

The 30th Internationalization and Unicode Conference  
November 15-17, 2006, Washington, DC USA

## Web Development Considerations for Unicode-based Text Processing in Uyghur<sup>1</sup> Language

**Prof. Imad Saleh**

*Paragraphe* Laboratory  
University of Paris VIII  
isaleh@wanadoo.fr

**Waris Abdukerim Janbaz**

*Paragraphe* Laboratory  
Researcher & PhD student  
University of Paris VIII  
warisabdukерim@yahoo.com

**Abstract:** This paper focuses on text processing and accessibility of an agglutinative Turkic language — Uyghur. It describes the basic concept of Uyghur Unicode font developing, character displaying and character inputting methods in a non-Uyghur supporting environment. This article also highlights the problems caused by the absence of two Uyghur characters in the Unicode Standard, and more importantly, the article describes the proposed solutions to overcome such incompleteness.

**Keywords:** Arabic-Script Uyghur, Turkic language, font, Input Method.

### Introduction

The Uyghurs are a Turkic-speaking ethnic group, with a population of about nine million according to official figures and inhabiting an area of Central Asia including today's Xinjiang Uyghur Autonomous Region (hereafter: XUAR, also called Chinese Turkistan) as well as parts of Kazakhstan and urban regions in the Ferghana valley. Historical records indicate that Uyghurs have used various scripts, sometimes concurrently, during different periods over the past 2000 years. The most important ones are the Soghdian Script (5<sup>th</sup> -10<sup>th</sup> centuries<sup>2</sup> and 15<sup>th</sup>-16<sup>th</sup> centuries<sup>3</sup>), the Orkhon-Yenisei script (also known as Turkic Runic script, 6<sup>th</sup>-9<sup>th</sup> centuries<sup>4</sup>), the Uyghur Script (also known as Old Uyghur Script, 10<sup>th</sup>-18<sup>th</sup> centuries<sup>5</sup>) and, after their gradual adoption of Islam starting in A.D. 934, the Arabic Script (10<sup>th</sup> century - present). The official writing system of the XUAR Uyghurs is Arabic-Script Uyghur<sup>6</sup> (hereafter: ASU) whereas the Cyrillic-Script Uyghur (hereafter: CSU) is still used by the Uyghurs of the ex-Soviet Union Republics. The newly introduced transliteration<sup>7</sup> – Latin-Script Uyghur<sup>8</sup> (hereafter: LSU) has

---

<sup>1</sup> Also spelled Uighur, Uygur, Uighuir, Uigur, Uiguir, and Weiwuer. The Language and Script Committee of the Xinjiang Uyghur Autonomous Region is scheduled to hold a normalization conference in mid-October 2006 to select one spelling.

<sup>2</sup> Haji Nurhaji, *Qedimki Uyghurlar we Qaraxaniylar (Ancient Uyghurs and Qarakhanids)*, Xinjiang People's Press, 2001, p.122.

<sup>3</sup> Abdurêhim Rahman & Ömerjan, *Shinjang Tarix Matêriyalliri (Xinjiang Historical Documents)*, volume 40, Xinjiang People's Press, 1980.9.

<sup>4</sup> The Orkhon-Yenisei alphabet was used by the Orkhon Uyghur Empire (744 - 840) but also by other Turkic tribes of Eastern Central Asia. See <http://www.omniglot.com/writing/orkhon.htm>

<sup>5</sup> Adopted, in an adapted form, first by the Mongols, and then by the Manchus.

<sup>6</sup> See annex 2

<sup>7</sup> Using one writing system to represent words in another is called transliteration.

<sup>8</sup> called *Uyghur Kompyutêr Yêziqi (UKY) or Uyghur Latin Yêziqi (ULY)* in Uyghur, meaning “Uyghur Computer Writing” or “Latin-Script Uyghur”. See [www.ukij.org/teshwiq/UKY\\_Heqqide\(KonaYeziq\).htm](http://www.ukij.org/teshwiq/UKY_Heqqide(KonaYeziq).htm) .

become widely accepted among Uyghurs and Uyghurologists. Now it is a commonly used standard for the transliteration for both ASU and CSU. Modified to take into account the phonetic and grammatical characteristics of Uyghur language, the Arabic script has been reintroduced for use by XUAR Uyghurs 1983 after a twenty-year interruption<sup>9</sup>. In this paper, we draw your attention to a few essential issues related to Unicode-based text processing in Windows using the Arabic-Script Uyghur.

