

## A Study of Traditional Mongolian Script Encodings and Rendering: Use of Unicode in OpenType fonts

Biligsai Khan Batjargal<sup>1</sup>, Garmaabazar Khaltarkhuu<sup>2</sup>, Fuminori Kimura<sup>3</sup>, and  
Akira Maeda<sup>3</sup>

<sup>1</sup>Graduate School of Science and Engineering, Ritsumeikan University, 1-1-1 Noji-higashi,  
Kusatsu, Shiga, 525-8577, Japan

<sup>2</sup>Mongolia-Japan Center for Human Resources Development, Mongolia-Japan Center  
Building, P.O.Box-46A/190, Ulaanbaatar, Mongolia

<sup>3</sup>College of Information Science and Engineering, Ritsumeikan University, 1-1-1 Noji-  
higashi, Kusatsu, Shiga, 525-8577, Japan  
{biligsai Khan, garmaabazar}@gmail.com, {fkimura, amaeda}@is.ritsumei.ac.jp

---

### Abstract

*This article discusses the rendering issues of complex text layouts, particularly traditional Mongolian script. Some standards such as Unicode and OpenType format have been implemented and are supported widely. Traditional Mongolian script has been standardized in Unicode. We analyzed existing OpenType fonts and their rendering schemes for traditional Mongolian script. We found some errors, and discovered grammatical rules, which are not documented in international standard and guidelines. None of the existing OpenType fonts was complete. This article provides some improvements and recommendations for future development of traditional Mongolian OpenType fonts.*

*Furthermore, this article discusses the issues of traditional Mongolian script encoding. There are several non-Unicode encodings and code pages that have been developed, some of which are still commonly used. This diversity of character encodings, code pages and keyboard drivers make it difficult to incorporate enriched Unicode content. We analyzed existing non-Unicode encodings and code pages of traditional Mongolian script. Developing text conversion technique from diverse encodings to Unicode is becoming an essential demand in order to adopt already available digital content in traditional Mongolian script easily.*

*Authors encourage developers to use the standardized code points of Unicode and the Unicode based encodings in traditional Mongolian script.*

### Keywords

*Traditional Mongolian Script, Unicode, Character Encoding, OpenType, Complex Text Rendering*

---

## 1 Introduction

In recent years, the role and importance of computer systems that preserve cultural heritage increased worldwide. Internationalization and digital cultural heritage preservation in Mongolia require information systems to support various ancient Mongolian records. This article discusses the challenges for displaying and handling text in traditional Mongolian script, such as rendering and character encodings. Most of the writing systems in Asia require complex transformations between text input and text display for proper rendering. For these complex text scripts, the way text is stored is not mapped to the way it is displayed in a straightforward fashion like western scripts. Examples of such writing systems are Arabic, traditional Mongolian and Brahmic (Indic) families such as Devanagari or Dravidian scripts and also Thai alphabet. In the past few decades, several character encoding schemes and standards for traditional Mongolian script were developed. This article reviews the rendering and encoding issues of traditional Mongolian script, which is considered to be one of the most complex writing systems in the world.

## 2 Traditional Mongolian script and Mongolian language

Mongolian is spoken by most of the Mongolian population as well as Inner Mongolians and other groups of people who live in several provinces of China and the Russian Federation (Poppe 1954). Mongols have used numerous writing systems such as Square or Phags-pa script, Clear script (Todo), Soyombo script and Horizontal square script, Vaghintara script and others, which were not as prominent. Traditional Mongolian script was the most popular writing system (Shagdarsuren 2001).

### 2.1. Traditional Mongolian script

Traditional Mongolian script is written vertically from top to bottom in columns advancing from left to right. This script has four derivative scripts: Todo, Manchu, Vaghintara, and Sibe (Xibe). The Todo script was used by the Oirats and Kalmyks, and the Manchu script was a writing system in the Qing dynasty. The Sibe script is used in Xinjiang, in the northwest of China. The Vaghintara script was used by the Buryats. Traditional Mongolian script has thirty four letters, which represent the sounds of Mongolian language. Of these, seven are vowels, twenty one are core consonants and six are consonants for writing foreign words. Some letters, such as “t” and “d”; “o” and “u”; and “oe” and “ue” have the same appearance in the certain position of words (Chinggaltai 1963; Choimaa 1991). Similar to Arabic, traditional Mongolian is a contextual script where letters are cursively joined and have initial, medial, and final presentation forms for the same letter (Table 1). In some cases, the letters join together along a vertical stem, but in the case of certain consonants, which lack a trailing vertical stem, they may form a single ligature with a following vowel (Figure 1). Over two hundred ligatures are included in technical report No. 170 for Unicode (Erdenechimeg, Moore et al. 1999). Some fonts included much more traditional Mongolian ligatures. Usually, in ancient time Mongols did not separate the vowels and the consonants but used syllables consisting of the vowel and consonants, which form a single ligature. In addition to cursive and positional forms, many letters also have variant forms used in accordance with spelling and grammatical rules. The actual number of glyphs would be

much higher than thirty four letters in traditional Mongolian. Thus, the traditional Mongolian script is regarded as complex. Encoding complex script features, as well as understanding the layout features and rules exclusively related to traditional Mongolian script are crucial for researchers and developers.

Mongolian letters		Isolate forms	Initial forms	Medial forms	Final forms
Naming in Unicode	Code point in Unicode				
'a'	U+1820				
'o'	U+1823				
'u'	U+1824				
'oe'	U+1825				
'ba'	U+182A				
'ta'	U+1832				
'da'	U+1833				

**Table 1.** Initial, medial and final forms of some traditional Mongolian letters (Erdenechimeg, Moore et al. 1999; Erdenechimeg, Moore et al. 2000)

A consonant + a vowel	The correct forms	Incorrect forms	Examples of other ligatures			

**Figure 1.** Some ligatures in traditional Mongolian script (Erdenechimeg, Moore et al. 1999; Erdenechimeg, Moore et al. 2000)

## 2.2. Grammar and writing rules of traditional Mongolian script

Some important elements and grammatical rules, which are necessary for the character encoding of traditional Mongolian script are explained in 2.2.1, 2.2.2, 2.2.3, 2.2.4, and 2.2.5.

### 2.2.1. Syllables

A syllable is formed by joining together a vowel and a consonant. Without a vowel it is not considered as a syllable. However, mostly in the beginning of the word, a vowel forms a syllable. There are four types of syllables: 1) vowels only, 2) consonants followed by vowels, 3) vowels followed by consonants, and 4) syllables composed of three letters; a consonant and a vowel followed by another consonant. Depending on the combination of syllables, the glyph form of letters and ligatures vary (Chinggaltai 1963). Some syllables appear as a ligature, which are important to render traditional Mongolian script properly. Figure 1 shows some ligatures for the consonant followed by a vowel syllable of letters ‘ba’, ‘pa’, ‘qa’ and ‘ka’.

### 2.2.2. Vowel harmony

Traditional Mongolian script has a characteristic feature of “vowel harmony”, whereby a word can only contain either back vowels (‘a’, ‘o’, ‘u’) or front vowels (‘e’, ‘oe’, ‘ue’), but not both at the same time, with the exception only of a certain limited set of words, the majority of which are foreign words. In other words, whatever the first vowel a word has, that vowel decides the characters of the vowels in the following syllables of the word. That is to say, the vowels in a word are either all “masculine” and “neutral” (that is, back vowels plus ‘i’) or all “feminine” and “neutral” (that is, front vowels plus ‘i’). Words that are written with masculine/neutral vowels are considered to be masculine, and words that are written with feminine/neutral vowels are considered to be feminine. The vowel ‘i’ is considered neutral and can therefore occur in both front and back vowel words, but when ‘i’ occurs in all syllables the word is considered to be front vowel and behaves as feminine, for example, taking feminine suffixes.

Vowel harmony is an important element of the encoding model, as the gender of a word determines the glyph form of the velar series of consonant letters for traditional Mongolian script. In traditional Mongolian script, the velar letters (‘qa’ and ‘ga’) have both masculine and feminine forms. The masculine and feminine forms of these letters have different pronunciations. When one of the velar consonants precedes a vowel, it takes the masculine form before masculine vowels, and the feminine form before feminine or neutral vowels. In the latter case, a ligature of the consonant and vowel is required. When one of these consonants precedes another, or is the final letter in a word, it may take either a masculine or feminine glyph form, depending on its context (Chinggaltai 1963; Choimaa 1991; Choimaa and Zayabaatar 2007). Consequently, the rendering system should automatically select the correct gender form for these letters based on the gender of the word (Erdenechimeg, Moore et al. 1999; Erdenechimeg, Moore et al. 2000; Unicode 2007). Vowel harmony in traditional Mongolian script is illustrated in Figure 2.

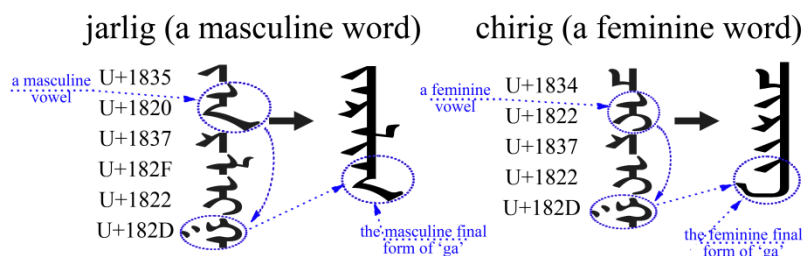


Figure 2. Traditional Mongolian script's gender forms (Unicode 2007)

### 2.2.3. The 'a' and 'e' in a word-final position

In traditional Mongolian script, the letters 'a' and 'e' in a word-final position may take an "offshoot forward tail" or "backward tail" form depending on the preceding consonant that they are attached to. Certain words, for example with a final letter 'a' or 'e' are separated from the preceding consonant including 'ga', 'qa', 'ma', 'na', and 'ra'; by a narrow gap; in that case the vowel always takes the "offshoot forward tail" form, and the 'a' or 'e' written in the "offshoot" form is an integral part of the word stem (Figure 3). Whether a final letter 'a' or 'e' is joined or separated is purely lexical and is not a question of varying orthography.


















