characters seem to treat U+0674 as a combining mark rather than a spacing character (placed after the base character).

```
0675;ARABIC LETTER HIGH HAMZA ALEF;Lo;0;AL;<compat> 0627 0674;;;N;ARABIC LETTER HIGH HAMZAH ALEF;;;
```

```
0676;ARABIC LETTER HIGH HAMZA WAW;Lo;0;AL;<compat> 0648 0674;;;N;ARABIC LETTER HIGH HAMZAH WAW;;;;
```

```
0677;ARABIC LETTER U WITH HAMZA ABOVE;Lo;0;AL;<compat> 06C7 0674;;;N;ARABIC LETTER HIGH HAMZAH WAW WITH DAMMAH;;;;
```

```
0678;ARABIC LETTER HIGH HAMZA YEH;Lo;0;AL;<compat> 064A 0674;;;N;ARABIC LETTER HIGH HAMZAH YA;;;;
```

### Desired outcome

We believe the glyph shapes for U+0674 and associated characters U+0675..U+0678 should be changed to all have the *high hamza* at the same vertical position.

A good model is to position the *high hamza* beside the top of the *alef* (not above) as in Figure 8.

The glyph for the compatibility character U+FBDD will be changed as a result of the section on U+0677 ARABIC LETTER U WITH HAMZA ABOVE and U+06C7 ARABIC LETTER U above.

Since U+0675..U+0678 are only used in isolate or initial positions, but their properties allow for medial or final positions as well, and in practice people usually use U+0674, it seems proper to declare U+0675..U+0678 as “discouraged”. If they are discouraged, the core spec should be updated. Suggested text:

Because of the compatibility decompositions for U+0675 ARABIC LETTER HIGH HAMZA ALEF, U+0676 ARABIC LETTER HIGH HAMZA WAW, U+0677 ARABIC LETTER U WITH HAMZA ABOVE, and U+0678 ARABIC LETTER HIGH HAMZA YEH these characters are not recommended for use. Instead, U+0674 ARABIC LETTER HIGH HAMZA followed by the base character should be used.

These characters (U+0675..U+0678) should have a heading marked “Discouraged”. Additionally, since they are now discouraged, and in order to avoid confusion, the annotation “Kazakh” should be removed.

## Discouraged

*Use of these characters is discouraged.*

0675 ARABIC LETTER HIGH HAMZA ALEF  
≈ 0627 0674

0676 ARABIC LETTER HIGH HAMZA WAW  
≈ 0648 0674

0677 ARABIC LETTER U WITH HAMZA ABOVE  
≈ 06C7 0674

0678 ARABIC LETTER HIGH HAMZA YEH  
≈ 064A 0674

Because the preferred order of representation of these sequences in text differs from the compatibility decompositions for U+0675..U+0678, there are implications for the Unicode Collation Algorithm. The weights in DUCET for U+0675..U+0678 should be adjusted.

The following Action Item could be closed: 125-A66 | Roozbeh Pournader, Liang Hai | Document recommendation number 9 (about high hamza) of L2/10-455 more thoroughly.

Based on a recent email communication on the Unicode list [3], “Jawi” should also be added to the annotation for U+0674 (this action has already been done).

- Kazakh, [Jawi](#)

## Samples

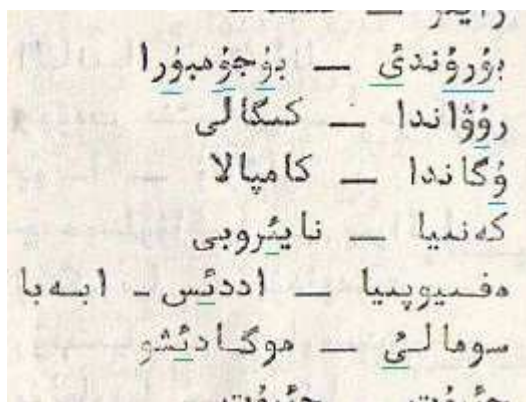
قمرغيز جازۇۇسۇنۇن المپبه تارتىبى جانا تامعالاردىن اتالمىشى