## Background

In the early 90's, Xinjiang local software developers worked out means to input ASU letters into computers under DOS. Starting in the late 90's, Microsoft Windows has become very popular in Uyghur society and many Windows add-on programs have entered the market to meet the needs of local people. Since the existing platforms do not supply any Uyghur input method nor any fonts that include all the glyphs of the ASU alphabet, inputting Uyghur text into different application software and correctly displaying Uyghur characters presented huge difficulties. Local software developers did not recognize the importance of having a common standard for different software. In spite of the fairly passive attitude of state-sponsored research groups and government authorities to the development of Uyghur information technology, many individuals started creating Uyghur software and websites using LSU or ASU. Webmasters created their websites mainly using ASU in image format. Whenever text format was used, each single website had to provide its own specific fonts<sup>10</sup>: every web site owner thus created and named his/her own fonts, and users had to download a different font for different website. Font makers and application software developers did not follow the Unicode standard because of some difficulties even if the most essential glyphs were already included in the Unicode table (this point will be discussed in more details in the following sections). No webmaster accepted the font names and coding rules of the others, and there was no common standard. Most of the fonts created during this period either replaced the ASCII characters or replaced the Arabic Unicode characters (0x600-0x6FF) with Uyghur characters without replacement agreement. Since the number of the Arabic letters in the code range 0x600-0x6FF is larger than the number of ASU letters, people made different choices in the replacement of some Arabic characters by specific ASU characters.

A large number of issues regarding non-standard fonts and their use were addressed in many different ways by individual computer scientists. Meanwhile, many of these problems were circumvented by using methods unrelated to the Unicode standard. Eventually, people expressed their strong desire to use the Unicode standard for Uyghur language processing in order to bring some order to a fairly confused situation.

In June 2002, the author developed the first Uyghur Unicode font and implemented both system-level and browser-level Input Method Editors(IME) for Windows. It was considered a revolutionary step forward, owing mostly to the new method and applications that are fully Unicode-compliant (as opposed to occasionally compatible).

---

<sup>9</sup> See "An Introduction to Latin-Script Uyghur", by Jean R. Duval and Waris A. Janbaz, September 2006.

<sup>10</sup> From now on in the text, the term "Uyghur font" refers to the Arabic-Script Uyghur font.

Hence, a campaign was launched to popularize and adapt the Unicode standard for Uyghur fonts. According to a survey carried out by the author, there were some 400 ASU processing software in use with no less than 300 different font encodings<sup>11</sup>. In 2004, under the leadership of the Uyghur Computer Science Association(UCSA)<sup>12</sup> founders, four conferences<sup>13</sup> were held on the Internet to introduce a unified encoding for Uyghur fonts. Among the attendees there were some leading Uyghur software experts from private companies, government-sponsored software research organizations and independent software engineers, namely, Alim Ehed, Erkin Batur, Gheyret Toxti Kenji, Muhemmed Abdulla, Yasin Imin and the author. The conferences reached an agreement to identify some ambiguous Uyghur glyphs in the Unicode and proposed solutions do handle such problems. In the following sections we describe the agreement and highlight the most essential steps to follow during Uyghur Unicode-based text processing and software development.

### Uyghur letters: characters and identification issues

ASU is a complex script<sup>14</sup> made up of 8 vowels<sup>15</sup> and 24 consonants (see annex 2), either borrowed or adapted from the original Arabic alphabet. Like Arabic, it is written from right to left and each letter has two to eight different shapes depending on its position in a word : ASU letters have initial, medial, final and isolated forms; some letters have conjunct forms<sup>16</sup>. In total, the ASU alphabet has 126 different glyphs. Besides, like in Arabic, the combination of “f” after “J” requires two more glyphs “ڤ” and “ڦ” depending on its position

The 108 basic glyphs<sup>17</sup> of the ASU letters were initially accepted by the Unicode Consortium/ISO. Another 18 glyphs<sup>18</sup> out of the 20 glyphs for conjunct vowel forms (in light grey in table 1 below) were added in 1998. Since the 20 conjunct glyphs can also be expressed as a sequence of two existing Unicode glyphs, the above-mentioned 2004 conference attendees suggested not to map the 18 glyphs marked in grey in table 1 to the code range: FBEA — FBEB. Using these 18 glyphs may cause problems such as reducing text inputting speed, increasing data storage redundancy, complicating data sorting operations.

