ISO/IEC JTC 1/SC 2/WG 2 N2745-1
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646[1]
Please fill all the sections A, B and C below.
Please read Principles and Procedures Document (P & P) from http://www.dkuug.dk/JTC1/SC2/WG2/docs/principles.html **for guidelines and details before filling this form.**
Please ensure you are using the latest Form from http://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html.
See also http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html **for latest *Roadmaps*.**

A. Administrative

1. **Title: HPhags-pa script encoding**
2. Requester's name: China National Information Technology Standardization Technical Committee, Mongolian Agency for Standardization and Metrology
3. Requester type (Member body/Liaison/Individual contribution): Member body
4. Submission date: 2004-04-02
5. Requester's reference (if applicable): Users' Agreement Related to HPhags-pa script encoding
6. Choose one of the following:
This is a complete proposal: Yes
or, More information will be provided later:

B. Technical - General

1. Choose one of the following:
a. This proposal is for a new script (set of characters): YES
Proposed name of script: HPhags-pa script
. b. The proposal is for addition of character(s) to an existing block:
Name of the existing block:
2. Number of characters in proposal: 57
3. Proposed category (select one from below - see section 2.2 of P&P document):
A-Contemporar_____ B.1-Specialized (small collection) B.1 B.2-Specialized (large collection)
C-Major extinct _____ D-Attested extinct _____ E-Minor extinct _____
F-Archaic Hieroglyphic or Ideographic _____ G-Obscure or questionable usage symbols _____
4. Proposed Level of Implementation (1, 2 or 3) (see Annex K in P&P document): 2
Is a rationale provided for the choice? NO
If Yes, reference:
5. Is a repertoire including character names provided? YES
a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document? YES
b. Are the character shapes attached in a legible form suitable for review? YES
6. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?
Beijing Founder Electronics Co., LTD
If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used: No.9, 5 Street, Shangdi, Information Industry Base, Haidian District, Beijing 100085, China . vih@founder.com , tel: 86-10 62982266 ext 1466, fax: 86-10 629 81438
7. References:
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided? Users' Agreement Related to HPhags-pa script encoding
b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached? No
8. Special encoding issues:
Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)?
No
9. Additional Information:
Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default

[1] Form number: N2652-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11)

Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at http://www.unicode.org for such information on other scripts. Also see http://www.unicode.org/Public/UNIDATA/UCD.html and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before? No

   If YES explain _____

2. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)? _____

   If YES, with whom? Yes, Chinese and Mongolian Research Institutions

   If YES, available relevant documents: see the attached Bibliography

3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? _____

   Reference: historical and linguistic research work, information technology use, or publishing

4. The context of use for the proposed characters (type of use; common or rare)

   Reference: common in historical and linguistic research work

5. Are the proposed characters in current use by the user community? _____Yes_____

   If YES, where? Reference: linguists

6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? Yes _____

   If YES, is a rationale provided? _____

   If YES, reference: Unicode and ISO/IEC 10646 Roadmaps

7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)? _____

8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? __No_____

   If YES, is a rationale for its inclusion provided? _____

   If YES, reference: _____

9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? No

   If YES, is a rationale for its inclusion provided? _____

   If YES, reference: _____

10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character? __No_____

    If YES, is a rationale for its inclusion provided? _____

    If YES, reference: _____

11. Does the proposal include use of combining characters and/or use of composite sequences? No

    If YES, is a rationale for such use provided? _____

    If YES, reference: _____

    Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? _____

    If YES, reference: _____

12. Does the proposal contain characters with any special properties such as control function or similar semantics? Yes. Included in Proposal.

    If YES, describe in detail (include attachment if necessary) _____

13. Does the proposal contain any Ideographic compatibility character(s)? __No_____

    If YES, is the equivalent corresponding unified ideographic character(s) identified? _____

    If YES, reference: _____

Proposal to Encode the HPhags-pa Script

Ⅰ. A Brief Account of the HPhags-pa[1)] Script

(1) The HPhags-pa script was invented by HPhags-pa by a special edict of Khubilai Khan, Emperor Shi Zu of the Yuan dynasty, and in 1269 the Emperor issued another edict to implement his script across the country. In a few cases, the HPhags-pa script was used as late as the late Qing dynasty. The HPhags-pa script, which was at first called "New Mongol Script", and later called "state characters of Yuan dynasty", is a set of phonetic symbols for multi-lingual writing that can be used "to write and translate all languages". According to material so far discovered, the HPhags-pa script has been used to " write and translate " Mongolian, Chinese (Han), Tibetan, Turkic as well as Sanskrit, and also other languages yet to be deciphered. Needless to say, of the above-mentioned languages, Mongolian as the official language of Yuan dynasty, and Chinese, as the language widely used by so many ethnic groups in China, are the two major languages for which HPhags-pa script has been used. However, the HPhags-pa script is used in different ways to "write/translate" different languages, thus, Mongolian and Chinese (Han) are transcribed according to their sounds, whereas Tibetan and Sanskrit are in principle transliterated letter for letter.

(2) As a special system of letters for "translating and writing" all languages, the HPhags-pa script is characterized by some striking features. The HPhags-pa alphabet is a very rich one, having more letters than enough for writing or translating any language and more often than not having certain compound letters to express a single sound. The HPhags-pa script is written from left to right, and from above downward, which follows the tradition of the Uighur-Mongol script and which is different from Tibetan though HPhags-pa adopted his letters from it, and is also different from Chinese, a major language which HPhags-pa letters are used to translate. The HPhags-pa system uses syllables as its writing unit, which differs from Mongol writing. In Tibetan, a symbol is used to indicate the delimiter of a syllable or a word, whereas the HPhags-pa script does not use any symbol to mark the limit of a syllable or a word. In HPhags-pa writing, the vowel letters are all combinations of a few elements, including isolate form, syllable-initial (or word-initial) form, syllable-medial (or word-medial) form, as well as syllable-final (or word-final) form, and which is quite different from Tibetan writing but is very much like Mongolian writing. In a word, the HPhags-pa writing has its own independent orthographic system, which as a whole differs from Tibetan, Mongolian and Chinese, hence is not a mere "variation" based on these languages though it has had direct or indirect relationship with them in its detail.

