Doc Type: Working Group Document

Title: Modifications to Mongolian Encoding in UCS

Source: China and Mongolia

Status: Member Bodies' Joint Proposal

Action: For consideration by JTC1/SC2/WG2

This proposal is for modifications to Mongolian encoding in UCS 2014. We generally agree that both Mongolian nominal characters and single presentation characters should be included in the new version of UCS. However, some additions and modification should be made.

Part One: Addition

Mongolian orthographic forms that devised in different periods and regions during the 400 years (1580's to 1921) should be included. These include special writings of more than 20 variation presentation characters, some words and other additional parts. Due to the reason that study on Mongolian writing of this period is really limited and hard to meet the requirement of encoding developing, we have organized scholars of great attainments on Mongolian literatures of this period to collect materials and study for three years. Through 3 symposiums on the issue, contents to be added are induced from quantitative and qualitative study: multi-variation forms of a, e, n, ng letters, variation of masculine and feminine consonant g, mixed use of consonant ǰ č y, multi-variation of final letters, writing of case suffixes variations etc., The purpose of adding Mongolian orthographic forms of this period is for digital preservation and study of the ancient literature of the period.

1.   new characters

The characters we are adding are:

Nominal characters: ᠀ ᠁ ᠂ ᠃ ᠄ ᠅ ᠆ ᠇ ᠈ ᠉ ᠊ ᠋

Variation Presentation characters: ᠀ ᠁ ᠂ ᠃ ᠄ ᠅ ᠆ ᠇ ᠈

᠀ ᠁ ᠂ ᠃ ᠄ ᠅ ᠆ ᠇ ᠈ ᠉ ᠊

Non-mandatory ligature： ᠀

1885 ᠀ 1886 ᠁ input of two AliGali letters

2.   (Free) variation selectors

The current three free variation selectorsare adequate for contemporary Mongolian but far from the the new contents (mentioned above) to be added . Therefore, we suggest to pick 5 common use variation selectors supplemen, while dealing with the presentation characters of written language of Middle Mongolian.

3.   BIRGA

We agree on the specifications about BIRGA in UCS of only leaving one BIRGA as a nominal character and others as "supplement" (11660-1166C).

4.   "Using variation selector characters' set within words"

The current ucs only shows the case of using variation selector character and hid the cases of not using variation selector characters. For the convenience of programming, it is suggested to include the cases of not using variation selector characters. It will be helpful to the programmers not familiar with Mongolian orthography. (See the table below)

Nominal Character			Free Variation Presentation Form which needs Free Variation Selectors			Collocation of Free Variation Selectors
Code	character	Name	Number	graphic	Name or Use	
(1820	ᠠ	MONGLIAN LETTER A			*[1] ml. a isolate form	ᠠ)
1821				ᠠ	* ml. a second isolate form	ᠠ FVS1
			(0004	ᠠ	ml. a first initial form	ᠠ ····)
			0006	᠇	ml. a second initial form	ᠠ FVS1 ····
			(0005	᠊	ml. a first medial form	···· ᠠ ····)
			0006	᠇	ml. a second medial form	····ᠠ FVS1 ····
			(0007	᠊	ml. a third medial form	▢U+202F ᠠ ····)
			(0005	᠊	ml. a first medial short form	····ᠠ ·180A ····)
			00CF	᠊	ml. a fourth medial short form	····ᠠ FVS3 ·180A ····
			00D0	᠊	ml. a sixth medial short form	····ᠠ ·180A ····
				᠊	ml. a seventh medial short form	····ᠠ ·180A ····
			(0008	᠊	ml. a first final form	····ᠠ)
			0009	᠊	ml. a second final form	····ᠠ FVS1
			0035	᠊	ml. a third final form	···· ᠊ FVS2

Code	character	Name	Number	graphic	Name or Use	Collocation of Free Variation Selectors
			00B8	━	ml. a fourth final form	···· ꓛ FVS3
			00B9	ꓛ	ml. a fifth final form	···· ꓛ
			00BA	━	ml. a sixth final form	···· ꓛ
			02E5	━	ml. a seventh final form	···· ꓛ
			02E3	━	ml. a eighth final form	···· ꓛ
（1826	ᠦ	MONGOLIAN LETTER UE			ml. ue first initial form	ᠦ ····)
（1825				ᠥ	ml. oe first isolate form	ᠥ)