---

<sup>11</sup> See “A proposal for the unification of Uyghur fonts”, <http://www.ukij.org/fonts/history/Teklip.pdf> .

<sup>12</sup> UCSA – The Uyghur Computer Science Association (or UKIJ – Uyghur Kompyutér Ilimi Jem’iyiti in Uyghur) is a non-profit association, founded by the author in Jan 2004. Web site: <http://www.ukij.org> .

<sup>13</sup> See “History of the unification of Uyghur fonts” <http://www.ukij.org/fonts>.

<sup>14</sup> “It refers to any writing system that requires some degree of character reordering and/or glyph processing to display, print or edit.” See [www.microsoft.com/typography/Glyph%20Processing/intro.msp](http://www.microsoft.com/typography/Glyph%20Processing/intro.msp) for details.

<sup>15</sup> The Arabic alphabet has 3 letters used to indicate long vowels or, in the case of the first one, used as a vowel support ا, و, ي. Short are not noted in normal writing. Given its phonetic characteristics, Uyghur notes all vowels: ئى, ئې, ئۆ, ئۇ, ئو, ئه, ئا, ئا, using derivatives of traditional Arabic letters.

<sup>16</sup> The initial form and, under some circumstances, the median form of all vowels are preceded by one “glottal stop sign ھ or ھ” (supported hamza) with which they form a common letter (treated by Uyghur as a single letter, see annex 2).

<sup>17</sup> Glyphs not including supported-hamza glyphs, See annex 2 and footnote 20.

<sup>18</sup> See Arabic Presentation Forms-A, glyph code range: FBEA – FBFB. See also table 1.

The 30th Internationalization and Unicode Conference  
November 15-17, 2006, Washington, DC USA

isolated'	final'	median'	initial'	isolated	final	medial	initial
				ا	ا	ئا	ئا
				ه	ه	هه	هه
						ئو	ئو
						ئۆ	ئۆ
						ئۇ	ئۇ
						ئۈ	ئۈ
پ	پ	پ	پ	ئپ	پپ	پپ	پپ
ى	ى	ه	د	ئى	ئى	ئى	ئى

Table 1. Composite forms of Uyghur (ASU) vowels

Unfortunately, the conjunct medial forms  $\text{پپ}^{19}$  and  $\text{ئى}^{20}$  of the Uyghur letters  $\text{پ}$  and  $\text{ئى}$  (indicated in dark gray in table 1 below) are still absent<sup>21</sup> in the Unicode Standard table<sup>22</sup> – Arabic Presentation forms-A. This omission renders the present Unicode standard incomplete and forces people to make up for it by “borrowing” two empty positions, FBD1 and FBD2. In order to avoid future confusion, it has been recommended by the UCSA, which is planning to submit a written request to the Consortium to that end, to supplement these two glyphs in future versions of the Unicode.

### Other Problem Areas

Microsoft Office 2000 comes with an international font called “Arial Unicode MS”. It includes nearly all the shapes (except the above-mentioned missing two) of the Uyghur letters, but it does not contain glyph substitution and positioning features of Uyghur characters. The same can be said of other fonts such as “Times New Roman” and “Traditional Arabic”, in which some substitution sequences lead to incorrect display. For example:

1. ئالەمدىكى ھەممە ئىنسان قەيىھ ئەمەس
2. ئالەمدىكى ھەممە ئىنسان قەيىھ ئەمەس
3. ئالەمدىكى ھەممە ئىنسان قەيىھ ئەمەس

(Not all human beings in the world are evil)

The first and second sentences above are considered illegal character combinations if they use existing “Arial Unicode MS” and “Times New Roman” fonts because their linked shapes for  $\text{ى}$ ,  $\text{ھ}$ ,  $\text{ئە}$  are not correct according to the ASU alphabet (see annex 2) shaping rules. Only sentence 3 above is correctly displayed since it uses a font,

<sup>19</sup> Character name for the Unicode Standard: ARABIC LIGATURE YEH WITH HAMZA ABOVE WITH E MEDIAN FORM. Ex: باغئېرىق (*Baghériq*).

