

**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

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 > =   ¹phags-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

¹ 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 VS < A860 PHAGS-PA LETTER E, FE00 > = ༎ ༎ phags-pa letter reversed shaping e

ཡ VS < 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: ZWZERO WIDTH JOINER (U+200D) and ZWZERO WIDTH NON-JOINER(U+200C).

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

- ① Isolate form: (letter) e.g. མ → མ; ཎ → ཎ
- ② Syllable-initial form: (letter) ZW e.g. མZW → མ; ཎ ZW → ཎ
- ③ Syllable-medial form: ZW (letter) ZW e.g. ZWམZW → མ; ZWཎ ZW → ཎ
- ④ Syllable-final form: ZW (letter) e.g. ZWམ → མ; ZWཎ → ཎ.

If both ZW ZERO WIDTH JOINER (U+200D) and VS 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 ZWཎ VS ZW. And while inputting ༡, the form should be ZWཡ VS ZW.

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

While inputting those glyphs which are in irregular positions, the ZW is used. For example, if the syllable-initial form is input at the end of the syllable, the form should be ཎZWཐ → ཎཐ; if ZW 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) ²
- (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.

² 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	☐ ³ འཇ ☐	≡(A85D)འཇ (A85E)འཇ
2	འཇ	Phags-pa letter i initial form	☐འཇ ☐ ... ⁴	≡ (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	འཇ	Phags-pa letter e isolate form	☐འཇ འཇ ☐	≡ (A85D) འཇ (A860) འཇ
6	འཇ	Phags-pa letter e initial form	☐འཇ འཇ ...	≡ (A85D) འཇ (A860) འཇ
7	འཇ	Phags-pa letter o isolate form	☐འཇ འཇ ☐	≡ (A85D) འཇ (A861) འཇ
8	འཇ	Phags-pa letter o initial form	☐འཇ འཇ ...	≡ (A85D) འཇ (A861) འཇ
9	འཇ	Phags-pa letter i isolate	☐འཇ འཇ ☐	≡ (A85D) འཇ (A867) འཇ

³ ☐ refers to space, numbers and non-Phags-pa script letters.

⁴ ... refers to any Phags-pa script glyph.

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

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.