تارتىبى	جههكه كه لىشى	باشتا كه لىشى	ورتودو كه لىشى	اياقتا كه لىشى	اتالمىشى	تارتىبى	جههكه كه لىشى	باشتا كه لىشى	ورتودو كه لىشى	اياقتا كه لىشى	اتالمىشى
1	ا	ا	ا	ا	ا	16	م	م	م	م	م
2	ب	ب	ب	ب	ب	17	و	و	و	و	و
3	پ	پ	پ	پ	پ	18	ف	ف	ف	ف	ف
4	ن	ن	ن	ن	ن	19	ۇ	ۇ	ۇ	ۇ	ۇ
5	ت	ت	ت	ت	ت	20	ۋ	ۋ	ۋ	ۋ	ۋ
6	چ	چ	چ	چ	چ	21	ۋ	ۋ	ۋ	ۋ	ۋ
7	ج	ج	ج	ج	ج	22	س	س	س	س	س
8	ح	ح	ح	ح	ح	23	ش	ش	ش	ش	ش
9	ق	ق	ق	ق	ق	24	ى	ى	ى	ى	ى
10	ق	ق	ق	ق	ق	25	ر	ر	ر	ر	ر
11	ع	ع	ع	ع	ع	26	ز	ز	ز	ز	ز
12	ك	ك	ك	ك	ك	27	ه	ه	ه	ه	ه
13	م	م	م	م	م	28	ى	ى	ى	ى	ى
14	ك	ك	ك	ك	ك	29	ى	ى	ى	ى	ى
15	ل	ل	ل	ل	ل	30	ى	ى	ى	ى	ى

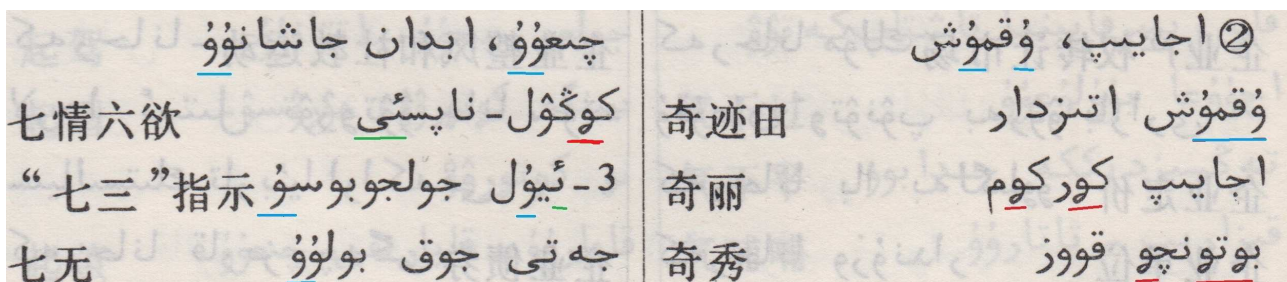
**Figure 1a:** KIRGHIZ OE with bar (not loop) (red). ARABIC LETTER U with filled in damma (blue). U+0626 (ALEF MAKSURA workaround is not used) (green). ARABIC LETTER FEH (orange). Kyrgyz language. [ky-1]



**Figure 1b:** KIRGHIZ OE with bar (not loop) (red). ARABIC LETTER U with filled in damma (blue). U+0626 (ALEF MAKSURA workaround is not used) (green). ARABIC LETTER FEH (orange). Kyrgyz language. [ky-1]



**Figure 1c:** ARABIC LETTER U with filled in damma (blue). U+0626 (ALEF MAKSURA workaround is not used) (green). Kyrgyz language. [ky-1]