(3) So far as we know, extant monuments written in HPhags-pa letters include the following kinds: (a) Inscriptions of the Yuan court like imperial edicts, the empress dowager's decrees, the prince's orders, the imperial master's instructions and various regulations concerning monasteries exempt from taxes and corvee issues by government departments; (b) Religious inscriptions. In all the above-mentioned important monuments and relics, HPhags-pa script is used to "write or translate" various languages; (c) Various gold, silver, copper and iron *pai-tzu* (tallies) for diplomatic envoys or officials bestowed by the emperor or military or administrative organizations; (d) Governmental seals and commanding officers' seals of various levels of the Yuan dynasty provided by *Li Bu* (i.e., Ministry of Education) under the *Zhong Shu* (State Council), paper banknotes and coins distributed by Yuan court as well as great numbers of private seals and signatures; (e) Books like *Menggu Ziyun* (Book of Chinese Rhymes) in HPhags-pa and Chinese (Han) writings and *Baijiaxing* (A Complete Guide to Surnames) remain relatively intact, while *Subhā ṣ itaratnanidhi* is block-printed fragments in Mongolian.

(4) The HPhags-pa script has a square seal script designed specially for transliterating Tibetan characters. Judged by its form of script, this system of glyphs belongs to the HPhags-pa writing system: its forms were invented in imitation of the decorative style of HPhags-pa script; and just like HPhags-pa writing itself, this system of glyphs is also written from left to right, from above downward. Tibetan, however, has its own independent system inasmuch as it differs from HPhags-pa writing both in form of letters and structure of strokes. Specifically, Tibetan has a number of forms of letters which are either altogether wanting in HPhags-pa writing, or not quite the same as

[1)] HPhags-pa was the inventor of the script. His is a Tibetan name and should be transliterated into HPhags-pa on the principle to be close to his name in his native language.

HPhags-pa writing. On the contrary, many forms of HPhags-pa letters are not adopted by this Tibetan system of square seal characters. All its glyphs for writing correspond with the 30 letters and 4 vowel glyphs in Tibetan. However, for lack of enough material on hand and in view of little research about it, in the present Proposal to Encode the HPhags-pa Script, we will not include its characters.

(5) In the past, not so much material of HPhags-pa writing had been discovered, and what's more, its research has been mainly carried out outside China and Mongolia. Since the 1980's, however, emerged an unprecedented upsurge of research for HPhags-pa script in its native land China and Mongolia. In the past 20 years, a great amount of new material has been discovered; e.g., scholars in China and Mongolia have so far discovered more than 40 Mongolian written monuments in HPhags-pa letters instead of about 10 official monuments of the Yuan court before 1980. These discoveries include quite a few stone inscriptions as well as a number of extremely valuable first-hand monuments. Based on deepened research into HPhags-pa writing as recorded in such new material, scholars in China and Mongolia have published highly valuable monographs and treatises, in which they put forward a series of new ideas. Those all-sided and rather abundant materials possessed by Chinese and Mongolian scholars and the latest research results they have achieved, provide us with adequate scientific ground for the encoding of the HPhags-pa script.

The present situation, however, is that on the one hand now and then we discover new texts of HPhags-pa writing, on the other hand our research about the discoveries does not seem sufficient, and there still exist a large number of monuments, e.g., Chinese monuments written in HPhags-pa letters, which require our further research. That's why scholars have reached a common understanding on quite a number of problems concerning HPhags-pa script, but their opinions still vary on certain important problems. Therefore our attitude in formulating the code is to include as much as possible the problems on which opinions do vary. And as soon as those academic problems are solved, we can, according to the then conditions, revise, replenish and improve our HPhags-pa encoding. Major books we have consulted for formulating the present proposal are listed in the attached *Bibliography.*

Ⅱ.Views on How to Handle Certain Problems

As is mentioned above, though the HPhags-pa script is a system based on Tibetan and Sanskrit writings with reference to Mongolian spelling, it is an independent system with its own characteristic features quite different from any other writing system. Consequently, our emphasis ought to focus on these features of the HPhags-pa script itself in preparing our proposal to encode it. The following are roughly our views on how to handle certain problems in its encoding:

(1) Method of Encoding: The majority of letters in HPhags-pa writing have their respective "variant presentation forms", the only difference being that some letters have relatively more variant presentation forms, others have fewer. So, according to relevant regulations in the *UCS*,a method called "nominal glyph" should be used to encode the HPhags-pa script, i.e., at first, we have to divide each letter in HPhags-pa script into "nominal glyph" and "variant presentation glyph", and then to encode each "nominal glyph." At the same time, we have to prepare "a variant presentation glyph set" and "regulations on the use of control symbol".

(2) Nominal Glyph: According to the *UCS* ,the "nominal glyph" is a glyph "to be used to denote, transmit, exchange, handle, input and present the written forms as well as additional signs of various languages". The majority of HPhags-pa letters have a few variants, from which one should be chosen as its "nominal glyph". It is above all the isolate form of a letter which is chosen, or, in the case of there being a few isolate forms, the one which appears the most frequently is to be chosen.

(a) Each of the letters 𗍫 A858, 𗍬 A859, 𗍭 A85A, 𗍮 A85B and 𗍯 A863 has only one isolate form, which is naturally chosen as its nominal glyph.

(b) Letters like 𗍬 A86B and 𗍫 A855 have each two isolate forms 𗍫 𗍬, out of which we choose as nominal glyphs 𗍬 and 𗍫 which have a high frequency of appearance.

(c) The three letters 𗍰 A851, 𗍱 A879 and 𗍲 A87A have no isolate forms, so we adopt forms 𗍰, 𗍱 and 𗍲 which appear the most frequently.

(d) 𗍳 A856, 'OE' and 𗍴 A857 'UE' are two indispensable basic vowels in the Mongolian language. Although their variant presentation glyphs are "compound letters" each consisting of two or three

"elements", they are regarded in the majority of academic monographs (e.g., N. Poppe, 1941,1957; L. Ligeti, 1964, 1972; Č. Šagdarsüreng, 1981, 2001; B. Bulag, 1983; Bao Xiang, 1984; A. Damdinsüreng, 1985, Tulgaguri, 1998; Y. Jančiw, 2002 et al) as compound yet independent letters listed in their respective alphabet. Other scholars who have not included these two letter in alphabet, also say that "sometimes a double letter is used to represent one sound". The forms of the two vowel letters  and  in the HPhags-pa system are based on the forms that represent the two vowel letters in Mongolian writing. The sounds they mark are not the mere combinations of their original glyphs A+E+O and A+E+U, instead what they mark are two other simple sounds ***ö*** and ***ü*** which have nothing to do with them. Therefore, it is necessary to distinguish them as two independent vowel letters instead of anything like "compound vowels" even in terms of information treatment; if not, these two vowel letters which take up 14.8% of the total number of letters, would cause numerous inconveniences in retrieval and sequential arrangement. The same is true if we take into consideration the "automatic transliteration of HPhags-pa text into Latin alphabet".