### 2.2.4. Closed syllable consonants

Traditional Mongolian script has another important rule – closed syllable consonants. Consonants, which exist at the end of a syllable or a word, and the following letter is not a vowel, are considered as the closed syllable consonants. This rule could be interpreted as "a consonant final syllable rule".

Depending on whether a consonant is syllable closed or not, the glyph form of that consonant varies, taking either a masculine, feminine or variant glyph form. There are two types of closed syllable consonants: "soft" and "hard". Hard consonants are the traditional Mongolian letters 'ba', 'ga', 'ra', 'sa', 'ta', and 'da'. Letters 'ma', 'la', 'na', 'ang', 'wa', and 'ya' are soft closed syllable consonants. If a letter is closed syllable consonant, the glyph must have a form which is shown in Table 2. Closed syllable consonants affect the form of the following suffixes (Choimaa 1991). In this way, the rendering system should automatically select the correct glyph form for closed syllable consonants.

Table 2. Basic forms of the closed syllable consonants (Choimaa 1991).

Positions	Soft closed syllable consonants					
	'ba' U+182A	masculine form of 'ga' U+182D	feminine form of 'qa' U+182D	'ra' U+1837	'sa' U+1830	'da' U+1833
Initial form	ᠪ	ᠭ	ᠠ	ᠷ	ᠰ	ᠳ
Medial form	ᠪ	ᠭ	ᠠ	ᠷ	ᠰ	ᠳ
Final form	ᠪ	ᠭ	ᠠ	ᠷ	ᠰ	ᠳ





Positions	Hard closed syllable consonants					
	'ma' U+182E	'la' U+182F	'na' U+1828	'ang' U+1828	'wa' U+1838	'ya' U+1836
Initial form				-		
Medial form						
Final form						

### 2.2.5. Suffixes

In traditional Mongolian script, some suffixes are written separately from the stem of a word or from other suffixes. Any attached suffixes are considered to be an integral part of the word as a whole. A suffix affects the form of the preceding letters. In other words, the glyph of the final letter of the stem (or the preceding suffix) and the first letter of the particular suffix could vary, depending on the particular suffix. Suffixes have masculine and feminine pairs (for example, -dur/-tur and -dür/-tür), and a stem may receive multiple suffixes. Only separately written suffixes are explained here, because these are vital for rendering traditional Mongolian script properly.

#### 2.2.5.1. Plural suffixes

Plural nouns use the plural suffixes shown in Table 3, which are written separately from the stem.
















Plural suffixes	Shape of the initial glyph	Special attention
 nugud	the first initial form of 'na'-U+1828	Added to masculine words ending in the consonant 'na' and vowels.
 nügüd	the first initial form of 'na'-U+1828	Added to feminine words ending in the consonant 'na' and vowels.
 ud/üd	the first medial form of 'u'-U+1824	Added to words ending in consonants except 'na'. Pronunciation and encoding varies according to vowel harmony.
 nar/ner	the first initial form of 'na'-U+1828	Used for humankind. Pronunciation and encoding varies according to vowel harmony.

**Table 3.** Some plural suffixes of traditional Mongolian (Chinggaltai 1963; Choimaa 1991; Choimaa, Bayarsaikhan et al. 2005; Choimaa and Zayabaatar 2007).

#### 2.2.5.2. Case suffixes

Almost all case suffixes are separated from the stem of a word or from other suffixes by a narrow gap. The final letter of the stem or suffix preceding the particular suffix takes the final positional form, whereas the first letter of the particular suffix usually takes a medial form or a final form (single letter suffixes), depending on the particular suffix.

Some special cases where the first letter of the suffix takes a normal initial form and a variant initial form are listed in Table 4.

Case	Case suffixes	Shape of the initial glyph	Special attention	
Genitive case		u/ü	the first final form of 'u'	Added to words ending in the letter "na". Pronunciation and encoding varies according to vowel harmony.
		un/ün	the first medial form of 'u'	Added to words ending in consonants other than "na". Pronunciation and encoding varies according to vowel harmony.
		yin	the first initial form of 'ja'	Added to words ending in vowels or diphthongs.
Accusative case		i	the final form of 'i'	Added to words ending in consonants.
		yi	the first initial form of 'ja'	Added to words ending in vowels and diphthongs.
Dative-locative case		dur/ dür	the initial form of 'da'	Added to words ending in vowels and soft closed syllable consonants. Pronunciation and encoding varies according to vowel harmony.
		du/dü		
		tur/tür	the initial form of 'ta'	Added to words ending in hard closed syllable consonants. Pronunciation and encoding varies according to vowel harmony.
		tu/tü	the second final form of 'e'	An ancient form of dative-locative case. Pronunciation and encoding varies according to vowel harmony.
Ablative case		ača/ eče	the initial form of 'e'	Pronunciation and encoding varies according to vowel harmony.
Comitative case		luya	the initial form of 'la'	Added to masculine words.
		lüge		Added to feminine words.
		tai/tei	the initial form of 'da'	Pronunciation and encoding varies according to vowel harmony.
Instrumental case		iyar/ iyer	the first initial form of 'ja'	Added to words ending in consonants. Pronunciation and encoding varies according to vowel harmony.
		bar/ ber	the initial form of 'ba'	Added to words ending in vowels and diphthongs. Pronunciation and encoding varies according to vowel harmony.

**Table 4.** Some case suffixes of traditional Mongolian (Chinggaltai 1963; Choimaa 1991; Choimaa, Bayarsaikhan et al. 2005; Choimaa and Zayabaatar 2007).

### 2.2.5.3. Possessive suffixes

The following possessive suffixes as shown in Table 5 are written separately from the stem. In traditional Mongolian, the reflexive-possessive suffix is added to the end of the stem with no case suffix. This serves as the accusative form. Other forms, which the reflexive-possessive suffix is applied after the case suffixes are shown in Table 6. Moreover, as shown in Table 6, there are alternative ways to apply a single suffix, which combines a case suffix and reflexive-possessive suffix.




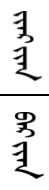



Knowing the separately written suffixes is important for rendering traditional Mongolian script properly. Depending on the particular suffix, the glyph of the final letter of the stem (or the preceding suffix) and the first letter of the particular suffix vary. Any rendering system should consider all the above rules and also must select the correct glyph form for a letter according to the grammatical rules of traditional Mongolian.

Suffixes		Shape of the initial glyph	Special attention		
Reflexive-possessive suffixes	ban	the first initial form of 'ba'	Added to words ending in vowels and diphthongs. Pronunciation and encoding varies according to vowel harmony.		
	iyen	the first initial form of 'ja'			
Possessive suffixes for possessive pronouns	mini	the first initial form of 'ma'	Added to all words.		
	minu				
	mani				
	manu				
	second-person	čini		the initial form of 'cha'	
		činu		the initial form of 'ta'	
		tani			
		tanu			
	third-person	anu			the initial form of 'a'
		ni		the second medial form of 'na'	Added to all words.
inu		the initial form of 'i'	Added to feminine words.		
Possessive suffix	ki	the ligature 'ki'	Added to all words.		
	kin	the ligature 'ki'			

**Table 5.** Some possessive suffixes of traditional Mongolian(Chinggaltai 1963; Choimaa 1991; Choimaa, Bayarsaikhan et al. 2005; Choimaa and Zayabaatar 2007).

Suffixes		Abbreviated version	Shape of the initial glyph	Special attention
Genitive case with reflexive-possessive suffixes	u-ban ü-ben	-	the first final form of 'u' and the initial form of 'ba'	Added to words ending in consonants
	ün-iyen un-iyen	-	the first medial form of 'u' and the first initial form of 'ja'	Added to words ending in vowels and diphthongs
	yin-iyen yin-iyen	-	the first initial form of 'ja'	Added to words ending in consonants. Pronunciation and encoding varies according to vowel harmony
Dative-locative case with reflexive-possessive suffixes	du-ban d ü-ben	dagan degen	the initial form of 'da' and the first medial form of 'ba' or the initial form of 'da' and the first initial form of 'ja' or the initial form of 'da'	Added to words ending in vowels and soft closed syllable consonants ('na', 'la', 'ma', 'wa', and 'ang'). Pronunciation and encoding varies according to vowel harmony.
	tur-iyen t ür-iyen	tagan tegen	the initial form of 'ta' and the first medial form of 'ba' or the initial form of 'ta' and the first initial form of 'ja' or the initial form of 'ta'	Added to words ending in hard closed syllable consonants ('ba', 'qa', 'ga', 'ra', 'sa', 'ta', and 'da'). Pronunciation and encoding varies according to vowel harmony.
	i-ben i-ban	yügen yugan	the final form of 'i' and the first initial form of 'ba' or the first initial form of 'ya'	Added to words ending in consonants. Pronunciation and encoding varies according to vowel harmony.
	yi-ben yi-ban	yügen yugan	the final form of 'i' and the first initial form of 'ba' or the first initial form of 'ya'	Added to words ending in vowels and diphthongs. Pronunciation and encoding varies according to vowel harmony.