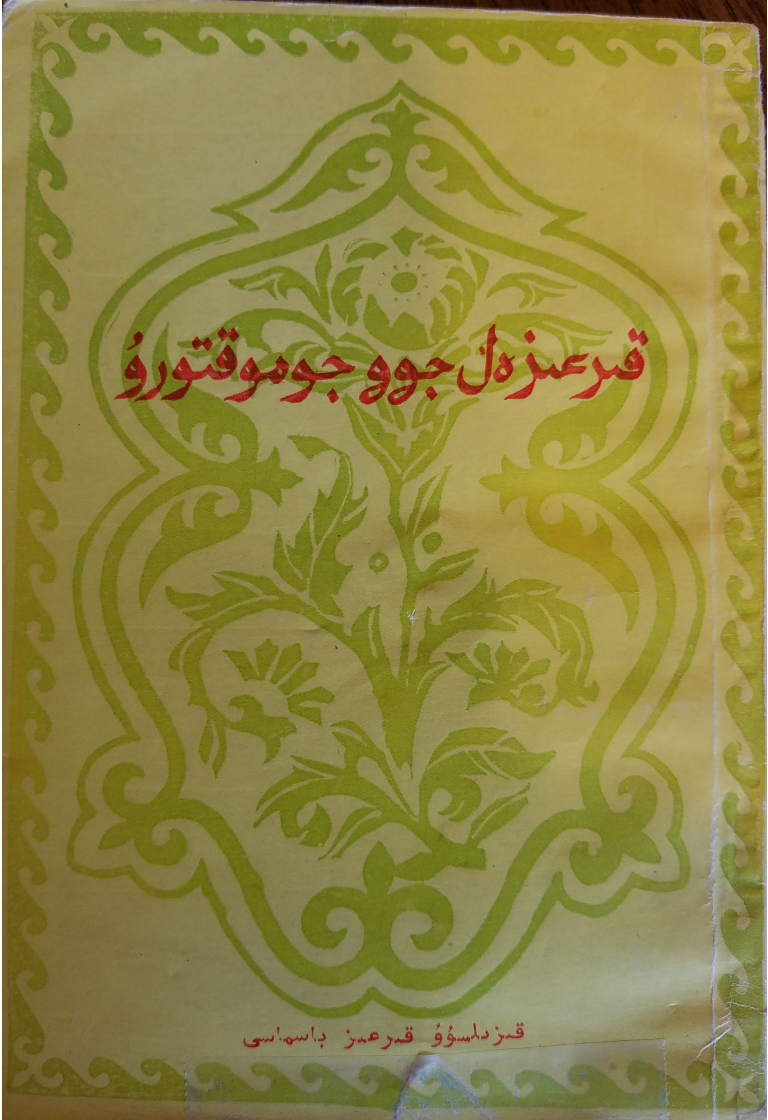


**Figure 2:** ARABIC LETTER U with filled in damma (blue). KIRGHIZ OE with loop instead of bar (red). U+0626 followed by ALEF MAKSURA using workaround mentioned above (green). Kyrgyz language. [ky-3, p. 213]



وبلاستىبىز 31 - مايدا كادىرلار قالايىق - قالقىنا امالئى ، جاقشى ئىش الدىرۇۇ كەرەك .  
 جىينالئىشنىن اچىپ ، حۇجىنتاۋنۇن قىلىپ بەرۋۇ كەرەك .  
 بۇروردۇن ۇلۇتتار قىزىمات جىينالئىشى ۋچۇنچۇ ، ورتو - باشتالغىچ دىسەننىن 50 جىلدىن ۋگۇتتورگو  
 تورتۇنچۇ ، اپتونوم رايون قۇرۇلغان -

**Figure 3:** ARABIC LETTER U with filled in damma (blue). KIRGHIZ OE with loop instead of bar (red). U+0626 followed by ALEF MAKSURA (workaround mentioned above) (green). Kyrgyz language. [ky-4, title page]