Nominal Character			Free Variation Presentation Form which needs Free Variation Selectors			Collocation of Free Variation Selectors
Code	character	Name	Number	graphic	Name or Use	
			——	ᠥ	* ml. ue second isolate form	ᠥ FVS1
			0010	ᠦ	*ml. ue second initial form	ᠦ FVS1 ····
1824				ᠦ	*ml. ue third initial form	ᠦ FVS2 ····
			（000C	ᠥ	* ml. ue first medial form	····ᠥ ····)
1861			——	ᠥ	* ml. ue second medial form	···· ᠥ FVS1 ····
			0010	ᠦ	* ml. ue third medial form	····ᠦ FVS2 ····
			00B2	ᠥ	* ml. ue fourth medial form	···· ᠥ FVS3 ····
			（000E	ᠥ	* ml. ue first final form	···· ᠥ)
			0011	ᠥ	*ml. ue second final form	···· ᠥ FVS1

Code	character	Name	Number	graphic	Name or Use	Collocation of Free Variation Selectors
			00BD		* ml. ue third final form	···· [FVS2]
			00B3		* ml. ue fourth final form	···· [FVS3]
(1828		MONGOLIAN LETTER			*ml. na ferst initial form	)
			0097		* ml. na second initial form	[FVS1]
			（0005		* ml. na first medial form	···· Mongolian consonants ····)
			（0014		ml. na second medial form	···· Mongolian vowel

	Nominal Character		Free Variation Presentation Form which needs Free Variation Selectors			Collocation of Free Variation Selectors
Code	character	Name	Number	graphic	Name or Use	
			0015		ml. na third medial form	···· [FVS2] ····
			0008		*ml. na fourth medial form	···· [FVS1] ····
			（0008		*ml. na first final form	····)
			0015		*ml. na second final form	···· [FVS1] ····
			00BA		*ml. na third final form	···· [FVS2]
			02E5		*ml. na fourth final form	···· [FVS3]
			02E3		*ml. na fifth final form	····
			00C0		*ml. na sixth final form	····

Code	character	Name	Number	graphic	Name or Use	Collocation
(182D	𐭂	MONGOLIAN LETTER GA			*ml. ga first initial form	𐭂 ····)
182C			——	𐭂	* ml. ga second initial form	𐭂 [FVS1] ····
			(001C	𐭂	ml. ga first medial form	···· 𐭂 Mongolian consonants ····)
			(0006	𐭂	* ml. ga second medial form	···· 𐭂 Mongolian consonants ····)
			(001D	𐭂	ml. ga third medial form	···· [MVS] ····)
			001A	𐭂	*ml. ga fourth medial form	····𐭂 [FVS1] [MVS] ····
			(001E	𐭂	ml. ga medial form	···· 𐭂 ····)
			(001A	𐭂	ml. ga first final form	···· 𐭂)

Nominal Character			Free Variation Presentation Form which needs Free Variation Selectors			Collocation of Free Variation Selectors
Code	character	Name	Number	graphic	Name or Use	
			001D	𐭂	ml. ga second medial f	···· 𐭂 [FVS1]
			(001F	𐭂	ml. ga third final form	···· Feminine vowel 𐭂)
			00B4	𐭂	ml. ga fourth final form	···· 𐭂
			02E5	𐭂	ml. ga fifth final form	···· 𐭂
			00B5	𐭂	ml. ga sixth final form	···· 𐭂

			0035	ꡏ	ml. ga third final form	•••• ꡏ
			1889	ꡏ	*ml.ga femininie isolate with dots	ꡏ [FVS1]
			001B	ꡏ	*ml. ga femininie isolate with dots	ꡏ [FVS2]

ꡏ ꡏ ꡏ ꡏ ꡏ ꡏ ꡏ ꡏ ꡏ ꡏ ꡏ in Mongolian and ꡏ ꡏ ꡏ ꡏ ꡏ ꡏ ꡏ ꡏ ꡏ ꡏ ꡏ ꡏ ꡏ ꡏ ꡏ ꡏ ꡏ ꡏ ꡏ in AliGali script will not appear in Table A in this way and might cause inconvenience in programming. It is suggested to include these letters Table A.

5. 3 levels of implementation

Adding "the 400 years Mongolian" orthography will make the encoding complicated and huge. Therefore, our suggestion is to divide GB/T 26226-2010 and MNS4932:2000 into three levels of basic, fortified and simplified. Should the proposal of 3-level division take into consideration in the new version of UCS?

6. Contents needed to be added:

① There is no "syllable-ending ya wa (debisger)" consonant in UCS 2014, it needs to be added.

ⓐ Change the two variations of 1838 ꡏ wa： medial form " ꡏ "和 final

form " ꡏ " not to use variation selector character; change medial form " ꡏ "

final form " ꡏ " to use variation selector characters.