(e)  is a sign marking a nasal sound. It appears above a relevant syllable in writing, but does not occupy any space in the syllable, and the letter below it still keeps the syllable-initial position as its variant form.

(3) Variant Presentation Glyph: As is regulated in the *UCS*, the "variant presentation glyph" shows, "in presenting certain characters, the graphic symbol of an individual glyph; this symbol depends on the position of the given glyph with relation to other glyphs."

(a) Strictlly speaking, each HPhags-pa letter has several "variant presentation forms". The majority of consonants can be divided into four variants which somewhat differ from each other, viz., isolate form, syllable-initial (or word-initial) form, syllable-medial (or word-medial) form and syllable-final (or word-final) form. For example, the letter A85E has its isolate form , syllable-initial form , syllable-medial form and syllable-final form . A few letters each have less than four variants. Owing to different styles of script, there may appear cases in which several variants have the same form.

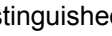
(b) The "variant presentation forms" of the HPhags-pa script are divided into "conditional variants" and "free variants" so far as different conditions of their appearance are concerned.
Conditional variants are those that appear "with relation to the position of other glyphs", which include in turn "positional variant" and "postpositive variant". The four variants of the letter A85E are "positional variants", while the six variants , , , ,  and  which appear only after the five "reversed letters" , , ,  and （） are "postpositive variants" conditioned by their preceding glyphs.
"Free variants" are those of one and the same letter that appear under the same condition.  and , or  and which differ in strokes but mark the same sound, are "free variants".

(c) Certain glyphs mark two sounds in one language, but one sound in another language. The former are two different letters; the latter two free variants of one letter. Such pairs of glyphs as  and , and  and mark one sound in Mongolian, but two sounds in Chinese texts. In our encoding, we'd better handle such glyphs as individual letters.

(d) By the "stroke variants" are meant variants slightly different in their size, thickness, length or sharpness of turn in the stroke. Glyph figures like , , and , which do not mark different sounds, nor any variant of a letter in certain cases, show only slight difference in calligraphy. Such "stroke variants" are not expressed in our encoding system.

(e) A vowel letter has more "positional variants". Apart from its normal isolate, syllable-initial, syllable-medial and syllable-final forms, it also has "postpositive variants" which appear after "reversed letters". One may distinguish a postpositive variant through its preceding letter. Still, there are cases in which normal letter variants (i.e., variants of non reversed letters) appear after "reversed letters", e.g., , which is to be chosen with the "variant selector".

(f) Various "free variants" are to be chosen with the "variant selector"

(g) With the "variant selector", we can distinguish various "free variants" of a certain letter. Thus, the letter  has its isolate-1st syllable-final form , its 1st syllable-initial &-medial form , its 2nd syllable-initial form , its 2nd syllable-medial form , and its 2nd syllable-final form . The "free variants" are distinguished with the "variant selector" practically like this: ; .

(h) For the figures of variant presentation glyphs and the use of the control symbol, please see the regulations in *" A Users' Agreement Related to HPhags-pa Script."*, which is designed for common use by various kinds of users (See the *Supplement*).

(4) Joiner. Since in HPhags-pa script, the syllable is the unit for joining, so the letters in one syllable should be joined together. While the majority of joined parts are on the right side of a letter, there are individual letters whose joined parts are on each letter's left side. The right joined parts after the non reversed letters and the left joined parts after them are the most common joining, without need for special selection, e.g., , , and , , , . But the right joined parts after the "reversed letters" are special joining which requires special selection, e.g., , , . Only when the letter is to be joined with its preceding glyph or with its following glyph, middle part connection is required; however, seeing that it is difficult to standardize the length of the joiner, we can design no middle joiner, and instead, handle it by means of such variants as . To use the joiner will facilitate the connection of various letters, and also enable us to select the joined parts, moreover, it will decrease the number of variant presentation glyphs.

(a) The right joiner (') is on the right side of a glyph and is used between glyphs joined on the right.

(b) The left joiner (') is on the left side of a glyph and is used between glyphs joined on the left side.

For details of the two joiners, see *A Users'Agreement Related to HPhags-pa Script.*.

(5) Control Symbols: The encoding of HPhags-pa script requires several control symbols:

(a) Three "free variant" selectors are needed in order to distinguish different "free variants" of a glyph that appear under the same conditions. The three variant selectors in *UCS* may be used, viz., VS1 (U+FE00), VS2 (U+FE01) and VS3 (U+FE02).

(b) A "syllable delimiter" should be used so as to distinguish the limit of a syllable and that of a word. We may use the narrow non-break space in the *UCS* NNB/SP (U+202F).  (of which is a disyllabic word, and is a trisyllabic word) ought to be recorded like NNB/SPSPNNB/SP NNB/SPSP .

(c) In order to present the positions of certain letters, we choose from the *General Punctuations* of the *UCS* the following two punctuation marks: ZWJ (ZERO WIDTH JOINER, U+200D) and ZWNJ (ZERO WIDTH NON-JOINER, U+200C).

ZWJ (ZERO WIDTH JOINER, U+200D) is used as follows:

1. Isolate form: (letter), e.g.,  → ; →

2. Syllable-initial form: (letter) ZWJ, e.g., → ; →

3. Syllable-medial form: ZWJ(letter) ZWJ, e.g., ZWJ ZWJ →; ZWJ ZWJ →

4. Syllable-final form: ZWJ(letter), e.g., ZWJ →; ZWJ →

ZWNJ ZERO WIDTH NON-JOINER (U+200C) is used as follows:

ZWNJ is used in recording when normal positions are violated. Thus, a syllable-initial form at the end of a syllable can be recorded like: ZWNJ →; or the common when ZWNJ is not used.

