

**ORGANISATION INTERNATIONALE DE NORMALISATION  
ISO/IEC JTC 1/SC 2/WG 2**

**Universal Multiple-Octet Coded Character Set  
(UCS)**

**ISO/IEC JTC 1/SC 2/WG 2 *N2964***

**2005-08-05**

**Title:** A User's Agreement Related to Phags-pa Script

**Source:** China and Mongolia

**References:**

**Action:** FYI

**Distribution:** ISO/IEC JTC 1/SC 2/WG 2

## A User’s Agreement Related to Phags-pa Script

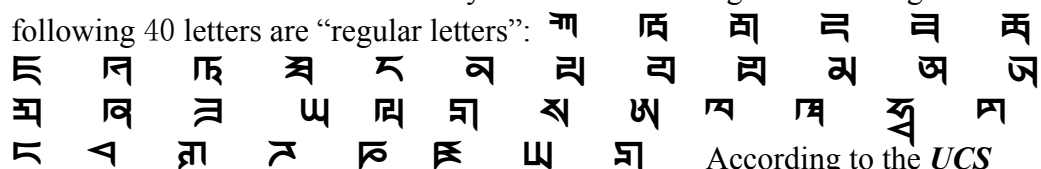

The Mongolian and Chinese experts suggested the necessity of the user’s agreement related to Phags-pa Script in order to unify the technical details which were unable to be included in *UCS* while actually used in the Phags-pa Scripts encoding, at the meeting held in Huhhot, China, in April of 2003. Then, the experts revised the agreement at the second joint meeting on the Phags-pa Script was held in Changsha, China, in October, 2004. In reference to the decision concerning the Phags-pa Script passed at the Phags-pa Ad hoc of WG2#46 held in Xiamen, China in Jan. 2005, the final revision of the agreement was carried out at the meeting on the Phags-pa Script encoding in Ulaanbaatar, Mongolia, in July 2005. Hence, the present agreement.

This Agreement includes: 1). The variant presentation glyphs and the detailed rules for the use of the free variation selector; 2). The input format for the single variant presentation glyphs; 3). The punctuation marks quoted from “Mongolian” in the UCS; 4). The syllable signs and the detailed rules for their use; 5). The alphabet needed to be handled with UNICODE Collation Algorithm; 6). The alphabet needed to be handled with rich text markup; 7). The way to handle the position of CANDARABINDU.


### **1). The variant presentation glyphs and the detailed rules for the use of the free variation selector**


---





Theoretically, each Phags-pa letter has two or more “variant presentation glyphs”. Generally speaking, the majority of consonant letters have fewer variants and differ slightly in their figures. The vowel letters, however, have more variants and differ more clearly in their figures. From such a point of view, the Phags-pa letters can be divided into “regular letters” and “multi-variant letters”.







By regular letters we mean those variant glyphs which do not differ much in their figures and have at most “isolate-syllable-final form” and “syllable-initial and syllable-medial form” (Due to the features of various styles of writing, a letter may have different numbers of variant glyphs in writing), but do not have “free variant”, so there is no need to use “control symbol” letters for regular recording. The following 40 letters are “regular letters”:  According to the *UCS* regulations, it suffices to use the  (U+200D) to record individual variant presentation glyph of such letters.

The multi-variant letters are those letters which differ greatly in their figures. They have more than three “positional variants” and the control symbols are not needed





while they are recorded in a regular way. And the reversed variants refer to those reversed forms which are common glyphs in terms of their figures. The variation selector-1  (U+FE00) is sometimes used while they are recorded .





In Phags-pa Script there are six “reversed variants”. In order to distinguish different “reversed variants” under the same condition the variation selector-1  (U+FE00) in the UCS is used. The detailed rules for the use of the selector are as follows:





① “The variation selector” is not used while the “reversed variants” which are regularly connected in the words or syllables are recorded. The use of “reversed variants ” depends on the preceding character. For example, the regularly-connected  should be recorded as , and also the regularly-connected  should be recorded as  ;





② The “variation selector” shall be also used while the “reversed variants” which are irregularly connected in the words or syllables are recorded. For example, the irregularly-connected  should be recorded as  +  (U+FE00) , and the irregularly-connected  should be recorded as  +  (U+FE00) .