**Figure 4:** ARABIC LETTER U with filled in damma. KIRGHIZ OE with loop instead of bar. Kyrgyz language. [ky-2, cover].

ئۇيغۇر تىلىدىكى سۆز ياسغۇچى قوشۇمچىلار

**Figure 5:** ARABIC LETTER U with filled in damma. Uyghur language. [ug-1, p 1452].

唇 圓		唇 展		状 唇 合开的口 后前位部
音 圓 窄	音 元 寬	音 元 窄	音 元 寬	
ؤ	و	ا	ا	音 元 后
ؤ	و	ى	ى، ا	音 元 前

• (v) ؤ , (f) ف , (m) م , (w) و , (b) ب , (p) پ : 个四十二有共位音音辅 . 二  
• (j) ي , (r) ر , (l) ل , (n) ن , (z) ز , (s) س , (d) د , (t) ت  
• (h) ه , (q) ق , (k) ك , (g) گ , (x) ح , (c) ع , (j) ج , (ts) چ

**Figure 6:** ARABIC LETTER U, ARABIC LETTER U WITH HAMZA ABOVE with filled in damma. ARABIC LETTER HIGH HAMZA ALEF, ARABIC LETTER HIGH HAMZA WAW, ARABIC LETTER HIGH HAMZA YEH, ARABIC LETTER FEH. Kazakh language. [kk-1, p 1452].

• 态动使的 - ؤرپيت -  
• 式定不的 - ؤرپيؤ  
• 态互交的 - ؤرپيس -

**Figure 7:** ARABIC LETTER U and ARABIC LETTER U WITH HAMZA ABOVE with filled in damma. Kazakh language. [kk-1, p 1370].

## قازاق جازۇنىڭ الپبەسى

رەتى	اربېتەرى	اتالۋى	رەتى	اربېتەرى	اتالۋى
1	ا	ا	18	و	و
2	اٲ	اٲ	19	وٲ	وٲ
3	ب	بى	20	پ	پى
4	ۆ	ۆى	21	ر	رى
5	گ	گى	22	س	سى
6	ع	عى	23	ت	تى
7	د	د	24	ؤ	ؤ
8	ه	ه	25	ۇ	ۇ
9	ج	جى	26	ۇٲ	ۇٲ
10	ز	زى	27	فى	فى
11	ي	ي	28	ح	حى
12	ك	ك	29	ھ	ھى
13	ق	قى	30	چ	چى
14	ل	لى	31	ش	شى
15	م	مى	32	ى	ى
16	ن	نى	33	ئى	ئى
17	ڭ	ڭى	34	داپەكشە	

17

**Figure 8:** ARABIC LETTER U, ARABIC LETTER U WITH HAMZA ABOVE with filled in damma. ARABIC LETTER HIGH HAMZA ALEF, ARABIC LETTER HIGH HAMZA WAW, ARABIC LETTER HIGH HAMZA YEH, ARABIC LETTER FEH. Kazakh language. [kk-2, p 17].

نىڭ جوعارى دەڭگەيى مەن وزگەشەلىگن  
 تولق ايگىلەدى. جۇڭگو مەن رەسەيدىڭ  
 قول ۇستاسا وترىپ نىدەتكە قارسى اتتاز-  
 ەان، ۋاقتتىق قىندىقتى بىرگە ەڭسەرگەن  
 تابىستى امالياتنىڭ، ەسوز جوق، ەكى  
 ەلدىڭ اۋدىس - كۈيسى مەن سەلبەستىگىن  
 كۈشەيتۋدىڭ قۋانتى قوزعاۋشى كۈشەنە  
 اينالاتىندىعىنا سەنەمەن.  
 (جالعاسى 2 - بەتتە)