Ablative case with reflexive-possessive suffixes		ača-ban		acagan	the initial form of ‘e’	Added to masculine words
		eče-ben		etsegen		Added to feminine words
Instrumental case with reflexive-possessive suffixes		iyar- iyan iyer- iyen	-	-	the first initial form of ‘ja’	Added to words ending in consonants. Pronunciation and encoding varies according to vowel harmony
		bar- iyan ber- iyen	-	-	the initial form of ‘ba’ the first initial form of ‘ja’	Added to words ending in vowels and diphthongs. Pronunciation and encoding varies according to vowel harmony
Comitative case with reflexive-possessive suffixes		tai-ban		taigaan	the initial form of ‘ta’ or the initial form of ‘ba’	Added to masculine words
		tei-ben		teigaan		Added to feminine words

**Table 6.** Some possessive suffixes with case suffixes of traditional Mongolian (Chinggaltai 1963; Choimaa 1991; Choimaa, Bayarsaikhan et al. 2005; Choimaa and Zayabaatar 2007).

### 3 Character encodings for traditional Mongolian script

This section discusses the character encoding issues of the traditional Mongolian script. Solving the encoding issues of the traditional Mongolian script is the key fundamental challenge to succeed in the future development of Mongolian information processing. Several non-Unicode character encodings and code pages were developed and some of them are still commonly used. As a result, diverse character encodings, code pages and keyboard layouts/locales preclude enrichment of the common text corpora for traditional Mongolian script. Although traditional Mongolian script was standardized in ISO/IEC 10646 and Unicode in 1999, there was no support for Unicode of traditional Mongolian from the major vendors until the release of Windows Vista in 2007. Prior to the Unicode standard, Mongolia did not have a national character encoding standard like American ASCII or the Japanese JIS for the character encoding of traditional Mongolian script. In China (PRC), GuoBiao established 7 and 8-bit encoding standard GB 8045-87 for traditional Mongolian character set. In Mongolia GB 8045-87 was not used. The traditional Mongolian script has been standardized in the Chinese standard–GB18030-2000 by adopting the previous standard – GB8045-87. Next section discusses the complexities of traditional Mongolian script encoding, disparities between Unicode and other character encodings, and their implementations.

#### 3.1. Differences on character encodings







In the past, several character encoding schemes and code pages were developed for traditional Mongolian script. These character encoding schemes and code pages depended on what sets the designer chose for canonical letters. In fact, the designer of an encoding scheme decides whether few representative letters or all shape variants should be used as canonical letters. When the ASCII was the main character encoding of computer systems, most systems were not able to include all shape variants of traditional Mongolian script. In

most cases, the selected representative letters were used as canonical letters by overwriting Basic Latin (U+0020~U+007E) and the Latin-1 Supplement (U+00A0~U+00FF) code range.

Several encoding schemes code pages were introduced, including the "Sudar" package of the National University of Mongolia (1991-92), MBE (Mongol Bichig Editor) for MS-DOS by Peter Cheung, Mongolian Language Support (Corff 1999) for UNIX/MS-DOS, Boljoo Input Method Editor (IME) (Boljoo 2006) Saiyin IME (Saiyinbilig 2006) and Menksoft Mongolian IME (MüנגgeGal 2005). Some approaches use transcriptions and transliteration techniques in addition to code page encoding. Other approaches include all the canonical letters in code page while canonical letters may be composed out by several glyphs in some approaches. Some implementations and research utilized encoding schemes above. For instance, BabelPad Unicode text editor for Windows (West 1999) uses Unicode, Mongolian-English dictionary by Lingua Mongolia (Pugh 2009) uses the encodings of Peter Cheung, a method for electronizing the traditional Mongolian script (Dula, Fujii et al. 2005) and Mongolia language system by (Chigen 2004) utilised the encodings of Oliver Corff. Sample encoding of the Mongolian word "father /abu/" is shown in Table 7.

Encoding Scheme (Codepage)	Encoded Text	Codes	Note
CMs (Classical Mongolian script) by Peter Cheung	FeB	U+0046, U+0065, U+0042	Widely used in Mongolia
GB8045-87	@M	U+0040, U+004D	Replaced by GB 18030-2000
Mongolian Language Support by Oliver Corff	Ab ü	U+0041, U+0062, U+00FC	Stores transcriptions, converts to Mongolian at the time to process
Boljoo IME	Private Use Area (PUA)	U+E61E, U+E686	Uses the PUA at the range of U+E610-U+7BD.
Menksoft Mongolian IME		U+E266, U+E2C6, U+E287	Uses the PUA at the range of U+E234-U+E34F. Widely used in Inner Mongolia.
Saiyin IME		U+E246, U+E247, U+E25C	Uses the PUA at the range of U+E235-U+E354.

Table 7. The Mongolian word abu:  father in different encoding schemes.

Among many codepages, Classical Mongolian Scripts (CMs) by Peter Cheung (aka dEgi, Taiwan) became de facto for desktop publishing applications, digital typesetting, and word processing in Mongolia until Microsoft Vista's built-in IME support was released. A set of TrueType fonts CMs Ulaanbaatar, CMs Huree and CMs Urga is widely used because of its simplicity. In China, legacy encodings such as the Private Use Area (PUA) mappings and GB18030 mappings are more commonly used instead of Unicode, for making web pages and electronic documents in Mongolian. Menksoft Mongolian IME, which uses the PUA at the range of U+E234-U+E34F, became de facto for Inner Mongolians for storing traditional Mongolian text. Saiyin IME and Boljoo IME are also commonly used in Inner Mongolia. Approaches such as GB8045-87 and Boljoo IME use ligatures to encode text, whereas Menksoft Mongolian IME and Saiyin IME encode all the positional and variant glyph forms of a letter separately. For instance, Menksoft encodes the initial form of Mongolian letter 'ue'-  in the U+E2A2, the medial form-  in the U+E2AB, and the final form -  in the U+E2A3. Similarly, encoded code points for the letter 'o' are as follows: the initial form  is U+E28C, the medial form  is U+E291, and the final form  is U+E2A3. Instead of using the Unicode basic character sets, Menksoft uses those variant forms for storing traditional Mongolian content.

An appropriate character encoding scheme for traditional Mongolian script will have positive consequences in the field of natural language processing, data exchange, building of library catalogues, indexing, information retrieval, etc. However, any texts in non-standard character encoding schemes that utilized PUA; or overwrote Basic Latin or Latin-1 Supplement code range are not indexed properly by retrieval systems or web search engines, such as Google, Yahoo, Bing, etc. Although traditional Mongolian script has been standardized in Unicode, use of Unicode standard is not widely spread among researchers and developers due to poor support for rendering traditional Mongolian script at the operating systems level, complexities of the Unicode Mongolian encoding model and the lack of a clear definition for the use of special characters. Moreover, in this study, we found that unicoded traditional Mongolian texts were not rendering correctly in any popular platforms.

### **3.2. Traditional Mongolian script in the Unicode standard**

From the Unicode standard version 3.1 to the latest Unicode standard version 6.0, traditional Mongolian script has been standardized in Unicode and isolated form for the vowels, and the initial form for the consonants are encoded at the range of U+1800-U+18AF (Unicode 2007). The Unicode standard includes basic character sets and special punctuation symbols and numerals, but does not explicitly encode the variant forms or the ligatures, although the correct variant form or ligature can, in most cases, be determined from the context

#### **3.2.1. Representative glyphs**

The encoded characters in the Unicode range at U+1800-U+18AF are the isolated forms for the vowels and the initial forms for the consonants. Letters that share the same glyph forms are distinguished by using different positional forms for the Mongolian code range. For example, the representative glyph for U+1823 (Mongolian letter ‘o’) is in the isolated form, whereas the representative glyph for U+1824 (Mongolian letter ‘u’) is in the initial form. The various positional and variant glyph forms of a letter are considered as presentation forms. It is the responsibility of the rendering system to select the correct glyph form for a letter according to its context. Thus, having a robust rendering algorithm is vital for displaying traditional Mongolian script correctly. All the representative glyphs are listed in technical report No. 170 for traditional Mongolian script in the ISO/Unicode standards (Erdenechimeg, Moore et al. 1999), but not in Unicode standard.

#### **3.2.2. Variant forms: Free Variation Selectors**

Free variation selectors are encoded in Unicode for traditional Mongolian script, when a glyph form cannot be predicted algorithmically by the rendering system. Those are:

- U+180B Mongolian Free Variation Selector One (FVS1);
- U+180C Mongolian Free Variation Selector Two (FVS2); and
- U+180D Mongolian Free Variation Selector Three (FVS3).

Users need to append an appropriate variation selector to the letter for indicating to the rendering system which glyph form is required. These format characters normally have no visual appearance. When required, a free variation selector immediately follows the base character it modifies. This combination of base character and variation selector is known as a standardized variant<sup>1</sup> (Unicode 2007).

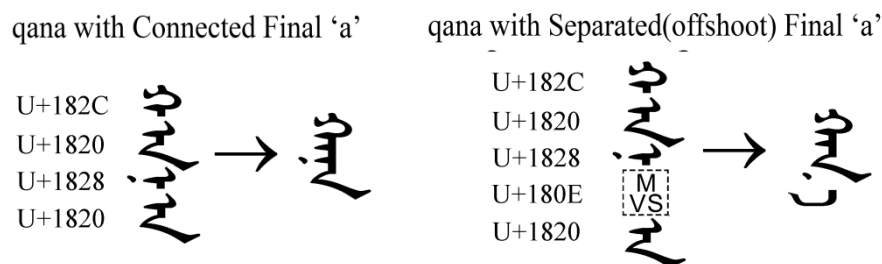
### 3.2.3. Narrow No-Break Space

The Narrow No-Break Space (NNBSP)–U+202F is encoded to define traditional Mongolian suffixes as an integral part of the word as a whole. Basically, a line break opportunity does not occur before a suffix, and whitespace is represented when using NNBSP (Unicode 2007).