ⓑ Add final form " ꡏ " to 1836 ꡏ ya.

② Add second final form " ᠠ " to letter u.

Part Two: Modifications

1. Partial adjustments to UCS

①Delete ᢇ18A7 of AliGali script from "Nominal Character set"

②Add rules of using free variation selector character in mandatory ligature.

2. Direction of characters in code table

We agree on the change of transferring direction of Mongolian character from vertical to horizontal in UCS 2014 on condition that Mongolian words will be presented and printed in vertical direction in any word-processing system.

3. On punctuation

Punctuation used in Mongolian script will be changed accordingly to the change of direction of Mongolian script in UCS.

4. On encoding position

As suggested by some software developers, there should be a complete variations table for four positions (independent, initial, medial, final) for each word. We should provide a complete character set of variation forms for four positions.

(light green indicates newly-added variations)

1829	ᠩ	MONGOLIAN LETTER ANG	——	ᠩ	*ml. ang first medial form	ᠨᠵ ᠩ ᠨᠵ
			00C6	ᠩ	ml. ang isolate form	ᠩ

			00C6	ᠿ	ml. ang initial form	ᠿ ᠌
			00C6	ᠿ	*ml. ang second medial form	᠌ᠿ ᠌ ᠌
			0017	᠍	* ml. ang first final form	᠌ ᠿ
1842	᠌	MONGOLIAN LETTER CHI	——	᠌	ml. chi isolate form	᠌
				᠌	ml. chi initial form	᠌
				᠌	*ml. chi medial form	᠌ ᠌ ᠌
				᠌	* ml. chi final form	᠌ ᠌

5. Questions of grouping by scripts

There is an annex in UCS 2014, which includes all presentation forms of Mongolian (includes Todo, Sibe, and Manchu scripts). Although this approach has been improved since the previous version (UCS 2000), it is still very complicated to put four kinds of scripts in one set.

For the purpose of users' convenience and clarity, the revision separates 4 sets (Within Word Presentation Character set, Single Presentation forms, mandatory Mongolian Ligature set and non-mandatory Mongolian Ligature set) of 4 scripts into 16 sets with different contents and scripts. However, the study of Todo, Sibe and Manchu are not yet mature, we submit only 4 tables of " the use rules of

traditional Mongolian free variation selectors " to the ISO/IEC JTC 1/SC 2/WG 2 and the remaining 12 forms of three scripts and the "the international standard of Uighur Mongolian encoding scheme" will be gradually submitted to ISO/IEC JTC 1/SC 2/WG 2.

Part Three: Background information

It has been 17 years since the formulation work of International Standard of Mongolian Character Encoding. The relevant provisions of the international organization for the standardization has been continuously improved with the rapid development of computer technology during this period. At the same time, software developers and customers of all circles also put forward some suggestions for improvements. Therefore, we assumed that it is necessary to modify UCS 2014 after full consultation of standardization department of China and Mongolia and soliciting expert opinions. We held several successive meetings in order to submit a more scientific, acceptable and prospective scheme after listening widely to the views of experts and users in different fields and regions."Traditional Mongolian Character Encoding and Standardization Meeting" hosted by Mongolian Studies of Inner Mongolia University was held in Hohhot from 10 to 12 of August, 2016. Experts and Scholars of 38 people from universities, research institutions, research and development units in Beijing, Shanghai, Shandong and Inner Mongolia attended the seminar.

The development process of Mongolian character is divided into three stages after serious discussion in the meeting: the Middle Mongolian orthography is added to the new version of traditional Mongolian encoding standard; the newly added and revised contents of 2010 Chinese National Standard is identified; "Expansion and Adjustment " such as the rules of independent form, word initial form, mid-word form, word final form and a relatively complete set of

characters of "transformation rules" are finally reached the consensus.

"The 2016 China and Mongolia Mongolian Character Encoding and Standardization Meeting" was held in Beijing from August 31 to September 2, 2016. Over 10 experts from Relevant departments of China such as National Information Technology Standardization Technical Committee, Ethnic Affairs Commission of Inner Mongolia Autonomous Region, Inner Mongolia University, Founder Electronics Co., Ltd., Weifang Beida Jade Bird Imagsetter Co, Inner Mongolia Menksoft Software Co, Ltd and Experts of Mongolia National Academy of Sciences and Mongolia National Bureau of standards attended the seminar.

After the reports, experts from China and Mongolia reached the consensus on the basis of thorough discussion and agreed to regard this scheme as China and Mongolia scheme to be submitted to the ISO/IEC JTC 1/SC 2/WG2.