كورشى ەلدەر. ەمدە جاڭا تىپتى ايدارشا  
 ۋىرۋىس وكپە قابىنۋ نىدەتى ۋورشىپ  
 تۇرعان ساتتە، جۇڭگو مەن رەسەي ەكى  
 مەملەكەتنىڭ تىزە قوسا سەلبەسىپ، ەبىر -  
 بىرىنە قازايلاسىپ، جۇڭگو - رەسەي  
 ەلىم - تەحنىكالىق جاڭالىق اشۋ جىلىن  
 ۋاعىندا وتكىزۋى جۇڭگو - رەسەي جاڭا  
 داۋىردەگى جان - جاقتى ستراتەگىيالىق  
 سىتەستىك سەرىكتەستىك قارىم - قاتىناسە-

Figure 9: ARABIC LETTER U (blue). HIGH HAMZA (green). Kazakh language. [kk-3].

541

تەلەپپۇزى	ئىملاسى	تەلەپپۇزى	ئىملاسى
ùtùx	ئۇتۇش	u	ئۇ
ùtùxi	ئۇتۇشى	ùpa	ئۇپا
ùtidu	ئۇتىدۇ	ùpi - ەڭlik	ئۇپا - ئەڭلىك
		ni - ەڭligi	ئۇپا - ئەڭلىكى

Figure 10: ARABIC LETTER U. Uyghur language. [ug-2, p. 541]

مۇندەرىجە

439 .....	گ	1 .....	كېرىش سۆز
451 .....	ل	1 .....	لۇغەتنىڭ تۈزۈلۈشى ھەققىدە
459 .....	م	1 .....	ئىزاھات
498 .....	ن	11 .....	ھازىرقى زامان ئۇيغۇر ئەدەبىي تىلىنىڭ
510 .....	ھ	11 .....	تەلەپپۇز قائىدىسى (لايمھە)
526 .....	ئو	116 .....	ھازىرقى زامان ئۇيغۇر ئەدەبىي تىلىنىڭ
541 .....	ئۇ	183 .....	تەلەپپۇز لۇغىتى
552 .....	ئۆ	198 .....	ئا
563 .....	ئو	227 .....	ئە
570 .....	ۋ	241 .....	ب
577 .....	ئې	265 .....	پ
587 .....	ئى	274 .....	ت
606 .....	ي	282 .....	ج
		283 .....	چ
		324 .....	خ
		340 .....	د
		351 .....	ر
		353 .....	ز
		402 .....	ژ
			س
			ش
			غ
			ف
			ق
			ك

642 .....	قوشۇمچە I
	خەمىيىۋى ئېلېمېنتلار جەدۋىلى
644 .....	قوشۇمچە II
	ئاپتونوم رايونىمىزدىكى مۇھىم يەر
	ناملىرى
	قوشۇمچە III
	”ئال-، سال-، چال-، قال-، كەل-،
	بول-، قىل-“ پېئىللىرىنىڭ ئاخىرىدىكى
	”ل“ تاۋۇشىنىڭ تەلەپپۇزدا چۈشۈپ
	قېلىش- قالماسلىقىغا مۇناسىۋەتلىك
660 .....	شارائىتلار بويىچە تۈرلىنىشى

Figure 11: ARABIC LETTER U. ARABIC LETTER FEH. [ug-2, table of contents]

