

Title:	Consensus on Encoding Phags-pa Script
Source:	Phags-pa ad hoc, Meeting #46: Chojingzhab (China), Chen Zhuang (China), Hexigeduren (China), Fang Jun (China), Yang Yucheng (China), Lu Qin (China), Sekiguchi (Japan), D. Tumurtogoo (Mongolia), Andrew West (UK), Asmus Freytag (Unicode Consortium), Peter Constable (USA), Ken Whistler (USA)
Doc Type:	Working Group Document
Status:	Ad Hoc Report
Action:	For consideration by JTC1/SC2/WG2
Date:	2005-01-25

The ad hoc committee on Phags-pa met to address PDAM2 ballot comments submitted by China and other outstanding issues, as enumerated in N2912. Consensus was reached on all issues.

The ad hoc committee on Phags-pa recommends that China's PDAM2 ballot comments be disposed as follows:

- T.1** Partially accepted. The four consonant letters requested by China will be added. The two vowel letters will not be added, as they can be represented using sequences of other existing characters.
- T.2** Partially accepted. Contributions from China in N2869, N2870 and N2871 enumerated many variation-selector sequences. Six of these are accepted for addition to Clause 20.4, as required to represent reversed connecting forms of letters. Other variation-selector sequences will not be added.
- T.3** Rejected. The joiner and its variant have been determined not to be necessary to represent the needed connecting forms. Syllable delimitation can be represented using the existing space characters (breaking or non-breaking) already encoded in the UCS.
- T.4** Accepted in principle. China has agreed to provide an appropriate font of the Khubilai Khan style provided it is made available prior to the close of Meeting 46.
- T.5** Partially accepted. The ad hoc determined that only one variation selector is required, and can use one of the variation-selector characters already encoded in the UCS.
- T.6** Partially accepted. The ad hoc reaffirmed the principle that Phags-pa contextually-determined positional variants can be represented in isolation using ZERO WIDTH JOINER and ZERO WIDTH NON-JOINER. The ad hoc determined that variation-selector sequences should be explicitly listed in Clause 20.4. The ad hoc determined that separate listing of free-variant forms was not required. The ad hoc determined that other entries in the Reference Table in N2870 do not need to be listed as the text elements can be represented as sequences of other existing characters.

As a result of these suggested dispositions, the ad hoc does not recommend that any characters be removed from PDAM2. Furthermore, in accordance with these dispositions and other suggestions resulting from the consensus agreement of the ad hoc committee, it is recommended that the contents of PDAM2 be changed as follows:

- Character name change: U+A85A PHAGS-PA LETTER -A should be renamed PHAGS-PA LETTER SMALL A.
- Four characters are to be added, with names as follows:
 - PHAGS-PA LETTER ALTERNATE YA
 - PHAGS-PA LETTER VOICELESS SHA
 - PHAGS-PA LETTER VOICED HA
 - PHAGS-PA LETTER ASPIRATED FA
- The ad hoc agreed that the code table should in principle be re-ordered in accordance with the recommendations of the Chinese NB as specified in N2870, with modifications to reflect the decision not to encode some of the characters listed in N2870, and to include characters in PDAM2. The revised ordering of characters is as specified in the draft revised code charts for amendment 2 in N2924.

In addition to the above changes to the code table, make the following changes to Amendment 2:

- In the specification of new tables, change the specification for the Phags-pa range from A840—A873 to A840—A877.
- In the specification for collection 307 UNICODE 5.0, change the final cell for row A8 from 73 to 77.
- Add the four new Phag-pa letters to the *New names* list.
- Add the following variation-selector-sequence entries as appropriate to existing or new tables in Clause 20.4 *Variation Sequences*:
 - < A856 PHAGS-PA LETTER SMALL A, FE00 > = 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
 - < A860 PHAGS-PA LETTER E, FE00 > = phags-pa letter reversed shaping e
 - < A868 PHAGS-PA LETTER SUBJOINED YA, FE00 > = phags-pa letter reversed shaping subjoined ya

Note: these variation-selector sequences do not select absolute glyph forms; rather, they select a form that is reversed from the normal form predicted by the preceding letter.

In addition to the aforementioned recommended changes to the amendment and recommended disposition of comments, the ad hoc reached the following consensus decisions:

- In reference to issue 6 in N2912, punctuation of Mongolian or Chinese origin, it was agreed that U+02DA was not the appropriate character to represent the small-circle punctuation mark found in Phags-pa texts. Further, it was agreed that the Chinese ideographic full stop, U+3002, may not be appropriate as punctuation mark used in Phags-pa text is centered. It was determined that the rare use of the small-circle

punctuation in Phags-pa does not require separate encoding of an additional punctuation character specific to Phags-pa script. China and Mongolia may consider developing a proposal for encoding a separate centered small-circle punctuation mark for Todo Mongolian, which if encoded could be used for the rare instances of this punctuation used in Phags-pa.

- In reference to issue 8 in N2912, inter-syllable delimitation, it was determined that the user community could establish recommendations or conventions on what characters are to be used to represent the delimiting space between syllables or words. The Unicode Consortium would be willing to incorporate such recommendations from China and Mongolia in the block description for the Phags-pa script to be published in the Unicode Standard.
- In reference to issue 11 in N2912, status of letter A, it was agreed that ISO/IEC 10646 does not determine that characters in the UCS must be given any one linguistic interpretation over another, and that Phags-pa letter A would not be identified as being specifically a vowel or a consonant.
- In reference to issue 12 in N2912, encoding of A-vowel forms, it was agreed that the characters available in PDAM2 are adequate to represent the text elements in question as character sequences. The representation as sequences does not preclude any alternative linguistic analysis of the text elements, and would not prevent software implementations for Phags-pa from processing the text elements as single entities.
- In reference to issue 14 in N2912, encoding of vowels /ö/ and /ü/, it was agreed that the characters available in PDAM2 are adequate to represent the text elements in question as character sequences. The representation as sequences does not preclude any alternative linguistic analysis of the text elements, and would not prevent software implementations for Phags-pa from processing the text elements as single entities.
- In reference to issue 15 in N2912, other glyph variants, it was agreed that these kinds of free-variation glyph variants do not need encoding as variation-selector sequences.
- In reference to issue 16 in N2912, visual versus logical encoding order for CANDRABINDU, it was agreed that the CANDRABINDU would be encoded within a syllable sequence in its visual order, that users would typically enter it in visual order, but that sorting algorithms would typically process it in reading order.
- With respect to the issue of the representation of the Phags-pa letters FA and ASPIRATED FA, there was agreement that, despite the similarity of these characters to sequences < HA, SUBJOINED WA > or < VOICED HA, SUBJOINED WA >, the best resolution was to encode distinct characters FA and ASPIRATED FA.