### 3.2.4. Mongolian Vowel Separator

The Mongolian Vowel Separator (MVS)–U+180E is used to represent the whitespace that separates a final letter ‘a’ or ‘e’ from the rest of the word. MVS is very similar in function to NNBSP, as it divides a word with a narrow non-breaking whitespace. Whereas NNBSP marks off a grammatical suffix, the ‘a’ or ‘e’ following MVS is not a suffix but an integral part of the word stem. For example, the word ‘qana’ or ‘hana’ without a gap before the final letter ‘a’ means, “the outer casing of a vein,” whereas the word ‘qana’ with a gap (MVS) before the final letter ‘a’ means, “the wall of a tent”, as shown in Figure 3. The words ‘qana’ are encoded “U+182C, U+1820, U+1828, U+1820” and “U+182C, U+1820, U+1828, U+180E, U+1820” respectively.

The MVS always selects the offshoot forward tail form of a following vowel ‘a’ or ‘e’. In addition, it may affect the form of the preceding letter. The particular form that is taken by a letter preceding an MVS depends on the particular letter and in some cases on whether traditional or modern orthography is being used (Unicode 2007).



**Figure 3.** Use of Mongolian Vowel Separator (Unicode 2007).

<sup>1</sup> <http://www.unicode.org/Public/UNIDATA/StandardizedVariants.html>

### **3.3. Rendering traditional Mongolian script**

Recently, rendering traditional Mongolian script has begun to be supported in most Windows applications (including Microsoft Office Publisher, most Adobe applications, and Microsoft Office from version 2003), Windows Vista and Windows 7. Microsoft Windows' multilingual text rendering engine – Uniscribe (Unicode script processor) – is able to render the traditional Mongolian script which is encoded in Unicode with the support of OpenType fonts. In older versions of Windows, OpenType support for traditional Mongolian script could be added by updating the Uniscribe driver to the latest version.

Some open source software, such as OpenOffice uses the International Components for Unicode (ICU)<sup>2</sup> libraries, which is developed by IBM, for Unicode algorithms and glyph layout in complex text layout including traditional Mongolian script.

In Unix-like systems, there are active developments, such as Pango<sup>3</sup> and HarfBuzz<sup>4</sup>. In Macintosh or Mac, the Apple Type Services for Unicode Imaging (ATSUI) is used for rendering Unicode-encoded text. Apple Advanced Typography (AAT) support with ATSUI enables applications to provide users with AAT fonts having rules of advanced font rendering and supporting internationalization and complex text layout features. Furthermore, XenoType Technologies is working to release a Mongolian Language Kit<sup>5</sup> for Mac OS X.

Rendering engines including Uniscribe, ATSUI, Pango and HarfBuzz are essential to render Unicode-encoded text in different platforms, especially complex text layout. However, fonts in OpenType or AAT format are more vital to embed the language specific rules in the fonts. Therefore, handling the grammatical rules in the OpenType or AAT fonts is crucial to include the error-free Mongolian rules. Several OpenType fonts were developed and so far, no single AAT font is available yet for traditional Mongolian.

Moreover, no attempt has been carried out to check the rendering systems of traditional Mongolian script and shaping algorithms of the OpenType fonts. Although simple Mongolian words and ligatures render correctly in some systems, we found that there are still problems with complex features of the traditional Mongolian script such as MVS, FVS1-3, NNBS, contextual gender forms, etc.

### **3.4. OpenType format**

The OpenType format is a cross-platform compatible font format developed jointly by Adobe and Microsoft. The OpenType supports widely expanded multilingual character sets and layout features, which provides richer linguistic support and advanced typographic control, such as ligatures, glyph substitution, swash variants, kerning, and more. OpenType fonts allow the embedding of traditional Mongolian script rules in a single file. Microsoft developed the guidelines for creating and supporting OpenType fonts for traditional Mongolian Script (Microsoft 2002).

Several OpenType fonts – Code2000, Simsun-18030, Daicing fonts, Manchu Font 2005, Mongolian Baiti, MongolUsug and MongolianScript were developed for traditional

---

<sup>2</sup> <http://www.icu-project.org/>

<sup>3</sup> <http://www.pango.org/>

<sup>4</sup> <http://www.freedesktop.org/wiki/Software/HarfBuzz>

<sup>5</sup> <http://www.xenotypetech.com/osxMongolian.html>

Mongolian script. We surveyed possible candidate fonts Mongolian Baiti, MongolUsug, and MongolianScript for testing to render traditional Mongolian script correctly. Other fonts failed to render simple Mongolian words and ligatures.

### 3.4.1. Mongolian Baiti

This font was developed by the Founder Corporation, Peking University. It is important to note that Microsoft distributed Mongolian Baiti with Windows Vista and Windows 7. The latest version is 5.01. Isolated forms for the vowels and the initial form for the consonants are encoded at the range of U+1800–U+18AF. The rest of the variant forms of glyphs and presentation forms are stored within the font file and indexed by Glyph Index (GID).

### 3.4.2. MongolianScript

This font was developed by Erdenechimeg Myatav, the task force member of the proposal “Traditional Mongolian Script in the ISO/IEC 10646 and Unicode Standards”<sup>6</sup> (Erdenechimeg, Moore et al. 1999). The latest version is 2.0. Variant forms of glyphs or presentation forms are encoded at the range of U+F300–U+F3B0. The ligature set is encoded at the range of U+F400–U+F4C1.

### 3.4.3. MongolUsug

This font was distributed with traditional Mongolian script editor–VertNote<sup>7</sup>. The latest version is 2.37. Variant forms of glyphs, presentation forms, and ligatures are encoded in the PUA at the range of U+E000–U+E811.

## 4 Test on rendering traditional Mongolian script

We conducted a preliminary experiment to examine the rendering algorithm of OpenType fonts as well as to check that traditional Mongolian script content is displayed correctly. Rendering algorithms of the three Mongolian OpenType fonts Mongolian Baiti, MongolianScript, and MongolUsug, were reviewed. In addition to checking the basic rules of traditional Mongolian script, we tested the following complex grammatical rules: 1) vowel harmony, feminine and masculine words; 2) syllables and closed syllable consonants; 3) suffixes; and 4) usage of free variant selectors.

We rendered text files with the same words of traditional Mongolian script in the different OpenType fonts, and compared the rendered results with the correct forms. We selected masculine and feminine words with closed syllable consonants or suffixes. We tested all 43 standardized variants of traditional Mongolian script. The experimental setup is for the English version of Windows 7, build 7600.16617 with Internet Explorer version 8.0 and Uniscribe version 1.626.

<sup>6</sup> <http://www.iist.unu.edu/newrh/III/1/docs/techreports/report170a.tgz>

<sup>7</sup> <http://www.uukhai.com/archives/1152>

#### 4.1. Findings and errors

The rendered results of our experiment, which surveys the rules of traditional Mongolian script, were irregular in all fonts as illustrated in Table 8. For instance, the closed syllable consonant ‘ga’-U+182D has rendered correctly in some words, but not in all. All fonts failed to render the closed syllable consonants ( ⑦, ⑩ and ⑬ ). Vowel harmony was not considered in several words and the velar letter ‘ga’-U+182D of the feminine word was rendered incorrectly( ① and ⑦ ). In addition, MongolianScript failed to render several suffixes ( ④, ⑧, ⑨, ⑫, ⑬ and ⑯ ), though MongolUsug ( ②, ⑤, ⑥ and ⑪ ) and Mongolian Baiti ( ② and ⑩ ) failed on few suffixes. Other errors which encountered are: 1) forming ligatures or syllables incorrectly ( ③, ⑱ and ⑳ ); 2) forming diphthongs incorrectly ( ⑨ and ⑳ ); and 3) rendering the standardized variant glyphs incorrectly( ⑬, ⑲ and Section 4.2). Errors which occurred are explained in the last column in Table 8.

Furthermore, some minor errors were found. For instance Unicode standard defines the glyph of the Mongolian letter ‘o’-U+1823 as  $\mathfrak{a}$ , ‘u’-U+1824 as  $\mathfrak{a}$ , ‘oe’-U+1825 as  $\mathfrak{a}$ , and ‘ue’-U+1826 as  $\mathfrak{a}$ . However, Mongolian Baiti and MongolianScript used the single glyph  $\mathfrak{a}$  for both letters ‘o’-U+1823 and ‘u’-U+1824; and  $\mathfrak{a}$  for ‘oe’-U+1825 and ‘ue’-U+1826.