③ The “variation selector” shall be used while the single “reversed variants” are recorded. For example,

  A856 PHAGS-PA LETTER SMALL A, FE00 > =  <sup>1</sup>phags-pa letter  
reversed shaping small a

  < A85C PHAGS-PA LETTER HA, FE00 > =   phags-pa letter  
reversed shaping ha

  < A85E PHAGS-PA LETTER I, FE00 > =   phags-pa letter  
reversed shaping i

  < A85F PHAGS-PA LETTER U, FE00 > =   phags-pa letter  
reversed shaping u

<sup>1</sup> These two forms shall use “variation selector-1” (FE00) . The syllable-initial, word-initial and word medial forms of a character should be differentiated by the position of the character in the given syllables or words. If a single variant is recorded,  ZERO WIDTH JOINER (U+200D) should be used

ཎ VS1 < A860 PHAGS-PA LETTER E, FE00 > = ༎ ༎ phags-pa letter reversed shaping e

ཡ VS1 < A868 PHAGS-PA LETTER SUBJOINED YA, FE00 > = ༡ ༡ phags-pa letter reversed shaping subjoined ya

## 2). The input format for the single variant presentation forms

Generally, the “positional variants” of Phags-pa Script can be automatically designated according to the positions of the variants in the syllables or words, and they do not need to be dealt in a special way. However, those single variant presentation forms which are not in the syllables or words need to be specially handled. Therefore, two control characters are transferred from *General Punctuation* of USC. They are: ZWJZERO WIDTH JOINER (U+200D)and ZWNJZERO WIDTH NON-JOINER(U+200C).

### The way to use ZWJZERO WIDTH JOINER (U+200D)

- ① Isolate form: (letter) e.g. མ → མ; ཎ → ཎ
- ② Syllable-initial form: (letter) ZWJ e.g. མZWJ → མ; ཎ ZWJ → ཎ
- ③ Syllable-medial form: ZWJ (letter) ZWJ e.g. ZWJམZWJ → མ; ZWJཎ ZWJ → ཎ
- ④ Syllable-final form: ZWJ (letter) e.g. ZWJམ → མ; ZWJཎ → ཎ.

If both ZWJ ZERO WIDTH JOINER (U+200D) and VS1 Variation Selector -1(FE00) are needed to be used after the same character, the latter should be put before the first. For example, while ༎ is input, the form should be ZWJཎ VS1 ZWJ. And while inputting ༡, the form should be ZWJཡ VS1 ZWJ.

### The way to use ZWNJZERO WIDTH NON-JOINER (U+200C)

While inputting those glyphs which are in irregular positions, the ZWNJ is used. For example, if the syllable-initial form is input at the end of the syllable, the form should be ཎZWNJཐ → ཎཐ; if ZWNJ is not used, it is the regular form of ཎཐ.


### 3). The punctuation marks quoted from Mongolian in the UCS



Few punctuation marks are used in Phags-pa Script. Based on the existing documents, there are following punctuation marks. And their shapes are the same as those of punctuation marks of Mongolian and Todo. Therefore, the punctuation marks used in Phags-pa Script can be transferred from Mongolian in the USC.

- (1) ◈ (1802)
- (2) ✦ (1803)
- (3) ✧ (1805)
- (4) ◦ (18XX) <sup>2</sup>
- (5) ’ (18XX)
- (6) ˘ (18XX) .

### 4). The syllable signs and the detailed rule of their use

A syllable sign is needed to define the boundaries of words and syllables. A syllable sign indicates the boundaries of syllables in a same word, and the space of a sign has

1/3 of a regular space. It is suggested to use  (U+202F, Narrow No-Break Space)

of UCS. For example,  (Phags-pa word with two syllables) is a word with two syllables;  is a word with three syllables). The way to input it is as

follows: .

### 5). The alphabet needed to be handled with UNICODE Collation Algorithm in A Tie-in way

In Phags-pa Script, there exist some cases that some combined letters are input with two or three nominal characters while they are processed as one letter for searching and sorting. Therefore, these characters need to be clarified so as to handle with UNICODE Collation Algorithm.