7.2 点阵字型表

本部分提供的 GB 21669—2008 中 190 个维吾尔文、哈萨克文、柯尔克孜文图形字符的 32 点阵正文白体字型见表 1。

表 1 维吾尔文、哈萨克文、柯尔克孜文字型表

060C	061B	061F	0626	0627	0628	062A	062C	062D	062E	062F	0631	0632	0633	0634	0639
،	؛	؟	ئى	ا	ب	ت	ج	ح	خ	د	ر	ز	س	ش	ع
063A	0640	0641	0642	0643	0644	0645	0646	0648	0649	064A	066A	0674	0675	0676	0677
غ	-	ف	ق	ك	ل	م	ن	و	ى	ي	%	”	۴	۵	۶
0679	067E	0686	0698	06AD	06AF	06BE	06C5	06C6	06C7	06C8	06C9	06CB	06D0	06D5	FB56
پ	پ	چ	ژ	ڭ	گ	ھ	و	ۆ	ۈ	ۈ	ۈ	ۈ	ې	ە	پ
FB57	FB58	FB59	FB7A	FB7B	FB7C	FB7D	FB8A	FB8B	FB92	FB93	FB94	FB95	FBAA	FBAB	FBAC
پ	پ	پ	چ	چ	چ	چ	ژ	ژ	گ	گ	گ	گ	ھ	ھ	ھ
FBAD	FBD3	FBD4	FBD5	FBD6	FBD7	FBD8	FBD9	FBDa	FBD8	FBDc	FBDd	FBDe	FBDf	FBE0	FBE1
ھ	ڭ	ڭ	ڭ	ڭ	ۈ	ۈ	ۈ	ۈ	ۈ	ۈ	ۈ	ۈ	ۈ	ۈ	ۈ
FBE2	FBE3	FBE4	FBE5	FBE6	FBE7	FBE8	FBE9	FBEA	FBEB	FBEC	FBED	FBEE	FBEF	FBF0	FBF1
ۈ	ۈ	ې	ې	پ	پ	د	د	ئا	ئا	ئە	ئە	ئو	ئو	ئو	ئو
FBF2	FBF3	FBF4	FBF5	FBF6	FBF7	FBF8	FBF9	FBFA	FBFB	FE89	FE8A	FE8B	FE8C	FE8D	FE8E
ئو	ئو	ئو	ئو	ئى	ئى	ئى	ئى	ئى	ئى	ئى	ئى	ئى	ئى	ئى	ئى
FE8F	FE90	FE91	FE92	FE95	FE96	FE97	FE98	FE9D	FE9E	FE9F	FEA0	FEA1	FEA2	FEA3	FEA4
ب	ب	ب	ب	ت	ت	ت	ت	ج	ج	ج	ج	ح	ح	ح	ح
FEA5	FEA6	FEA7	FEA8	FEA9	FEAA	FEAD	FEAE	FEAF	FEB0	FEB1	FEB2	FEB3	FEB4	FEB5	FEB6
خ	خ	خ	خ	د	د	ر	ر	ز	ز	س	س	س	س	ش	ش
FEB7	FEB8	FEC9	FECA	FECB	FECc	FECd	FECe	FECf	FED0	FED1	FED2	FED3	FED4	FED5	FED6
ش	ش	ع	ع	ع	ھ	غ	غ	غ	ف	ف	ف	ف	ق	ق	ق
FED7	FED8	FED9	FEDA	FEDB	FEDC	FEDD	FEDE	FEDF	FEE0	FEE1	FEE2	FEE3	FEE4	FEE5	FEE6
ق	ق	ك	ك	ك	ك	ل	ل	ل	ل	م	م	م	ن	ن	ن
FEE7	FEE8	FEE9	FEEA	FEED	FEED	FEED	FEED	FEF0	FEF1	FEF2	FEF3	FEF4	FEFB	FEFC	
ن	ن	ه	ه	و	و	ى	ى	ى	ى	ى	ى	ى	لا	لا	

Figure 12: Examples of all characters discussed in this document. [GB/T 25892.1-2010]

附录 A  
(资料性附录)  
柯尔克孜文  
柯尔克孜文

柯尔克孜文字母和对应编码字符集国家标准编码如表 A.1 所示。

表 A.1 柯尔克孜文

顺序	字母	GB 21669 编码	顺序	字母	GB 21669 编码
1	а *	0627	16	м	0645
2	б.	0628	17	о *	0648
3	в.	067E	18	о *	06C5
4	г.	0646	19	у *	06C7
5	д.	062A	20	у *	06C9
6	ж	062C	21	ў	06CB
7	ч	0686	22	с	0633
8	х	062D	23	ш	0634
9	ф.	0641	24	д	062F
10	қа	0642	25	р	0631
11	м	0639	26	з	0632
12	к	0643	27	е *	06D5
13	гы	06AF	28	ы *	0649
14	лш	06AD	29	ы *	0626
15	л	0644	30	и	064A