Correct form	Character Sequence		Pronunciation and English meaning	Rendered forms in OTFs			Errors
	Basic Glyph	Unicode		Mongolian Script	Mongol Usug	Mongolian Baiti	
ᠪᠢᠴᠢᠭ	ᠪᠢᠴᠢᠭ	U+182A	bičig	ᠪᠢᠴᠢᠭ	ᠪᠢᠴᠢᠭ	ᠪᠢᠴᠢᠭ	①: Vowel harmony is not considered. The final velar letter ‘ga’-U+182D of the feminine word is rendered incorrectly.
	ᠪᠢᠴᠢᠭ	U+1822					
	ᠪᠢᠴᠢᠭ	U+1834					
	ᠪᠢᠴᠢᠭ	U+1822					
ᠡᠭᠢᠰᠢᠭ ᠢᠨᠦ	ᠡᠭᠢᠰᠢᠭ ᠢᠨᠦ	U+1821	egesig in ü	ᠡᠭᠢᠰᠢᠭ ᠢᠨᠦ	ᠡᠭᠢᠰᠢᠭ ᠢᠨᠦ	ᠡᠭᠢᠰᠢᠭ ᠢᠨᠦ	②: The first letter ‘i’-U+1822 of the possessive suffix ‘inü’ is rendered incorrectly.
	ᠡᠭᠢᠰᠢᠭ ᠢᠨᠦ	U+182D					
	ᠡᠭᠢᠰᠢᠭ ᠢᠨᠦ	U+1821					
	ᠡᠭᠢᠰᠢᠭ ᠢᠨᠦ	U+1830					
	ᠡᠭᠢᠰᠢᠭ ᠢᠨᠦ	U+1822					
	ᠡᠭᠢᠰᠢᠭ ᠢᠨᠦ	U+182D					
	NNBSP	U+202F					
	ᠡᠭᠢᠰᠢᠭ ᠢᠨᠦ	U+1822					
	ᠡᠭᠢᠰᠢᠭ ᠢᠨᠦ	U+1828					
	ᠡᠭᠢᠰᠢᠭ ᠢᠨᠦ	U+1826					
ᠪᠢᠶᠢᠭᠢᠨ ᠢᠶᠦᠭᠡᠨ	ᠪᠢᠶᠢᠭᠢᠨ ᠢᠶᠦᠭᠡᠨ	U+182A	b ijig-i-ben(y ügen)	ᠪᠢᠶᠢᠭᠢᠨ ᠢᠶᠦᠭᠡᠨ	ᠪᠢᠶᠢᠭᠢᠨ ᠢᠶᠦᠭᠡᠨ	ᠪᠢᠶᠢᠭᠢᠨ ᠢᠶᠦᠭᠡᠨ	③: The ligature ‘bü’-U+182A U+1826 or syllable ‘bü’ is formed incorrectly. ④: The accusative case suffix ‘i’ is rendered incorrectly ⑤: The accusative case with reflexive-possessive suffix ‘yügen’ is rendered incorrectly
	ᠪᠢᠶᠢᠭᠢᠨ ᠢᠶᠦᠭᠡᠨ	U+1826					
	ᠪᠢᠶᠢᠭᠢᠨ ᠢᠶᠦᠭᠡᠨ	U+1835					
	ᠪᠢᠶᠢᠭᠢᠨ ᠢᠶᠦᠭᠡᠨ	U+1822					
	ᠪᠢᠶᠢᠭᠢᠨ ᠢᠶᠦᠭᠡᠨ	U+182D					
	NNBSP	U+202F					
	ᠪᠢᠶᠢᠭᠢᠨ ᠢᠶᠦᠭᠡᠨ	U+1822					
	NNBSP	U+202F					
	ᠪᠢᠶᠢᠭᠢᠨ ᠢᠶᠦᠭᠡᠨ	U+1826					
	ᠪᠢᠶᠢᠭᠢᠨ ᠢᠶᠦᠭᠡᠨ	U+1828					
ᠴᠢᠷᠢᠭ ᠮᠠᠨᠢ	ᠴᠢᠷᠢᠭ ᠮᠠᠨᠢ	U+1834	čirig mani	ᠴᠢᠷᠢᠭ ᠮᠠᠨᠢ	ᠴᠢᠷᠢᠭ ᠮᠠᠨᠢ	ᠴᠢᠷᠢᠭ ᠮᠠᠨᠢ	⑥: The first letter ‘ma’-U+182E of the possessive suffixes ‘mani’ and ‘mini’ is rendered incorrectly
	ᠴᠢᠷᠢᠭ ᠮᠠᠨᠢ	U+1822					
	ᠴᠢᠷᠢᠭ ᠮᠠᠨᠢ	U+182F					
	ᠴᠢᠷᠢᠭ ᠮᠠᠨᠢ	U+1822					
	ᠴᠢᠷᠢᠭ ᠮᠠᠨᠢ	U+182D					
	NNBSP	U+202F					
	ᠴᠢᠷᠢᠭ ᠮᠠᠨᠢ	U+182E					
	ᠴᠢᠷᠢᠭ ᠮᠠᠨᠢ	U+1820					
	ᠴᠢᠷᠢᠭ ᠮᠠᠨᠢ	U+1828					
	ᠴᠢᠷᠢᠭ ᠮᠠᠨᠢ	U+1822					

ᠪᠠᠵᠤ	ᠪᠠᠵᠤ	U+1821	egče	ᠪᠠᠵᠤ	ᠪᠠᠵᠤ	ᠪᠠᠵᠤ	ᠪᠠᠵᠤ									
	ᠪᠠᠵᠤ	U+182D														
ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ (degen)	ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ	U+1834	vertically, abruptly	ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ	ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ	ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ	ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ									
	ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ	U+182C	kögjīm dūr iyen (degen)					ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ	ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ	ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ	ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ					
	ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ	U+1825														
	ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ	U+182D														
	ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ	U+1835														
	ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ	U+1822														
	ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ	U+182E														
	ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ	U+202F														
	ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ	U+1833										“music, sound, musical instrument” are taken the dative-locative case with reflexive-possessive suffixes	ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ	ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ	ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ	ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ
	ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ	U+1826														
	ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ	U+1837														
	ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ	U+202F														
ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ	U+1822															
ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ	U+1836															
ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ	U+1821															
ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ	U+1828															
ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ	U+1826	b üridkel iyer	ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ	ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ	ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ	ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ										
ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ	U+1826															
ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ	U+1837															
ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ	U+1822						Instrumental case of the “record, list, registration”	ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ	ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ	ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ	ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ					
ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ	U+1833															
ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ	U+182C															
ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ	U+1821															
ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ	U+182F															
ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ	U+202F															
ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ	U+1822															
ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ	U+1836															
ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ	U+1821															
ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ	U+1837															
ᠰᠡᠳᠻᠢᠯᠢ ᠮᠢᠨᠢ	ᠰᠡᠳᠻᠢᠯᠢ ᠮᠢᠨᠢ	U+1830	sedkil mini	ᠰᠡᠳᠻᠢᠯᠢ ᠮᠢᠨᠢ	ᠰᠡᠳᠻᠢᠯᠢ ᠮᠢᠨᠢ	ᠰᠡᠳᠻᠢᠯᠢ ᠮᠢᠨᠢ						ᠰᠡᠳᠻᠢᠯᠢ ᠮᠢᠨᠢ				
	ᠰᠡᠳᠻᠢᠯᠢ ᠮᠢᠨᠢ	U+1821														
	ᠰᠡᠳᠻᠢᠯᠢ ᠮᠢᠨᠢ	U+1833														
	ᠰᠡᠳᠻᠢᠯᠢ ᠮᠢᠨᠢ	U+182C														
	ᠰᠡᠳᠻᠢᠯᠢ ᠮᠢᠨᠢ	U+1822					“mind or heart, spirit, thought” with the possessive suffix ‘mini’	ᠰᠡᠳᠻᠢᠯᠢ ᠮᠢᠨᠢ	ᠰᠡᠳᠻᠢᠯᠢ ᠮᠢᠨᠢ	ᠰᠡᠳᠻᠢᠯᠢ ᠮᠢᠨᠢ						
	ᠰᠡᠳᠻᠢᠯᠢ ᠮᠢᠨᠢ	U+182F														
	ᠰᠡᠳᠻᠢᠯᠢ ᠮᠢᠨᠢ	U+202F														
	ᠰᠡᠳᠻᠢᠯᠢ ᠮᠢᠨᠢ	U+182E														
	ᠰᠡᠳᠻᠢᠯᠢ ᠮᠢᠨᠢ	U+1822														
	ᠰᠡᠳᠻᠢᠯᠢ ᠮᠢᠨᠢ	U+1828														
	ᠰᠡᠳᠻᠢᠯᠢ ᠮᠢᠨᠢ	U+1822														
	ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ	ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ									U+1826		tüledb üri d ü (n.)	ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ	ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ	ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ
ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ		U+1822														
ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ		U+182F														
ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ		U+1821														
ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ		U+1833	“factory, manufactory, mill, plant” with the dative-locative case suffix ‘dü’	ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ	ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ	ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ										
ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ		U+182A														
ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ		U+1826														
ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ		U+1837														
ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ		U+1822														
ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ		U+202F														
ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ		U+1833														
ᠲᠡᠭᠭᠢᠮᠢ ᠳᠦᠷ ᠶᠡᠨ		U+1826														
ᠵᠡᠯᠢᠭᠦᠲᠻᠠᠨ ᠦ	ᠵᠡᠯᠢᠭᠦᠲᠻᠠᠨ ᠦ	U+1835					jelig ütken ü	ᠵᠡᠯᠢᠭᠦᠲᠻᠠᠨ ᠦ	ᠵᠡᠯᠢᠭᠦᠲᠻᠠᠨ ᠦ	ᠵᠡᠯᠢᠭᠦᠲᠻᠠᠨ ᠦ	ᠵᠡᠯᠢᠭᠦᠲᠻᠠᠨ ᠦ					
	ᠵᠡᠯᠢᠭᠦᠲᠻᠠᠨ ᠦ	U+1821														
	ᠵᠡᠯᠢᠭᠦᠲᠻᠠᠨ ᠦ	U+182F														
	ᠵᠡᠯᠢᠭᠦᠲᠻᠠᠨ ᠦ	U+1822										“wilderness, empty area” with the genitive case ‘ü’	ᠵᠡᠯᠢᠭᠦᠲᠻᠠᠨ ᠦ	ᠵᠡᠯᠢᠭᠦᠲᠻᠠᠨ ᠦ	ᠵᠡᠯᠢᠭᠦᠲᠻᠠᠨ ᠦ	
	ᠵᠡᠯᠢᠭᠦᠲᠻᠠᠨ ᠦ	U+182D														
	ᠵᠡᠯᠢᠭᠦᠲᠻᠠᠨ ᠦ	U+1826														
	ᠵᠡᠯᠢᠭᠦᠲᠻᠠᠨ ᠦ	U+1833														
	ᠵᠡᠯᠢᠭᠦᠲᠻᠠᠨ ᠦ	U+182C														
	ᠵᠡᠯᠢᠭᠦᠲᠻᠠᠨ ᠦ	U+1821														
	ᠵᠡᠯᠢᠭᠦᠲᠻᠠᠨ ᠦ	U+1828														
	ᠵᠡᠯᠢᠭᠦᠲᠻᠠᠨ ᠦ	U+202F														