<sup>20</sup> Character name for the Unicode Standard: ARABIC LIGATURE UIGHUR KIRGHIZ YEH WITH HAMZA ABOVE WITH ALEF MAKSURA MEDIAN FORM. Ex: قەتئىي (certainly, doubtlessly)

<sup>21</sup> The XUAR’s delegation members, Prof. Hoshur Islam and Yasin Imin, who have submitted the proposition also admit this omission. See also Arabic Presentation Forms-A (code range: FBEA – FBFB).

<sup>22</sup> <http://www.unicode.org/charts/PDF/UFB50.pdf>

specifically developed for ASU, “UKIJ Tuz Tom”. In order to create the correct connected forms for Uyghur, it is necessary to take special glyph processing measures for the three problem letters **ى**, **ه**, **ئ** and two “glottal stop signs **ئ**, **ء**” (supported hamza), during the creation of Uyghur fonts. These problems can be overcome by adding all the information that control the substitution and positioning of Uyghur glyphs during font development. Notably by using Open Type Layout features of Open Type font technology.

As we can see from the Arabic character code tables and list of character names<sup>23</sup> for the Unicode standard, the Uyghur letters do not appear in the ASU alphabetical order. The positions of Uyghur letters **ى**, **ه**, **ئ** are ambiguous, and the two “glottal stop signs **ئ**, **ء**” represent the initial and medial form of the Arabic letter **ئ** which does not exist in the ASU alphabet.

Representative shapes	ى				ه	ه				ئ	
	06CC <sup>24</sup>		0649 <sup>25</sup>		06D5 <sup>26</sup>	0647 <sup>27</sup>		06BE <sup>28</sup>		0626 <sup>29</sup>	
isolated	FBFC	ى	FEEF	ى		FEE9	ه	FBAA	ه	FE89	ئ
final	FBFD	ى	FEF0	ى		FEEA	ه	FBAB	ه	FE8A	ئ
initial	FBFE	ئ	FBE8	د		FEEB	ه	FBAC	ه	FE8B	ئ
medial	FBFF	ئ	FBE9	ه		FEEC	ه	FBAD	ه	FE8C	ئ

Table 2. Problem-letters in the Unicode chart.

Obviously, two separate code points are given to each **ى** and **ه** in Unicode. The Arabic character **ه** *hah* has four different shapes, which correspond to two shapes in each of two different ASU letters, **ئ** and **ه**. ASU letter **ئ** has a unique code 06D5 but the positions of its four different shapes are not indicated in the Arabic Presentation forms. In the ASU alphabet, the three letters **ى**, **ه**, **ئ** have the following different shapes (see also annex 2).

isolated'	final'	medial'	initial'	isolated	final	medial	initial
				ه	ه	ه	ه
ى	ى	ه	د	ئ	ئ	ئ	ئ
				ه	ه	ه	ه

Table 3. Different shapes of **ى**, **ه**, **ئ** in Uyghur alphabet

<sup>23</sup> See <http://www.unicode.org/charts> .

<sup>24</sup> ARABIC LETTER FARSI YEH, Initial and medial forms of this letter have dots.

<sup>25</sup> ARABIC LETTER ALEF MAKSURA, represents YEH-shaped letter without dots in any positional form

<sup>26</sup> ARABIC LETTER AE (Uighur Kazakh, Kirghiz)

<sup>27</sup> ARABIC LETTER HEH

<sup>28</sup> ARABIC LETTER HEH DOACHASHMEE (Urdu)

<sup>29</sup> ARABIC LETTER YEH WITH HAMZA ABOVE

The 30th Internationalization and Unicode Conference  
November 15-17, 2006, Washington, DC USA

Given that Open Type font technologies allow font developers to integrate shaping information inside fonts, the following decisions are made by above mentioned conference attendees and implemented all over XUAR since late 2004.

ى: Uyghur letter *i* as in *ishik* (ئىشىك, door). It has 8 different shapes as listed in the table 3 above. For the initial', medial', final' and isolated' forms of this letter we use the initial', medial, final and isolated forms of the Arabic letter ى 0649, respectively; The isolated' and final' shapes should not be confused with the isolated and final shapes of the Farsi letter YEY 06CC. The composite forms should be expressed as a combination of the initial and medial shapes of Arabic letter YEY ى 0626 followed by the medial and final forms of the Arabic letter ى 0649. For example: Open Type font shaping engine should combine ى (FEF0) with ى (FE8B) to obtain ى.