(6) Punctuation Marks: Very few punctuation marks are used in HPhags-pa writing. Judging by the monuments and sources now extant, we find the following six marks, whose figures are largely the same as Mongolian punctuation marks, and, having different functions, they have to be given different terms:

(a) HPHAGS-PA ONE DOT (A840)

(b) PHAGA-PA A DOUBLE DOT (A841)

(c) HPHAGS-PA FOUR DOTS (A842)

(d) 。 HPHAGS-PA CIRCLE (A843)

(e) PAGS-PA JOINER (A844)

(f) HPHAGS-PA TIBETAN ANVSVARA （A845）

(7) Signs of Indication:

(a) The HPhags-pa punctuation marks are preceded by the sign "HPHAGS-PA".

(b) HPhags-pa letters used only to record Mongolian or to record Mongolian together with other languages are preceded by "HPHAGS-PA LETTER"; HPhags-pa letters not used to record Mongolian, but are used to record Chinese together with other languages are preceded by "HPHAGS-PA LETTER HAN"; HPhags-pa letters not used to record Mongolian or Chinese, but are used to record Tibetan or Tibetan together with other languages are preceded by "HPHAGS-PA LETTER TIBETAN"; HPhags-pa letters used only to record Sanskrit, but not to record Mongolian, Chinese or Tibetan, are preceded by "HPHAGS-PA LETTER SANSKRIT".

(c) Letters of a given language are spelt in Latin letters close to the pronunciation of that language.

(d) Glyphs which have the same meaning as those in the Mongolian Encoding should adopt as much as possible the signs of indication in Mongolian Encoding, otherwise the glyphs have to use different signs, e.g., ONE DOT, A DOUBLE DOT, FOUR DOTS.

(8) Order of Encoding: The letters of the HPhags-pa alphabet are arranged in quite different orders by various schools; hence, there has never existed, so to speak, a standard and universally acknowledged alphabet for the Hphagsp-ps script.

The HPhags-pa script is a writing system designed for multilingual purposes, just like Latin, Cyrillic, Arabic and Mongolian. So we may well consider the following principles:

(a) Following the example of Mongolian encoding, the vowel letters are to be arranged before the consonant letters.

(b) Consonant letters found in the “original alphabet” are arranged according to their accepted order, while the “newly added consonant letters” are arranged in the order of their similar letters and after the consonant letters of the “original alphabet”.

(c) The punctuation marks are put in the preceding line of letters (A840-A84F), as is the regular practice in the UCS.

(9) Style of Script:

The regular style is used in official monuments, tallies, stone tablets and books of the Yuan dynasty, with texts in Mongolian, Chinese, Tibetan or other languages. This is the most widely used style of the script. The regular style can be subdivided into Khubilai style (1277/1289) , , , and ; Togan style (1345) , , and ; Juyong Pass style (1345) , , and ; wood-engraved block style , , and ; and fine style , , and , etc. Besides, there appear in recent publications other styles of script, such as styles of HPhags-pa script in Soviet and Chinese academic publications.

The decorative style is mainly used in official seals, tablets and tallies of the Yuan dynasty.These are mainly Chinese materials written in decorative style. The HPhags-pa decorative style first appeared in *Menggu Ziyun* (Mongolian Rhymes) revised in 1308. The regular style and the decorative style of the HPhags-pa script are closely dependent upon each other in their figures.

The square style was mainly used in Tibetan Buddhist monasteries or on their seals, and continued to be in use in certain scope as late as modern times. Its figures differ greatly from the regular and the decorative styles.

We think that it is preferable to adopt a style which has been used in monuments written in HPhags-pa letters and at the same is to some extent standardized and looks smooth and beautiful. Hence, we have adopted for our present encoding the very style of HPhags-pa script found in Emperor Khubilai's edicts (1277/1289).

References:

1) Andrew C. West： Response to comments on Phags-pa proposal in N2706， ISO/IEC JTC1/SC2/WG2 N2719，March, 2004.
2) Proposal to encode the Phags-pa script，ISO/IEC JTC1/SC2/WG2 N2622，September, 2003.
3) ISO/IEC JTC1/SC2：Project subdivision proposal for ISO/IEC 10646：2003/Amendment 1，WG2 N2706，October, 2003.
4) Principles on Encoding Phags-pa Script, ISO/IEC JTC1/SC2/WG2 N2666，October, 2003.
5) 1984（ Bao Xiang: A Study on Mongolian Script, Educational Press of Inner Mongolia, 1984)
6) 1983 (B. Bulag: A Brief History of Mongolian Script, People's Press of Inner Mongolia, 1983）
7) Н. Поппе: Квадратная письменность, Издательство академии наук СССР, 1941（ N. Poppe: HPhags-pa Script, Press of U.S.S.R. Academy of Sciences, 1941)
   — The Mongolian monuments in Hphags-pa script, edited by N. Poppe & John R.Krueger, Wiesbaden, 1957.
   — 八思巴字蒙古语碑铭，内蒙古文化出版社，1986 (Tablet Inscriptions in ' Phags-pa-Mongolian Script, Cultural Press of Inner Mongolia, 1986)
8) Marian Lewiciki, Newly Discovered Inscriptions in Phags-pa Script, Wilno, 1937.
9) L. Ligeti：Monuments en écriture 'phags-pa，akadémiai kiadó， Budapest，1973 (L. Ligeti: Monuments in HPhags-pa Script, Budapest: Akademiai Kiado, 1973.)
10) 罗常培、蔡美彪：八思巴字与元代汉语资料汇编，科学出版社，1959（Luo Changpei，Cai Meibiao：The HPhags-pa script and Yuan dynasty Chinese）
11) Ц. Шагдарсүрэн：Монгол үсэгзүй, Улаанбаатар, 1981 (Ts. Šagdarsüreng, Mongol Paleography, Ulan-Bator 1981; Beijing, 1989)
   — Монголчуудын үсэг бичгийн товчоон，Улаанбаатар, 2001 (Ts. Šagdarsüreng, A Study of Mongolian Scripts, 2nd enlarged edition, Ulan-Bator, 2001)
12) 1983， 4 (Šonghor: A Study on the Religious Spells Inscribed on the Stone Walls of Ju Yong Pass)
13) 陶宗儀：书史会要，上海书店，1984 (Tao Zongyi: Shu Shi Hui Yao, Shanghai Publishing- house, 1984.)
14) 1998 (Tulgaguri: A Short History of Mongolian Script, Cultural Press of Inner Mongolia, 1998)
15) Д. Төмөртогоо: Монгол дөрвөлжин үсгийн дурсгалын судалга， Улаанбатар, 2002 (D. Tumurtogoo: Studies on Mongolian Monuments in HPhags-pa Script, Ulan-Bator, 2002)
   — Дөрвөлжин бичгийн мөнгке гэдэг үгийн тухай, SM IV, 1967 (D. Tumurtogoo: Graphical Peculiarity of the Word "Mongke" in HPhags-pa Writing, in: Studia Mongolica, Vol. IV: 1967)

— The Graphical System of Vowel Letters in HPhags-pa Script, in “Theoretical and Historical Problems in Mongolian Linguistics”, Ulan-Bator, 2002, pp. 357-361.