注：表中标记“\*”号的字母为元音字母。

6

Figure 13: Examples of U+0626, U+0641, U+06C5 (with loop), U+06C7. [GB/T 31917-2015]





表 1 正文白体

060C	061B	061F	0626	0627	0628	062A	062C	062D	062E	062F	0631
،	؛	؟	ئى	ا	ب	ت	ج	ح	خ	د	ر
0632	0633	0634	0639	063A	0640	0641	0642	0643	0644	0645	0646
ز	س	ش	ع	غ	-	ق	ف	ك	ل	م	ن
0648	0649	064A	066A	0674	0675	0676	0677	0678	067E	0686	0698
و	ى	ي	%	'	۳	ۋ	ۋ	ۋ	ۋ	پ	چ
06AD	06AF	06BE	06C5	06C6	06C7	06C8	06C9	06CB	06D0	06D5	FB56
پ	ه	ې	ۋ	ۋ	ۋ	ۋ	ۋ	ۋ	ۋ	ۋ	ك
FB57	FB58	FB59	FB7A	FB7B	FB7C	FB7D	FB8A	FB8B	FB92	FB93	FB94
گ	گ	گ	چ	چ	چ	چ	ر	ر	ر	ر	پ
FB95	FBAA	FBAB	FBAC	FBAD	FBDD	FBDA	FBDB	FBDC	FBDD	FBDE	FBDF
گ	ه	ه	ه	ه	ك	ك	ك	ك	ك	ك	ك
FBDA	FBDB	FBDC	FBDD	FBDE	FBDF	FBE0	FBE1	FBE2	FBE3	FBE4	FBE5
ۋ	ۋ	ۋ	ۋ	ۋ	ۋ	ۋ	ۋ	ۋ	ۋ	ۋ	ۋ
FBE6	FBE7	FBE8	FBE9	FBEA	FBEB	FBEC	FBED	FBEE	FBEF	FBF0	FBF1
ئو	ئو	ئو	ئو	ئا	ئا	ئا	ئا	ئو	ئو	ئو	ئو
FBF2	FBF3	FBF4	FBF5	FBF6	FBF7	FBF8	FBF9	FBFA	FBFB	FB89	FB8A
ئو	ئو	ئو	ئو	ئو	ئو	ئو	ئو	ئو	ئو	ئو	ئو
FB8B	FB8C	FB8D	FB8E	FB8F	FB90	FB91	FB92	FB93	FB94	FB95	FB96
ت	ت	ت	ت	ب	ب	ب	ب	ت	ت	ت	ت
FE9D	FE9E	FE9F	FEA0	FEA1	FEA2	FEA3	FEA4	FEA5	FEA6	FEA7	FEA8
ج	ج	ج	ج	ح	ح	ح	ح	خ	خ	خ	خ
FEA9	FEAA	FEAD	FEAE	FEAF	FEB0	FEB1	FEB2	FEB3	FEB4	FEB5	FEB6
د	د	ر	ر	ز	ز	س	س	س	س	س	س

3

表 1 (续)

FEB7	FEB8	FEC9	FECA	FECB	FECC	FECD	FECE	FECF	FED0	FED1	FED2
ش	ش	ع	ع	ع	ع	غ	غ	غ	غ	ف	ف
FED3	FED4	FED5	FED6	FED7	FED8	FED9	FEDA	FEDB	FEDC	FEDD	FEDE
ف	ف	ق	ق	ق	ق	ك	ك	ك	ك	ل	ل
FEDF	FEEO	FEF1	FEF2	FEF3	FEF4	FEF5	FEF6	FEF7	FEF8	FEF9	FEFA
ل	ل	م	م	م	م	ن	ن	ن	ن	ه	ه
FEED	FEEE	FEF0	FEF1	FEF2	FEF3	FEF4	FEF5	FEF6	FEF7		
و	و	ى	ى	ى	ى	ى	ى	لا	لا		