ᠮᠠᠵᠤ: Vowel harmony and closed syllable consonants rule are not considered. The middle velar letter ‘ga’- U+182D (hard closed syllable consonant) of the feminine word is rendered incorrectly.  
 ᠮᠠᠵᠤ: The first letter ‘d’- U+1833 of the dative-locative case suffix ‘dūr’ for the word ended in the soft closed syllable consonant is rendered incorrectly.  
 ᠮᠠᠵᠤ: The diphthong ‘iy’- U+1822 U+1836 is formed incorrectly.

ᠮᠠᠵᠤ: Closed syllable consonants rule is not considered. The closed syllable consonant ‘da’- U+1833 is displayed incorrectly.

ᠮᠠᠵᠤ: The final letter ‘ue’- U+1826 of the dative-locative case suffix ‘dü’ is rendered incorrectly.  
 ᠮᠠᠵᠤ: The first letter ‘d’- U+1833 of the dative-locative case suffix ‘dü’ is rendered incorrectly

ᠮᠠᠵᠤ: The genitive case ‘ü’ is rendered incorrectly.



ᠮᠠᠩᠭᠠᠯ ᠳᠤᠷ ᠶᠠᠨ ᠲᠠᠨ	ᠮᠠᠩᠭᠠᠯ	U+1826	mangval dur iyan (davan)	ᠮᠠᠩᠭᠠᠯ ᠳᠤᠷ ᠶᠠᠨ	ᠮᠠᠩᠭᠠᠯ ᠳᠤᠷ ᠶᠠᠨ	ᠮᠠᠩᠭᠠᠯ ᠳᠤᠷ ᠶᠠᠨ	<p>ᠮᠠᠩᠭᠠᠯ ᠳᠤᠷ ᠶᠠᠨ</p> <p>ᠮᠠᠩᠭᠠᠯ ᠳᠤᠷ ᠶᠠᠨ</p> <p>ᠮᠠᠩᠭᠠᠯ ᠳᠤᠷ ᠶᠠᠨ</p>	<p>ᠮᠠᠩᠭᠠᠯ ᠳᠤᠷ ᠶᠠᠨ</p> <p>ᠮᠠᠩᠭᠠᠯ ᠳᠤᠷ ᠶᠠᠨ</p> <p>ᠮᠠᠩᠭᠠᠯ ᠳᠤᠷ ᠶᠠᠨ</p>	<p>ᠮᠠᠩᠭᠠᠯ ᠳᠤᠷ ᠶᠠᠨ</p> <p>ᠮᠠᠩᠭᠠᠯ ᠳᠤᠷ ᠶᠠᠨ</p> <p>ᠮᠠᠩᠭᠠᠯ ᠳᠤᠷ ᠶᠠᠨ</p>	<p>ᠮᠠᠩᠭᠠᠯ ᠳᠤᠷ ᠶᠠᠨ</p> <p>ᠮᠠᠩᠭᠠᠯ ᠳᠤᠷ ᠶᠠᠨ</p> <p>ᠮᠠᠩᠭᠠᠯ ᠳᠤᠷ ᠶᠠᠨ</p>
	ᠳᠤᠷ	U+182E								
	ᠶᠠᠨ	U+1820								
	ᠲᠠᠨ	U+1829								
	ᠲᠠᠨ	U+182D								
	ᠲᠠᠨ	U+1820	The personal name "Mangal" is taken the dative-locative case with reflexive-possessive suffixes							
	ᠲᠠᠨ	U+182F								
	NNBSP	U+202F								
	ᠲᠠᠨ	U+1833								
	ᠲᠠᠨ	U+1824								
	ᠲᠠᠨ	U+1837								
	NNBSP	U+202F								
	ᠲᠠᠨ	U+1822								
ᠲᠠᠨ	U+1836									
ᠲᠠᠨ	U+1820									
ᠲᠠᠨ	U+1828									
ᠳᠠᠨᠭᠢ	ᠳᠠᠨᠭᠢ	U+1833	d ᠳᠠᠨᠭᠢ -i	ᠳᠠᠨᠭᠢ	ᠳᠠᠨᠭᠢ	ᠳᠠᠨᠭᠢ	<p>ᠳᠠᠨᠭᠢ</p> <p>ᠳᠠᠨᠭᠢ</p> <p>ᠳᠠᠨᠭᠢ</p>	<p>ᠳᠠᠨᠭᠢ</p> <p>ᠳᠠᠨᠭᠢ</p> <p>ᠳᠠᠨᠭᠢ</p>	<p>ᠳᠠᠨᠭᠢ</p> <p>ᠳᠠᠨᠭᠢ</p> <p>ᠳᠠᠨᠭᠢ</p>	
	FVS1	U+180B								
	ᠳᠠᠨᠭᠢ	U+1826	“total, sum, result” with the accusative case suffix							
	ᠳᠠᠨᠭᠢ	U+1829								
	NNBSP	U+202F								
ᠳᠠᠨᠭᠢ	U+1822									
ᠰᠣᠳᠨᠠᠮ ᠠᠴᠠ ᠪᠠᠨ	ᠰᠣᠳᠨᠠᠮ	U+1830	sodnam ača ban (ačavan)	ᠰᠣᠳᠨᠠᠮ ᠠᠴᠠ ᠪᠠᠨ	ᠰᠣᠳᠨᠠᠮ ᠠᠴᠠ ᠪᠠᠨ	ᠰᠣᠳᠨᠠᠮ ᠠᠴᠠ ᠪᠠᠨ	<p>ᠰᠣᠳᠨᠠᠮ ᠠᠴᠠ ᠪᠠᠨ</p> <p>ᠰᠣᠳᠨᠠᠮ ᠠᠴᠠ ᠪᠠᠨ</p> <p>ᠰᠣᠳᠨᠠᠮ ᠠᠴᠠ ᠪᠠᠨ</p>	<p>ᠰᠣᠳᠨᠠᠮ ᠠᠴᠠ ᠪᠠᠨ</p> <p>ᠰᠣᠳᠨᠠᠮ ᠠᠴᠠ ᠪᠠᠨ</p> <p>ᠰᠣᠳᠨᠠᠮ ᠠᠴᠠ ᠪᠠᠨ</p>	<p>ᠰᠣᠳᠨᠠᠮ ᠠᠴᠠ ᠪᠠᠨ</p> <p>ᠰᠣᠳᠨᠠᠮ ᠠᠴᠠ ᠪᠠᠨ</p> <p>ᠰᠣᠳᠨᠠᠮ ᠠᠴᠠ ᠪᠠᠨ</p>	
	ᠠᠴᠠ	U+1823								
	ᠠᠴᠠ	U+1833								
	ᠠᠴᠠ	U+1828								
	ᠠᠴᠠ	U+1820								
	ᠠᠴᠠ	U+182E	The personal name "Sodnom" is taken the ablative case with reflexive-possessive suffixes							
	NNBSP	U+202F								
	ᠠᠴᠠ	U+1820								
	ᠠᠴᠠ	U+1834								
	NNBSP	U+1820								
	ᠠᠴᠠ	U+202F								
ᠠᠴᠠ	U+182A									
ᠠᠴᠠ	U+1820									
ᠠᠴᠠ	U+1828									
ᠯᠬᠠᠪᠪᠠ ᠯᠭᠠ	ᠯᠬᠠᠪᠪᠠ	U+1840	Lhavba luga	ᠯᠬᠠᠪᠪᠠ ᠯᠭᠠ	ᠯᠬᠠᠪᠪᠠ ᠯᠭᠠ	ᠯᠬᠠᠪᠪᠠ ᠯᠭᠠ	<p>ᠯᠬᠠᠪᠪᠠ ᠯᠭᠠ</p> <p>ᠯᠬᠠᠪᠪᠠ ᠯᠭᠠ</p> <p>ᠯᠬᠠᠪᠪᠠ ᠯᠭᠠ</p>	<p>ᠯᠬᠠᠪᠪᠠ ᠯᠭᠠ</p> <p>ᠯᠬᠠᠪᠪᠠ ᠯᠭᠠ</p> <p>ᠯᠬᠠᠪᠪᠠ ᠯᠭᠠ</p>	<p>ᠯᠬᠠᠪᠪᠠ ᠯᠭᠠ</p> <p>ᠯᠬᠠᠪᠪᠠ ᠯᠭᠠ</p> <p>ᠯᠬᠠᠪᠪᠠ ᠯᠭᠠ</p>	
	ᠯᠭᠠ	U+1820								
	ᠯᠬᠠᠪᠪᠠ	U+182D	with Lkhagva: The personal name "Lkhagva" with the comitative case 'luga'.							
	ᠯᠬᠠᠪᠪᠠ	U+182A								
	ᠯᠬᠠᠪᠪᠠ	U+1820								
	NNBSP	U+202F								
	ᠯᠬᠠᠪᠪᠠ	U+182F								
	ᠯᠬᠠᠪᠪᠠ	U+1824								
	ᠯᠬᠠᠪᠪᠠ	U+182D								
	MVS	U+180E								
ᠯᠬᠠᠪᠪᠠ	U+1820									
ᠴᠡᠪᠡᠭᠡᠮᠡᠳᠡ ᠯᠢᠭᠡ	ᠴᠡᠪᠡᠭᠡᠮᠡᠳᠡ	U+1834	Čebegmed-lüge	ᠴᠡᠪᠡᠭᠡᠮᠡᠳᠡ ᠯᠢᠭᠡ	ᠴᠡᠪᠡᠭᠡᠮᠡᠳᠡ ᠯᠢᠭᠡ	ᠴᠡᠪᠡᠭᠡᠮᠡᠳᠡ ᠯᠢᠭᠡ	<p>ᠴᠡᠪᠡᠭᠡᠮᠡᠳᠡ ᠯᠢᠭᠡ</p> <p>ᠴᠡᠪᠡᠭᠡᠮᠡᠳᠡ ᠯᠢᠭᠡ</p> <p>ᠴᠡᠪᠡᠭᠡᠮᠡᠳᠡ ᠯᠢᠭᠡ</p>	<p>ᠴᠡᠪᠡᠭᠡᠮᠡᠳᠡ ᠯᠢᠭᠡ</p> <p>ᠴᠡᠪᠡᠭᠡᠮᠡᠳᠡ ᠯᠢᠭᠡ</p> <p>ᠴᠡᠪᠡᠭᠡᠮᠡᠳᠡ ᠯᠢᠭᠡ</p>	<p>ᠴᠡᠪᠡᠭᠡᠮᠡᠳᠡ ᠯᠢᠭᠡ</p> <p>ᠴᠡᠪᠡᠭᠡᠮᠡᠳᠡ ᠯᠢᠭᠡ</p> <p>ᠴᠡᠪᠡᠭᠡᠮᠡᠳᠡ ᠯᠢᠭᠡ</p>	
	ᠯᠢᠭᠡ	U+1821								
	ᠴᠡᠪᠡᠭᠡᠮᠡᠳᠡ	U+182A	with Tsevegmed: The personal name "Čebegmed" with the comitative case 'lüge'.							
	ᠴᠡᠪᠡᠭᠡᠮᠡᠳᠡ	U+1821								
	ᠴᠡᠪᠡᠭᠡᠮᠡᠳᠡ	U+182D								
	ᠴᠡᠪᠡᠭᠡᠮᠡᠳᠡ	U+182E								
	ᠴᠡᠪᠡᠭᠡᠮᠡᠳᠡ	U+1821								
	ᠴᠡᠪᠡᠭᠡᠮᠡᠳᠡ	U+1833								
	NNBSP	U+202F								
	ᠴᠡᠪᠡᠭᠡᠮᠡᠳᠡ	U+182F								
	ᠴᠡᠪᠡᠭᠡᠮᠡᠳᠡ	U+1826								
	ᠴᠡᠪᠡᠭᠡᠮᠡᠳᠡ	U+182D								
	ᠴᠡᠪᠡᠭᠡᠮᠡᠳᠡ	U+1821								