ئ: Uyghur letter *e* as in *eyneklerde* (ئەينەكلەردە, in the mirrors). This letter uses the final and isolated shapes (ە , ھ) of the Arabic letter ھ<sup>30</sup> 0647, in the same way as Persian does. This causes a special problem due to the fact that the glyphs of the same Arabic ھ in the initial and medial positions (ھ , ھ) correspond to those of Uyghur ھ (*h* as in *hélihem*, even now; *gunah*, sin or offense; *qebih*, odious), which, in turn, has different final and isolated shapes (ھ , ھ, see table 3). In order to deal with this inconsistency, we have chosen to use 06D5 for the Uyghur letter ئ and 06BE for the Uyghur letter ھ. Moreover, the Urdu letter ھ 06BE also has nearly the same shapes as Uyghur. Inside Uyghur fonts, substitution lookups of 06D5 use FEE9 and FEEA in isolated and final positions of ئ. The composite shapes ھ and ئ can be obtained by adding the final form of ھ 06D5 into the initial and medial forms of ى 0626, in the same way as is done for ى.

ئ and ء<sup>31</sup>: the glottal stop signs. This is a phoneme which is not listed separately in the ASU alphabet but still covered by its spelling rules. In Uyghur words, the glottal stop is not as strongly pronounced as it is in Semitic languages or in Uzbek for example, and it has weakened to become no more than a hiatus. Marked in ASU by a hamza on top of a “tooth”, it appears usually in words of Arabic origin and replaces an original ‘ain (ع) or a hamza (ء) in a medial or final position (e.g. ئالەم from Arabic , سائەت from Arabic ساعة, خائىن from Arabic , سوئال from Arabic سؤال). In initial position, the same sign is considered as part of the initial form of a vowel and does not have any phonetic value<sup>32</sup>. In the Unicode chart, the glottal stop signs correspond to the initial and medial forms of the Arabic letter ى 0626. These Arabic glyphs are not considered as different shapes of

<sup>30</sup> See also <http://www.unicode.org/standard/where/> , Variant shapes of the Arabic character *hah*.

<sup>31</sup> Initial and medial forms of ى 0626.

<sup>32</sup> It is often said that the decision of Uyghur linguists to add this sign as part of the initial form of letters is a link with the old Uyghur writing system, in which all initial vowels were preceded by a tooth. The Arabic alphabet has 3 letters, ا, و and ي which can be used to indicate long vowels. Short vowels can be indicated through the use of vowel marks above or under the consonants but which are dispensed of in normal writing. Uyghur notes down all vowels as in table 1 using derivatives of traditional Arabic letters.

any independent letter in the Uyghur alphabet (cf. annex 2). Since one glyph of each of the two letters ئې and ئى (shown in dark-gray in the table 1) is still missing in Unicode, we can use a sequence of either of these glyphs ( ﺉ or ﺊ ) followed by the final, isolated, medial' or final' forms of vowels ئې and ئى. Consequently, all the other conjunct forms of the vowels (cf. table 1) can be obtained combining with the Arabic letter ﺀ 0626 and a representative shape of a vowel respectively.

In spite of the above mentioned limitations (two glyphs instead of one conjunct glyph for ئې and ئى) the above mentioned conventions have now been widely accepted by the Uyghur Computer Science Association and, subsequently, by the Xinjiang University branch of the 863 Research Group<sup>33</sup>. Annex 1 gives more details about the Uyghur letters' representative shapes, glyph codes for the shaping during substitution and positioning.

## Uyghur font and input method development

After having learnt the specificities of those letters, it is easy to create Uyghur fonts using existing font-creating software. The inclusion of non-spacing combining marks, such as ZWJ (zero width joiner 200C), ZWNJ (zero width non-joiner; 200D), LTR (left to right mark; 200E), and RTL (right to left mark; 200F), is also recommended in any Uyghur font. The rest of the time-consuming repetitive font developing task is absolutely the same as when creating an Arabic script font<sup>34</sup>. Some Uyghur Unicode fonts that developed based on the above-mentioned conventions are available for free at the UCSA website<sup>35</sup>. Our recommended font-creating tools are: *Font Creator*<sup>36</sup> and *Fontographer*<sup>37</sup>. Glyph substitutions, positioning lookups and shaping features and Open Type tables of Arabic fonts can also be added with the help of software like Microsoft VOLT.