---

<sup>2</sup> These three punctuation marks are going to be added to the part of “Mongolian” in the UCS

No.	Glyph	Name	Binding and Condition	
			Character	Code Point
1	འཇ	Phags-pa letter i isolate form	☐ <sup>3</sup> འཇ ☐	≡(A85D)འཇ (A85E)འཇ
2	འཇ	Phags-pa letter i initial form	☐འཇ ☐ ... <sup>4</sup>	≡ (A85D) འཇ (A85E)འཇ
3	འཉ	Phags-pa letter u isolate form	☐འཉ ☐	≡ (A85D) འཉ (A85F) འཉ
4	འཉ	Phags-pa letter u initial form	☐འཉ ☐ ...	≡ (A85D) འཉ (A85F) འཉ
5	འམ	Phags-pa letter e isolate form	☐འམ ☐	≡ (A85D) འམ (A860) འམ
6	འམ	Phags-pa letter e initial form	☐འམ ☐ ...	≡ (A85D) འམ (A860) འམ
7	འམ	Phags-pa letter o isolate form	☐འམ འམ ☐	≡ (A85D) འམ (A861) འམ
8	འམ	Phags-pa letter o initial form	☐འམ འམ ...	≡ (A85D) འམ (A861) འམ
9	འམ	Phags-pa letter i isolate	☐འམ འམ ☐	≡ (A85D) འམ (A867) འམ

<sup>3</sup> ☐ refers to space, numbers and non-Phags-pa script letters.

<sup>4</sup> ... refers to any Phags-pa script glyph.

		form		△
10	ཨ	Phags-pa letter i initial form	□ ཨ △ ...	≡ (A85D) ཨ (A867) △
11	ཨ	Phags-pa letter oe isolate form	□ ཨ ཨ ཨ □	≡ (A85D) ཨ (A866) ཨ (A861) ཨ
12	ཨ	Phags-pa letter oe initial form	□ ཨ ཨ ཨ ཨ ...	≡ (A85D) ཨ (A866) ཨ (A861) ཨ
13	ཨ	Phags-pa letter oe final form	... ཨ ཨ □	≡ (A866) ཨ (A861) ཨ
14	ཨ	Phags-pa letter oe medial form	... ཨ ཨ	≡ (A866) ཨ (A861) ཨ
15	ཨ	Phags-pa letter ue isolate form	□ ཨ ཨ ཨ □	≡ (A85D) ཨ (A866) ཨ (A85F) ཨ
16	ཨ	Phags-pa letter ue initial form	□ ཨ ཨ ཨ ཨ ...	≡ (A85D) ཨ (A866) ཨ (A85F) ཨ
17	ཨ	Phags-pa letter ue final form	... ཨ ཨ □	≡ (A866) ཨ (A85F) ཨ
18	ཨ	Phags-pa letter ue medial form	... ཨ ཨ ...	≡ (A866) ཨ (A85F) ཨ

6). The alphabet needed to be handled with *rich text markup*

Some of the free variants of Phags-pa Scripts need to be indicated by *rich text markup*. They include:

No	Name	Form	Glyph
1	Phags-pa letter wa	second isolate form	𑍕
2	Phags-pa letter wa	second initial form	𑍕
3	Phags-pa letter e	second isolate form	𑍖
4	Phags-pa letter e	second initial form	𑍖
5	Phags-pa letter e	second final form	𑍖
6	Phags-pa letter e	second medial form	𑍖
7	Phags-pa letter e	reversed second final form	𑍖
8	Phags-pa letter e	reversed second medial form	𑍖
9	Phags-pa letter half i	second final form	𑍗
10	Phags-pa letter half i	second medial form	𑍗

## 7). The Position of Candarabindu

It is unique to handle the position of the Phags-pa Script ཨྵ (Candarabindu, U+A8730). It is as follows: 1). At the initial position of the syllable or word while



inputting; 2). At the end of the syllable or word while searching and sorting.

The way to input 𠄎 should be 𠄎𠄎 ; while its sequential arrangement should be 𠄎  
𠄎.