ᠮᠠᠩᠭᠠᠯ ᠳᠤᠷ ᠶᠠᠨ: The letter 'ga'-U+182D is rendered incorrectly, if it appeared after the letter 'ang'-U+1829 in the middle of a word.

ᠳᠠᠨᠭᠢ: The second initial form of the 'da'-U+1833 is rendered incorrectly.  
ᠳᠠᠨᠭᠢ: The medial form of the 'ue'-U+1826 is rendered incorrectly

ᠰᠣᠳᠨᠠᠮ ᠠᠴᠠ ᠪᠠᠨ: The first letter of the ablative case 'ača' (also ablative case with reflexive-possessive suffix 'ačavan') for the masculine word is rendered incorrectly.

ᠯᠬᠠᠪᠪᠠ ᠯᠭᠠ: Closed syllable consonants rule is not considered. The closed syllable consonant 'ga'-U+182D is rendered incorrectly

ᠴᠡᠪᠡᠭᠡᠮᠡᠳᠡ ᠯᠢᠭᠡ: The ligature 'gm'-U+182D U+182E is formed incorrectly.  
ᠴᠡᠪᠡᠭᠡᠮᠡᠳᠡ ᠯᠢᠭᠡ: The first letter of the comitative case 'luga' or 'lüge' is displayed incorrectly.

ᠴᠢᠮᠡᠨ ᠲᠠᠶᠢᠭᠠᠨ	ᠴ	U+183C	cēmēn(FVS1)t tayigan	ᠴᠢᠮᠡᠨ ᠲᠠᠶᠢᠭᠠᠨ	ᠴᠢᠮᠡᠨ ᠲᠠᠶᠢᠭᠠᠨ	ᠴᠢᠮᠡᠨ ᠲᠠᠶᠢᠭᠠᠨ	<p>21: The second medial form of 'na'-U+1828 is rendered incorrectly.</p> <p>22: The final form of 'ta'-U+1832 is rendered incorrectly in loan words.</p>
	ᠳ	U+1827					
	ᠳ	U+182E					
	ᠳ	U+1827	with my cement: "cement" is taken comitative case with reflexive-possessive suffix "tayigan"				
	ᠳ	U+1828					
	FVS1	U+180B					
	ᠳ	U+1832					
	NNBSP	U+202F					
	ᠳ	U+1832					
	ᠳ	U+1820					
	ᠳ	U+1836					
	ᠳ	U+1822					
ᠳ	U+182D						
ᠳ	U+1820						
ᠳ	U+1828						
ᠦᠨᠢᠶᠡ ᠲᠡᠢ ᠪᠡᠨ	ᠦ	U+1826	üniy-e tei ben (teyigen)	ᠦᠨᠢᠶᠡ ᠲᠡᠢ ᠪᠡᠨ	ᠦᠨᠢᠶᠡ ᠲᠡᠢ ᠪᠡᠨ	ᠦᠨᠢᠶᠡ ᠲᠡᠢ ᠪᠡᠨ	<p>23: The letter 'ya'-U+1836 is rendered incorrectly if it appeared before the "offshoot forward tail" form of the letter 'e'</p>
	ᠦ	U+1828					
	ᠦ	U+1822					
	ᠦ	U+1836	with my cow: "cow" is taken comitative case with reflexive-possessive suffix "teyigen"				
	MVS	U+180E					
	ᠦ	U+1821					
	NNBSP	U+202F					
	ᠦ	U+1832					
	ᠦ	U+1821					
	ᠦ	U+1836					
	ᠦ	U+1822					
	ᠦ	U+182D					
ᠦ	U+1821						
ᠦ	U+1828						
ᠦ	U+182C						
ᠶᠠᠶᠢᠨ	ᠶ	U+1823	qoyin(FVS1)-a	ᠶᠠᠶᠢᠨ	ᠶᠠᠶᠢᠨ	ᠶᠠᠶᠢᠨ	<p>24: The consonant 'na'-U+1828 is rendered incorrectly when it appeared before the "offshoot forward tail" form of the letter 'a'</p> <p>25: The diphthong 'yi'-U+1836 U+1822 is formed incorrectly</p>
	ᠶ	U+1836	after, later, behind, in the back or rear				
	ᠶ	U+1822					
	ᠶ	U+1828					
	MVS	U+180E					
	ᠶ	U+1820					
ᠠᠭᠨ	ᠠ	U+1820	angn(FVS1)-a	ᠠᠭᠨ	ᠠᠭᠨ	ᠠᠭᠨ	<p>26: The ligature 'ngn'-U+1829 U+1828 is formed incorrectly.</p>
	ᠠ	U+1829	to hunt				
	ᠠ	U+1828					
	MVS	U+180E					
	ᠠ	U+1820					

Table 8. Rendering the traditional Mongolian script in various OTFs

#### 4.2. Use of Free Variant Selectors

Mongolian Baiti failed to render 2 variants, MongolUsug failed 12 and MongolianScript failed to render 23 variants of all 43 forms. The rendered results of the selected standardized variant for traditional Mongolian script are illustrated in Table 9 and errors which occurred are highlighted.