Let us now assume that we already have Uyghur Unicode fonts that are developed based on above-mentioned conventions. Since the existing operating systems<sup>38</sup> do not provide an IME for Uyghur, we still need to create ways to input Uyghur letters. Windows 2000 and later versions have a Uniscribe Script Processor, USP10.DLL, that can handle complex script shaping both on the application and system levels. Besides, Windows 2000 and later versions' multilingual support for right-to-left written languages present an advantage for Unicode-based Text Processing in Uyghur. A combination of these two can be very useful for the development of an Uyghur IME. Before 2004, the majority of Uyghur software developers would use non-Unicode based fonts and application-level shaping techniques to render the correct forms of Uyghur letter. Later on, software developers learnt that 1.471 or later versions of USP10.DLL has a shaping engine that

---

<sup>33</sup> A National High-Tech Research Group, financed by the PRC government. The XJU branch is specialized in multilingual software development.

<sup>34</sup> See <http://www.microsoft.com/typography/OpenType%20Dev/arabic/intro.mspix> for more information about developing Open Type Fonts for Arabic Script

<sup>35</sup> See <http://www.ukij.org/fonts>

<sup>36</sup> <http://www.high-logic.com/fontcreator.html>

<sup>37</sup> <http://www.fontlab.com/Font-tools/Fontographer>

<sup>38</sup> Except Windows Vista, which is supposed to come out next year.

The 30th Internationalization and Unicode Conference  
November 15-17, 2006, Washington, DC USA

fully supports Uyghur Unicode. And therefore, they started releasing both system-level and application-level input methods<sup>39</sup>. Most of these IMEs either use a plug-in application to capture and modify keyboard events or a DLL file generated by Microsoft Keyboard Layout Editor to input Uyghur characters. Locale definitions for Uyghur is still absent in Windows XP and older versions, but it should not be an obstacle to adding an Uyghur IME. Given that the ISO two-letter code<sup>40</sup> is “ug” and it is not taken by others locales, Uyghur IMEs prefer using “ug” to represent the Uyghur language in Windows Language bar. More details on how to develop an Uyghur IME and multiscrypt converting tools can be found in the referenced articles.

## Conclusion

In this paper, we focused mainly on the background of Uyghur software development and implementation issues related to Uyghur Unicode fonts and IMEs. The more than two years of efforts invested in unifying Uyghur fonts by using the Unicode standard proved fruitful. Many application software and websites were purposefully created by UCSA members to develop Uyghur information technology. We expect more efficient support from the government authorities to enforce and implement the XUAR Language and Script Law to satisfy the rising demand for Uyghur software and Uyghur websites. At the same time, more efforts should be done by the computer software industry to expand compatibility. There are also other technical issues to resolve in order to add Uyghur locale definitions to Windows XP and earlier versions of Windows. Finally, we call on international software companies not to omit the Uyghur from their lists of supported languages in the future.