16)А. Дамдин сү р эн：Монгол дө рвө лжин бичиг, Улаанбатар, 1986 (A. Damdinsüreng: HPhags-pa Mongol Writing, 1986)

17) ᠴᠠᠭᠠᠨᠰᠠᠷᠠ：ᠪᠣᠮᠪᠠᠳᠠᠭᠤᠯᠬᠤ ᠠᠭᠤᠳᠠᠯᠲᠠᠢ ᠪᠢᠴᠢᠭ ᠤᠨ ᠠᠭᠤᠷᠢᠮᠳᠠᠭᠢ ᠭᠡᠪ ᠮᠣᠩᠭᠣᠯ ᠬᠡᠯᠡᠨ ᠳᠡᠬᠢ ᠰᠠᠨᠰᠺᠷᠢᠲ ᠦᠭ ᠢᠨᠡᠭᠡᠵᠡ ᠪᠢᠴᠢᠬᠦ ᠵᠢᠨᠬᠡᠨᠡ ᠰᠤᠷᠭᠠᠯᠲᠠ, ᠥᠪᠥᠷ ᠮᠣᠩᠭᠣᠯ ᠤᠨ ᠰᠤᠷᠤᠯᠴᠠᠭ ᠤᠨ ᠡᠷᠳᠡᠮ ᠤᠨ ᠰᠤᠷᠤᠯᠴᠠᠭ ᠤᠨ ᠰᠡᠳᠬᠦᠯ ᠤᠨ ᠰᠤᠷᠭᠤᠭᠤᠯᠢ ᠤᠨ ᠡᠷᠳᠡᠮ ᠤᠨ ᠰᠢᠯᠭᠡᠭᠡᠨ ᠵᠢᠨ ᠬᠡᠪᠯᠡᠯ ᠤᠨ ᠡᠬᠢ ᠪᠢᠴᠢᠭ ᠤᠨ ᠰᠤᠷᠤᠯᠴᠠᠭ

HPHAGS-PA

	A84	A85	A86	A87	A88
0	A940	A850	A860	A870	A880
1	A841	A851	A861	A871	A881
2	A842	A852	A862	A872	A882
3	A843	A853	A863	A873	A883
4	A844	A854	A864	A874	A884
5	A845	A855	A865	A875	A885
6	A846	A856	A866	A876	A886
7	A847	A857	A867	A877	A887
8	A848	A858	A868	A878	A888
9	A849	A859	A869	A879	A889
A	A84A	A85A	A86A	A87A	A88A
B	A84B	A85B	A86B	A87B	A88B
C	A84C	A85C	A86C	A87C	A88C
D	A84D	A85D	A86D	A87D	A88D
E	A84E	A84E	A86E	A87E	A88E
F	A84F	A84F	A86F	A87F	A88F

A840 ᠀ **HPHAGS-PA ONE DOT**

→ 1802 ᠂ mongolian comma

A841 ᠁ **HPHAGS-PA DOUBLE DOT**

→ 1803 ᠃ mongolian full stop

A842 ⁘ **HPHAGS-PA FOUR DOTS**

→ 1803 ⁘ mongolian four dots

A843 ꡀ **HPHAGS-PA CIRCLE**

→ 3002 〇 ideographic full stop

A844 ᠄ **HPHAGS-PA JOINER**

→ 189A ᠊ mongolian nirugu

A845 ꡁ **HPHAGS-P A TIBETAN ANUSVARA**

→ 1880 ᢀ mongolian letter ag anusvara one

→ 0901 ँ devanagari sign candrabindu

→ 0F83 ཿ tibetan sign sna i\ldan

A846 (Thes position shall not be used)
A847 (This position shall not be used)
A848 (This position shall not be used)
A849 (This position shall not be used)
A84A (This position shall not be used)
A84B (This position shall not be used)
A84C (This position shall not be used)
A84D (This position shall not be used)
A84E (This position shall not be used)
A84F (This position shall not be used)