Unicode Basic Glyph	Position	Description of variant appearance	Character Sequence	Correct Variant Glyph	Mongolian Baiti	Mongolian Script	Mongol Usug
ᠠ	isolate	second form of "a"	U+1820 U+180B	ᠠ	ᠠ	ᠠ	ᠠ
	medial			ᠠ	ᠠ	ᠠ	
	final			ᠠ	ᠠ	ᠠ	
ᠡ	medial	third form of "a"	U+1820 U+180C	ᠡ	ᠡ	ᠡ	ᠡ
	initial	second form of "e"	U+1821 U+180B	ᠢ	ᠢ	ᠢ	ᠢ
final	ᠢ			ᠢ	ᠢ		
ᠢ	medial	second form of "i"	U+1822 U+180B	ᠣ	ᠣ	ᠣ	ᠣ
	medial	second form of "o"	U+1823 U+180B	ᠤ	ᠤ	ᠤ	ᠤ
ᠣ	medial	second form of "u"	U+1824 U+180B	ᠥ	ᠥ	ᠥ	ᠥ
	medial	second form of "oe"	U+1825 U+180B	ᠦ	ᠦ	ᠦ	ᠦ
ᠤ	medial	third form of "oe"	U+1825 U+180C	ᠦ	ᠦ	ᠦ	ᠦ
	isolate	second form of "ue"	U+1826 U+180B	ᠨ	ᠨ	ᠨ	ᠨ
medial	ᠨ			ᠨ	ᠨ		
final	ᠨ			ᠨ	ᠨ		
ᠨ	medial	third form of "ue"	U+1826 U+180C	ᠨ	ᠨ	ᠨ	ᠨ
	initial	second form of "na"	U+1828 U+180B	ᠮ	ᠮ	ᠮ	ᠮ
medial	second form of "na"	ᠮ		ᠮ	ᠮ		
	third form of "na"	ᠮ		ᠮ	ᠮ		
ᠮ	medial	separate form of "na"	U+1828 U+180D	ᠮ	ᠮ	ᠮ	ᠮ
	final	alternative form of "ba"	U+182A U+180B	ᠪ	ᠪ	ᠪ	ᠪ
ᠪ	isolate	feminine second form of "qa"	U+182C U+180B	ᠪ	ᠪ	ᠪ	ᠪ
	initial	second form of "qa"		ᠪ	ᠪ	ᠪ	
	medial	third form of "qa"		ᠪ	ᠪ	ᠪ	
fourth form of "qa"		ᠪ	ᠪ	ᠪ			
ᠪ	initial	second form of "ga"	U+182D U+180B	ᠪ	ᠪ	ᠪ	ᠪ
	medial			ᠪ	ᠪ	ᠪ	
	final			feminine form of "ga"	ᠪ	ᠪ	ᠪ
ᠪ	medial	third form of "ga"	U+182D U+180C	ᠪ	ᠪ	ᠪ	ᠪ
	medial	feminine form of "ga"	U+182D U+180D	ᠪ	ᠪ	ᠪ	ᠪ
ᠪ	final	second form of "sa"	U+1830 U+180B	ᠪ	ᠪ	ᠪ	ᠪ
	final	third form of "sa"	U+1830 U+180C	ᠪ	ᠪ	ᠪ	ᠪ
ᠪ	medial	second form of "ta"	U+1832 U+180B	ᠪ	ᠪ	ᠪ	ᠪ
	initial	second form of "da"	U+1833 U+180B	ᠪ	ᠪ	ᠪ	ᠪ
medial	ᠪ			ᠪ	ᠪ		
final	ᠪ			ᠪ	ᠪ		
ᠪ	medial	second form of "ja"	U+1835 U+180B	ᠪ	ᠪ	ᠪ	ᠪ
	initial	second form of "ya"	U+1836 U+180B	ᠪ	ᠪ	ᠪ	ᠪ
ᠪ	medial	third form of "ya"	U+1836 U+180C	ᠪ	ᠪ	ᠪ	ᠪ
	final	second form of "wa"	U+1838 U+180B	ᠪ	ᠪ	ᠪ	ᠪ

Table 9. Rendering the variant glyphs of traditional Mongolian script in various OTFs

### 4.3. Usability survey

We conducted a usability survey among Mongolian users in order to identify actual problems of traditional Mongolian script in Unicode. In this survey, users were asked to compose, edit and format the unencoded traditional Mongolian document in Microsoft Word 2007 and Notepad. Most users' computer setup was the English version of Windows 7 with Internet Explorer version 8.0 and Uniscribe version 1.626.

Users' feedback of this study is summarized as follows:

- The font sizes are unequal in the fonts Mongolian Baiti, Mongolian Script and Mongol Usug. Majority of users criticized that the font size of Mongolian Baiti is

too small and they stated even 16 pt size is not satisfied for viewing traditional Mongolian text. In fact, Mongolian Baiti 17pt, Mongolian Script 11pt and Mongol Usug 13pt are used to display the content of Table 8;

- Breaks for word wrapping or line wrapping affect incorrect rendering of traditional Mongolian script. Users reported that Microsoft Word 2007 and Notepad render the words, which truncated to the next line, incorrectly. and;
- Use of the control characters (MVS, FVS1, FVS2, FVS3 and NNBS) is confusing, particularly when typing. In fact, NNBS should be used often when typing words with suffixes.

## 5 Recommendation

We surveyed the rendering issues of traditional Mongolian script and analyzed available OpenType fonts and their rendering schemes for traditional Mongolian script. The study produced some errors, and revealed grammatical rules, which are not documented in international standards. All fonts failed to display the variant glyphs with free variant selectors correctly, which were already standardized in Unicode. In general, Mongolian Baiti was better than others with few shortcomings on rendering variant forms, suffixes and feminine words.

In addition, all fonts need some improvements in their rendering algorithms. All the rendering rules, which explicitly define the traditional Mongolian script grammatical rules in full, need to be standardized in international use such as in ISO/IEC 19646 and Unicode. There are several excellent guides to traditional Mongolian script rendering behavior such as technical report No. 170 called "Traditional Mongolian Script in the ISO/IEC 19646 and Unicode Standards" (Erdenechimeg, Moore et al. 1999) and Mongolian Encoding (Choijinzhav 2000). All the variant forms and ligatures should be explicitly documented and guided using the appropriate control characters (MVS, FVS1, FVS2, FVS3 and NNBS), because the rendering of traditional Mongolian script is too much relied on the rendering engine or OpenType fonts to form the correct shape.

Furthermore, the guidelines for creating and supporting OpenType fonts for traditional Mongolian script (Microsoft 2002) need to be updated.

## 5 Acknowledgments

This work was supported in part by the Grant-in-Aid for the Global COE Program "Digital Humanities Center for Japanese Arts and Cultures (DH-JAC)" from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan, MEXT Grant-in-Aid for Strategic Formation of Research Infrastructure for Private University "Sharing of Research Resources by Digitization and Utilization of Art and Cultural Materials" (Grant Number: S0991041), and MEXT Grant-in-Aid for Young Scientists (B) "Research on Information Access across Languages, Periods, and Cultures" (Leader: Akira Maeda, Grant Number: 21700271).

The authors expressed deep appreciation to the following individuals for their contributions: Mrs Erdenechimeg Myatav, Denmark; Mrs Bayanzul Lodoisamba, Tokyo, Japan; and Mr, Badral Sanlig, OpenMN, Mongolia.

The authors also recognized the contributions made by the following individuals who provided comments on the background documents: Dr Andrew Glass, Microsoft Corporation; and Dr Deborah W. Anderson, Script Encoding Initiative, University of California, Berkeley, California, United States.

## 6 References

- Boljoo, 2006, Boljoo IME. <http://www.boljoo.com/>.
- Chigen, N., 2004, Electronizing Project of Mongolian.
- Chinggaltai, 1963, A Grammar of the Mongol Language, New York: Frederick Ungar Publishing Co.
- Choiijnzhab, 2000, Mengguwen Bianma (Mongolian Encoding), Hohhot: Inner Mongolia University Publishing House.
- Choimaa, S., 1991, Mongol bicigiin zov bicix dürmiin xuraangui. Bloomington, The Mongolia Society special papers.
- Choimaa; S., Bayarsaikhan; M., Munkh-Uchral, E. and Batkhishig, S., 2005, Mongol helnii hel züin toli bichig. Ulaanbaatar, Mongolia, Centre for Mongol Studies, National University of Mongolia.
- Choimaa, S. and Zayabaatar, D., 2007, Mongol bicigiin zov bicix dürmiin xuraangui, medleg shalgax soril. Ulaanbaatar, National University of Mongolia.
- Corff, O., 1999, Mongolian Language Support (MLS) for UNIX/MSDOS, Information on Mongolia and Academical Computing.
- Dula, M., Fujii, A. and Ishikawa. T., 2005, A Method for Electronizing the Traditional Mongolian Script and Its Application to Text Retrieval. *IEICE Transactions on Information and Systems*, Pt.2 (Japanese Edition), vol. J88-D-II, no. 10, pp. 2102-2111.
- Erdenechimeg, M., Moore, R. Namsrai, Y., 1999, Traditional Mongolian Script in the ISO/Unicode Standards. Macau, UNU/IIST, Macau.
- Erdenechimeg, M., Moore, R. Namsrai, Y., 2000,. Encoding and Implementation Issues in Standardising Traditional Mongolian Script. In 16th International Unicode Conference, Amsterdam, The Netherlands, The Unicode Consortium, 27-30 March, 2010.
- Microsoft, 2002, Creating and Supporting OpenType Fonts for the Mongolian Script, Microsoft Typography, <http://www.microsoft.com/typography/otfntdev/mongolot/>.
- MüngerGal, 2005, Menksoft Mongolian IME, Accessed 10 November, 2010, <http://www.menksoft.com/>.
- Poppe, N. N., 1954, Grammar of written Mongolian, Wiesbaden: Otto Harrassowitz.
- Pugh, E. R. E., 2009, Linguamongolia Dictionary, Linguamongolia.
- Saiyibilig, 2006, Saiyibilig IME, Accessed 10 November, 2010, <http://www.sainsoft.net/>.
- Shagdarsuren, T., 2001, Study of Mongolian scripts (Graphic study of grammatology), Ulaanbaatar, Mongolia, Urlakh Erdem Kheveleliin Gazar.
- Unicode, 2007, The Unicode Standard 5.0, Boston San Francisco New York Addison-Wesley.
- West, A., 1999, BabelPad: Unicode Text Editor for Windows.