## References

- [1] Waris A. Janbaz, 2002, *Online Uyghur Unicode processing technique and its implementation* (publication in Chinese), Xinjiang University Press.
- [2] Abdurêhim, Waris A. Janbaz, 2004, *Orthographic rules of the Latin-Script Uyghur (in Uyghur)*, [http://www.ukij.org/teshwiq/UKY\\_Heqqide\(KonaYeziq\).htm](http://www.ukij.org/teshwiq/UKY_Heqqide(KonaYeziq).htm).
- [3] The Unicode Consortium (2003). *The Unicode Standard, Version 4.0*, Addison-Wesley Professional, ISBN: 0321185781.
- [4] Xinjiang University, *Proceedings 2000 International Conference on Multilingual Information Processing*. Ürümchi ( publication in Chinese).
- [5] The Unicode Consortium Website <http://www.unicode.org>
- [6] Reinhard F. Hahn (1991), *Spoken Uyghur*, University of Washington, ISBN: 0-295-97015-4.
- [7] John Hudson, November 2000, *Windows Glyph Processing: an Open Type Primer*, <http://www.microsoft.com/typography/Glyph%20Processing/intro.msp> .
- [8] F. Avery Bishop, David C. Brown, David M. Meltzer, *Supporting Multilanguage text layout and complex scripts with Windows 2000*, [www.microsoft.com/typography/developers/uniscribe/intro.htm](http://www.microsoft.com/typography/developers/uniscribe/intro.htm)
- [9] Imad Saleh, Waris Abdukerim Janbaz, June 2006, *Uyghur language processing on the Web*, AIML 2006 International Conference, Sharm El Sheikh, Egypt.
- [10] The Uyghur Computer Science Association, December 2004, “*History of the unification of Uyghur fonts*”, <http://www.ukij.org/fonts> .
- [11] Jean Rahman Duval, Waris Abdukerim Janbaz, September 2006, *An Introduction to Latin-Script Uyghur*, 2006 Middle East & Central Asia Conference, University of Utah, Salt Lake City, USA.

---

<sup>39</sup> Free, open source versions of these IMEs are also available on the UCSA’s website <http://www.ukij.org> .

<sup>40</sup> See [http://www.loc.gov/standards/iso639-2/php/code\\_list.php](http://www.loc.gov/standards/iso639-2/php/code_list.php) .

The 30th Internationalization and Unicode Conference  
November 15-17, 2006, Washington, DC USA

**Annex 1. Representative shapes and codes by Uyghur alphabet order**

final	medial	initial	isolated	Rep.	Order
ئ FE8A	ئ FE8C	ئ FE8B	ئ FE89	ئ 0627	1
ا FE8E			ا FE8D	ا 0627	2
ە FE8A			ە FEE9	ە 06D5	3
ب FE90	ب FE92	ب FE91	ب FE8F	ب 0628	4
پ FB57	پ FB59	پ FB58	پ FB56	پ 067E	5
ت FE96	ت FE98	ت FE97	ت FE95	ت 062A	6
ج FE9E	ج FEA0	ج FE9F	ج FE9D	ج 062C	7
چ FB7B	چ FB7D	چ FB7C	چ FB7A	چ 0686	8
خ FEA6	خ FEA8	خ FEA7	خ FEA5	خ 062E	9
د FEAA			د FEA9	د 062F	10
ر FEAE			ر FEAD	ر 0631	11
ز FEB0			ز FEAF	ز 0632	12
ژ FB8B			ژ FB8A	ژ 0698	13
س FEB2	س FEB4	س FEB3	س FEB1	س 0633	14
ش FEB6	ش FEB8	ش FEB7	ش FEB5	ش 0634	15
غ FECE	غ FED0	غ FECF	غ FECD	غ 063A	16
ف FED2	ف FED4	ف FED3	ف FED1	ف 0641	17
ق FED6	ق FED8	ق FED7	ق FED5	ق 0642	18
ك FEDA	ك FEDC	ك FEDB	ك FED9	ك 0643	19
گ FB93	گ FB95	گ FB94	گ FB92	گ 06AF	20
ك FBD4	ك FBD6	ك FBD5	ك FBD3	ك 06AD	21
ل FEDE	ل FEE0	ل FEDF	ل FEDD	ل 0644	22
م FEE2	م FEE4	م FEE3	م FEE1	م 0645	23
ن FEE6	ن FEE8	ن FEE7	ن FEE5	ن 0646	24
ه FBAB	ه FBAD	ه FBAC	ه FBAA	ه 06BE	25
و FEEE			و FEED	و 0648	26
ۇ FBD8			ۇ FBD7	ۇ 06C7	27
ۆ FBDA			ۆ FBD9	ۆ 06C6	28
ۈ FBDC			ۈ FBDB	ۈ 06C8	29
ۋ FBDF			ۋ FBDE	ۋ 06CB	30
پ FBE5	پ FBE7	پ FBE6	پ FBE4	پ 06D0	31
ى FEF0	ى FBE9	ى FBE8	ى FEEF	ى 0649	32
ي FEF2	ي FEF4	ي FEF3	ي FEF1	ي 064A	33

Annex 2: Arabic-Script Uyghur Alphabet with pictures (used to teach Uyghur children)