A850 ꡐ **HPHAGS-PA LETTER A**

→ 0F68 ཨ tibetan letter a

A851 ꡑ **HPHAGS-PA LETTER E**

A852 ꡒ **HPHAGS-PA LETTER EE**

≡ 090F ए devanagari letter e

A853 ꡓ **HPHAGS-PA LETTER I**

→ 0F72 ི tibetan vowel sign I

A854 ꡔ **HPHAGS-PA LETTER O**

→ 0F7C ོ tibetan vowel sign o

A855 ꡕ **HPHAGS-PA LETTER U**

→ 0F74 ུ tibetan vowel sign u

A856 ꡖ **HPHAGS-PA LETTER OE**

≡ 0911 ऑ DEVANAGARI LETTER CANDRA O

A857 ꡗ **HPHAGS-PA LETTER UE**

A858 ꡘ **HPHAGS-PA LETTER KA**

→ 0F40 ཀ tibetan letter ka

A859 ꡙ **HPHAGS-PA LETTER KHA**

A85A ꡚ **HPHAGS-PA LETTER GA**

A85B ꡛ **HPHAGS-PA LETTER NGA**

A85C ꡜ **HPHAGS-PA LETTER HAN CA**

→ 0F45 ཅ tibetan letter ca

A85D ꡝ **HPHAGS-PA LETTER CHA**

→ 0F46 ཆ tibetan letter cha

A85E ꡞ **HPHAGS-PA LETTER JA**

→ 0F47 ཇ tibetan letter ja

A85F ꡟ **HPHAGS-PA LETTER HAN NYA**

→ 0F49 ཉ tibetan letter nya

A860 ꡠ **HPHAGS-PA LETTER TA**

A861 ꡡ **HPHAGS-PA LETTER THA**

→ 0F50 ཐ tibetan letter tha

A862 ꡢ **HPHAGS-PA LETTER DA**

→ 0F51 ད tibetan letter da

A863 ꡣ **HPHAGS-PA LETTER NA**

→ 0f53 ན tibetan letter na

A864 ꡤ **HPHAGS-PA LETTER PA**

→ 0F54 པ tibetan letter pa

A865 ꡥ **HPHAGS-PA LETTER HAN PHA**

A866 ꡦ **HPHAGS-PA LETTER BA**

→ 0F56 བ tibetan letter ba

A867 ꡧ **HPHAGS-PA LETTER MA**

→ 0F58 མ tibetan letter ma

A868 ꡨ **HPHAGS-PA LETTER HAN TSA**

→ 091E অ devanagari letter nya

A869 HPHAGS-PA LETTER HAN TSHA

→ 091C ज devanagari letter ja

A86A HPHAGS-PA LETTER HAN DZA

A86B HPHAGS-PA LETTER HAN WA

A86C HPHAGS-PA LETTER HAN ZHA

→ 0F5E ཞ tibetan letter zha

A86D HPHAGS-PA LETTER ZA

→ 0F5F ཟ tibetan letter za

A86E HPHAGS-PA LETTER MINUSCULE A

→ 0F60 ཀ tibetan letter -a

A86F HPHAGS-PA LETTER YA

→ 0F61 ཡ tibetan letter ya

A870 HPHAGS-PA LETTER RA

→ 0F62 ར tibetan letter ra

≡ 0F61 ཡ tibetan letter ya

A871 HPHAGS-PA LETTER LA

→ 0F63 ལ tibetan letter la

A872 HPHAGS-PA LETTER SHA

→ 0F64 ཤ tibetan letter sha

A873 HPHAGS-PA LETTER SA

→ 0F66 ས tibetan letter sa

A874 HPHAGS-PA LETTER HA

→ 0F67 ཧ tibetan letter ha

A875 HPHAGS-PA LETTER QHA

→ 0F41 ཁ tibetan letter kha

A876 HPHAGS-PA LETTER HAN QWA

A877 HPHAGS-PA LETTER HAN FA

A878 HPHAGS-PA LETTER QA

→ 0F42 ག tibetan letter ga

A879 HPHAGS-PA LETTER HAN HALF

→ 18A6 mongolian letter ag half u

→ 0FAD ྭ tibetan subjoined letter wa

A87A HPHAGS-PA LETTER HAN HALF YA

→ 18A7 mongolian letter ag half ya

→ 0FB1 ྱ tibetan subjoined letter ya

A87B HPHAGS-PA LETTER SANSKRIT TTA

A87C HPHAGS-PA LETTER SANSKRIT TTHA

→ 0F4B ཋ tibetan letter l tta

A87D HPHAGS-PA LETTER SANSKRIT DDA

→ 0F4C ཌ tibetan letter dda

A87E HPHAGS-PA LETTER SANSKRIT NNA

→ 0F4E ཎ tibetan letter nna

A87F HPHAGS-PA LETTER HAN YA

A880 HPHAGS-PA LETTER HAN SHA

→ 0F64 ཤ tibetan letter sha

A881 HPHAGS-PA LETTER HAN HA

≡ 0F67 ཧ tibetan letter ha

A882 HPHAGS-PA LETTER HAN FHA

A883 (This position shall not be used)

A884 (This position shall not be used)

A885 (This position shall not be used)

A886 (This position shall not be used)

A887 (This position shall not be used)

A888 (This position shall not be used)

A889 (This position shall not be used)

A88A (This position shall not be used)

A88B (This position shall not be used)

A88C (This position shall not be used)

A88D(This position shall not be used)

A88E(This position shall not be used)

A88F(This position shall not be used)

ISO
International Organization Standardization
Organization Internationale de Normalisation

ISO/IEC JTC1/SC2/WG2
Universal Multiple-Octer Coded Character Set
(UCS)

ISO/IEC JTC1/SC2/WG2 N2745-2
2004-04-02

Title: **Summary of the 1st Joint Meeting on the HPhags-Pa Script Encoding**
Source: **China National Information Technology Standardization Technical Committee**
Mongolian Agency for Standardization & Metrology
Status: **Joint Proposal**
Action: **for discussion at the meeting in Markham, Canada, June, 2004.**
Distribution: for ISO/IEC JTC1/SC2/WG2 experts

I. From March 31 to April 2, 2004 is held in Hohhot, China, the First Joint Meeting for the HPhags-pa Script Encoding, attended by experts from Mongolian Agency for Standardization and Metrology, National University of Mongolia, and Mongolian National University of Education on one side and from China IT Standardization Technical Committee, State Nationalities Affairs Committee, Mongolian Language Working Committee of Inner Mongolia Autonomous Region and Inner Mongolia University on the other.

II. After careful study and discussion of the Documents SC2 N3696, WG2 N2622, WG2 N2719, participants of the present meeting appreciate the arduous efforts Mr. Andrew C. West has made in preparing his proposal to encode the HPhags-pa script and certain results he has achieved, but feel regret that his proposal largely fails to reflect those numerous new academic monographs on HPhags-pa script published mainly in Mongolian by scholars in China and Mongolia, native land of HPhags-pa script in the past 20 years, and what's more, Mr. West completely negates China's preliminary comments (N 2706) on SC2 N3696. Therefore, we consider it necessary to submit our overall opinions, opinions of China and Mongolia, in the form of *Proposal to Encode the HPhags-pa Script.*

III. In order to regulate certain technical details which cannot have been included in the *UCS*, but are useful in the present Encoding of the HPhags-pa Script, at the present Meeting is formulated *A Users'Agreement Related to HPhags-pa Script* Both China and Mongolia will publish it in due time in the form of their respective "national standards". For the moment, we submit the Agreement separately to the ISO/IEC JTC1/SC2 and its WG2.

IV. An important meeting of decisive significance of the ISO/IEC JTC1/SC2/WG2 will be held in Canada in June, 2004, on the HPhags-pa script encoding. There is need for both China and Mongolia to send their experts to attend that meeting.