4

Figure 16: Example of U+0626, U+0641, U+0674..U+0678, U+06C5 (with loop), U+06C7. [GB/T 25900-2010]

## References

- [1] Eberhard, David M., Gary F. Simons, and Charles D. Fennig (eds.). 2020. *Ethnologue: Languages of the World*. Twenty-third edition. Dallas, Texas: SIL International. Online version: <http://www.ethnologue.com>. (accessed 14-Aug-2020).
- [2] Unicode CLDR Project. <http://cldr.unicode.org/index> (accessed 14-Aug-2020).
- [3] Unicode mailing list. 13-Oct-20. Subject: Question for Malay Jawi letter in Unicode - three quarter Hamza.

## Kazakh References

- [kk-1] 1985. "Qazaqsha-Khanzuwsha Sözdik". Kazakh/Chinese Official Government published dictionary. ISBN 7-105-00817-2/z.41.
- [kk-2] Kazakh Spelling Dictionary.
- [kk-3] 27-Aug-2020. Xinjiang Gazette. Newspaper number (inside China) CN 65-0003/-H

## Kyrgyz References

- [ky-1] Khīrghīz Jazuusunun Emila Sözdüğü قىرعىز جازۇۇسۇنۇن مەملا سوزدۇگۇ (Kyrgyz Literature Orthographic Dictionary).
- [ky-2] 1985. Kyrgyz Folktales.



- [ky-3] Ханзуча-кыргызча жаңы сөз жаңы аталамалар сөздүгү Шинжаң эл басмасы, 1998. ISBN: 7-228-03984-X/Н. Chinese-Kyrgyz New Word, New Names Dictionary.
- [ky-4] [KG] 2005-06-04  
Kyzylsuu Geziti June 4, 2005  
Kyzylsuu Newspaper  
Article 1  
Областыбыздын жер жерлеринде - «1 - июнь» балдар майрамы куттукталды  
"June 1" Children's Day celebrated in the territory of our region

## Uyghur References

- [ug-1] 1999. Uyghur Tilining Izahliq Lughiti (Qisqartilmisi). Uyghur Official Government published dictionary. Urumchi, Shinjang Xelq Neshiryati, ISBN : 7-228-05448-2/H.146
- [ug-2] 1988. Hazirqi Zaman Uyghur Edebiy Tilining Teleppuz Lughiti. Uyghur Pronunciation Dictionary. ISBN 7-105-00529-7/H-28

## China National Standard Documents (with thanks to Liang Hai for these references)

- [GB/T 25892.1-2010] Information technology - Uyghur, Kazak, Kirgiz coded character set - 32 Dot matrix font - Part 1: Tuz lean. <http://www.gb688.cn/bzgk/gb/newGbInfo?hcno=C7FF86E923218431F96A9EE1DAE63BD9>
- [GB/T 31917-2015] Information technology—Universal keyboard layout of the alphanumeric zone for Kyrgyz <http://www.gb688.cn/bzgk/gb/newGbInfo?hcno=0B09C0F5F7A1EEB65D4DB22DAAD2A125>
- [GB/T 21669-2008] Information technology - Uyghur, Kazak and Kirghiz coded character set <http://www.gb688.cn/bzgk/gb/newGbInfo?hcno=0E186E72C798CF977CEFEC297A620841>

[GB/T 25900-2010] Information technology - Uighur, Kazak, Kirgiz font for information process - Bai Ti & Hei Ti  
<http://www.gb688.cn/bzgk/gb/newGbInfo?heno=AC57297F8FDC21A9340E47DE15B31F2E>