SIGNATURES

On the Chinese side:

_____, Huchitu, Chairman of the Mongolian Language Working Committee of Inner Mongolia Autonomous Region.

_____, Chen Zhuang, Engineer, China National Information Technology Standardization Technical Committee.

_____, Choijinzhab, Professor in Inner Mongolia University.

_____, Jalsan, Professor in Inner Mongolia University.

_____, Wu Chizhe, Professor in Inner Mongolia University.

_____, Hesige, Department head under Mongolian Language Working Committee of Inner Mongolia.

On the Mongolian side:

_____, Gurragchaa Battsengel, secretary of Standardization Technical Committee for ICT , Mongolian Agency for Standardization & Metrology.

_____, Tsevel Shagdarsuren, Prof., Dr. in Humanistic Sciences, Scientific Secretary of the Centre for Mongol Studies, National University of Mongolia.

_____, Domii Tumurtogoo, Academician, Dr. Professor in Mongolian National University of Education, Ulaanbaatar, Mongolia.

ISO
International Organization Standardization
Organization Internationale de Normalisation

ISO/IEC JTC1/SC2/WG2
Universal Multiple-Octer Coded Character Set
(UCS)

ISO/IEC JTC1/SC2/WG2 N2745-3
2004-04-02

Title: Summary of the User's Agreement Related to HPhags-pa Script
Source: China National Information Technology Standardization Technical Committee., Mongolian Agency for Standardization & Metrology
Status: Joint Proposal
Action: for discussion at the meeting in Markham, Canada, June, 2004.
Distribution: for ISO/IEC JTC1/SC2/WG2 experts

A Users' Agreement Related to HPhags-pa Script

At the China-Mongolia Meeting held in Hohhot, China on March 31 – April 2, 2004, on the Encoding of the HPhags-pa Script, experts of both sides consider it necessary to stipulate *A Users' Agreement Related to HPhags-pa Script* in order to standardize certain technical details in the use of the HPhags-pa encoding which are not included in the *UCS*.

The present Agreement includes: (Ⅰ) General Principle for the Encoding of HPhags-pa Script; (Ⅱ) Nominal Glyphs and Their Names in the HPhags-pa Script; (Ⅲ) Variant Presentation Glyphs and Their Names in HPhags-pa Script; and (Ⅳ) The Reference Table.

Ⅰ.General Principle for the Encoding of HPhags-pa Script

(1) Method of Encoding: The majority of letters in HPhags-pa writing have their respective "variant presentation forms", the only difference being that some letters have relatively more variant presentation forms, others have fewer. So, according to relevant regulations in the *UCS*, a method called "nominal glyph" should be used to encode the HPhags-pa script, i.e.,at first, we have to divide each letter in HPhags-pa script into "nominal glyph" and "variant presentation glyph", and then to encode each "nominal glyph." At the same time, we have to prepare "a variant presentation glyph set" and "regulations on the use of control symbol".

(2) Nominal Glyph: According to the *UCS* ,the "nominal glyph" is a glyph "to be used to denote, transmit, exchange, handle, input and present the written forms as well as additional signs of various languages". The majority of HPhags-pa letters have a few variants, from which one should be chosen as its "nominal glyph". It is above all the isolate form of a letter which is chosen, or, in the case of there being a few isolate forms, the one which appears the most frequently is to be chosen.

(a) Each of the letters  A858, A859, A85A, A85C and A863 has only one isolate form ,which is naturally chosen as its nominal glyph.

(b) Letters like  A863 and A86F have each two isolate forms , , , . Out of which we choose as nominal glyphs and which have a high frequency of appearance.

(c) The three letters A851, A879 and A87A have no isolate forms, so we adopt forms , and which appear the most frequently.

(d) A856, 'OE' and A857 'UE' are two indispensable basic vowels in the Mongolian language. Although their variant presentation glyphs are "compound letters" each consisting of two or three "lexeme ", they are regarded in numerous academic monographs (e.g., N. Poppe, 1941,1957; L. Ligeti, 1964, 1972; Č. Šagdarsüreng, 1981, 2001; B. Bulag, 1983; Bao Xiang, 1984; A. Damdinsüreng, 1985, Tulgaguri, 1998; Y. Jančiw, 2002 et al) as compound yet isolate letters listed in their respective alphabet. Other scholars who have not included these two letters in alphabet, also say that "sometimes a double letter is used to represent one sound". The forms of the two vowel letters and in the HPhags-pa system are based on the forms that represent the two vowel letters in Mongolian writing. The sounds they mark are not the mere combinations of their original glyphs A+E+O and A+E+U, instead what they mark are two other simple sounds ö and u which have nothing to do with them. Therefore, it is necessary to handle them as two independent vowel letters instead of anything like "compound vowels" even in terms of information treatment; if not, these two vowel letters which take up 14.8% of the total number of letters, would cause numerous inconveniences in retrieval and sequential arrangement. The same is true if we take into consideration the "automatic transliteration of HPhags-pa text into Latin alphabet".

(e) is a sign marking a nasal sound. It appears above a relevant syllable in writing, but does not occupy any place in the syllable, and the letter below it still keeps the syllable-initial position as its variant form.

(3) Variant Presentation Glyph: As is regulated in the *UCS*, the "variant presentation glyph" shows, "in presenting certain characters, the graphic symbol of an individual glyph; this symbol depends on the position of the given glyph with relation to other glyphs."

(a) Strictlly speaking, each HPhags-pa letter has several "variant presentation forms". The majority of consonants can be divided into four variants which somewhat differ from each other, viz., isolate form, initial (in syllable or word) form, medial (in syllable or word) form and final (in syllable or word) form. For example, the letter A85E has its isolate form , syllable-initial form , syllable-medial form and syllable-final form . A few letters each have less than four variants. Owing to different styles of script, there may appear cases in which several variants have the same form.

(b) The "variant presentation forms" of the HPhags-pa script are divided into "conditional variants" and "free variants" so far as different conditions of their appearance are concerned.

“Conditional variants” are those that appear "with relation to the position of other

glyphs", which include in turn "positional variant" and "postpositive variant". The four variants of the letter A85E are "positional variants", while the six variants , , , ,  and  which appear only after the five "reversed letters", , ,  and （） are "postpositive variants" conditioned by their preceding glyphs.

“**Free variants**” are those of one and the same letter that appear under the same condition.  and , or  and  which differ in strokes but mark the same sound, are "free variants".

(c) Certain glyphs mark two sounds in one language, but one sound in another language. The former are two different letters; the latter two free variants of one letter. Such pairs of glyphs asand and and mark one sound in Mongolian, but two sounds in Chinese texts. In our encoding, we'd better handle such glyphs as individual letters.

(d) By the "stroke variants" are meant variants slightly different in their size, thickness, length or sharpness of turn in the stroke. Glyph figures like , ,  and , which do not mark different sounds, nor any variant of a letter in certain cases, show only slight difference in calligraphy. Such "stroke variants" are not expressed in our encoding system.

(e) A vowel letter has more "positional variants". Apart from its normal isolate, syllable-initial, syllable-medial and syllable-final forms, it also has "postpositive variants" which appear after "reversed letters". One may distinguish a postpositive variant through its preceding letter. Still, there are cases in which normal letter variants (i.e., variants of non reversed letters) appear after "reversed letters", e.g., , which is to be chosen with the "variant selector".

(f) Various "free variants" are to be chosen with the "variant selector".

(g) With the “variant selector”, we can distinguish various “free variants” of a certain letter. Thus, the letter  has its isolate-1st syllable-final form , its 1st syllable-initial &-medial form , its 2nd syllable-initial form , its 2nd syllable-medial form , and its 2nd syllable-final form . The “free variants” are distinguished with the “variant selector” practically like this: ; .

(4) Joiner. Since in HPhags-pa script, the syllable is the unit for joining, so the letters in one syllable should be joined together. While the majority of joined parts are on the right side of a letter, there are individual letters whose joined parts are on each letter's left side. The right joined parts after the non reversed letters and the left joined parts after them are the most common joining, without need for special selection, e.g., , ,  and , , , . But the right joined parts after the "reversed letters" are special joining which requires special selection, e.g., , , . Only when the letter is to be joined with its preceding glyph or with its following glyph, medial part connection is required; however, seeing that it is difficult to standardize the length of the joiner, we can design no medial joiner, and instead, handle it by means of such variants as . To use the joiner will

facilitate the connection of various letters, and also enable us to select the joined parts, moreover, it will decrease the number of variant presentation glyphs.

(a) The right joiner ( ) is on the right side of a glyph and is used between glyphs joined on the right.

(b) The left joiner ( ) is on the left side of a glyph and is used between glyphs joined on the left side.

(5) Control Symbols: The encoding of HPhags-pa script requires several control symbols:

(a) Three "free variant" selectors are needed in order to distinguish different "free variants" of a glyph that appear under the same conditions. The three variant selectors in *UCS* may be used, viz., VS1 (U+FE00), VS2 (U+FE01) and VS3 (U+FE02).

(b) A "syllable delimiter" should be designed so as to distinguish the limit of a syllable and that of a word. It means that the limit of the syllables in a word ought to have a gap equal to one third of a space.. We may use the narrow non-break space in the *UCS* NNB/SP (U+202F).  (of which is a disyllabic word, and is a trisyllabic word) ought to be recorded like NNB/SP SPNNB/SPNNB/SP SP .

(c) In order to present the positions of certain letters, we choose from the *General Punctuation* of the *UCS* the following two punctuation marks: ZWJ (ZERO WIDTH JOINER, U+200D and ZWNJ (ZERO WIDTH NON-JOINER, U+200C).

ZWJ (ZERO WIDTH JOINER, U+200D) is used as follows:

1. Independent form: (letter) , e.g., → ;  →

2. Syllable-initial form: (letter) ZWJ,e.g., → ;  →

3. Syllable-medial form: ZWJ(letter) ZWJ,e.g., →;  →

4. Syllable-final form: ZWJ(letter), e.g.,  →;  →

ZWNJ ZERO WIDTH NON-JOINER (U+200C) is used as follows:

ZWNJ is used in recording when normal positions are violated. Thus, a syllable-initial form at the end of a syllable can be recorded like: → ; or the common when ZWNJ is not used.

(6) Punctuation Marks: Very few punctuation marks are used in HPhags-pa writing. Their figures are largely the same as Mongolian punctuation marks, and, having different functions, they have to be given different terms:

(a) HPHAGS-pa ONE DOT (A840)

(b)  HPHAGS-PA A DOUBLE DOT（A841）

(c)  HPHAGS-PA FOUR DOTS（A842）

(d) ◦ HPHAGS-PA CIRCLE（A843）

(e) HPHAGS-PA JOINER（A844）

(f)  HPHAGS-PA TIBETAN ANVSVARA（A845）

(7) Signs of Indication:

(a) The HPhags-pa punctuation marks are preceded by the sign "HPHAGS-PA".

(b) HPhags-pa letters used only to record Mongolian or to record Mongolian together with other languages are preceded by "HPHAGS-PA LETTER"; HPhags-pa letters not used to record Mongolian, but are used to record Chinese together with other languages are preceded by "HPHAGS-PA LETTER HAN"; HPhags-pa letters not used to record Mongolian or Chinese, but are used to record Tibetan or Tibetan together with other languages are preceded by "HPHAGS-PA LETTER TIBETAN"; HPhags-pa letters used only to record Sanskrit, but not to record Mongolian, Chinese or Tibetan, are preceded by "HPHAGS-PA LETTER SANSKRIT".

(c) Letters of a given language are spelt in Latin letters close to the pronunciation of that language.

(d) Glyphs which have the same meaning as those in the Mongolian Encoding should adopt as much as possible the signs of indication in Mongolian Encoding, otherwise the glyphs have to use different signs, e.g., ONE DOT, A DOUBLE DOT, FOUR DOTS.

(8) Order of Encoding: The letters of the HPhags-pa alphabet are arranged in quite different orders by various schools; hence, there has never existed, so to speak, a standard and universally acknowledged alphabet for the Hphags-pa script.

The HPhags-pa script is a writing system designed for multilingual purposes, just like Latin, Cyrillic, Arabic and Mongolian. So we may well consider the following principles:

(a) Following the example of Mongolian encoding, the vowel letters are to be arranged before the consonant letters.

(b) Consonant letters found in the “original alphabet” are arranged according to their accepted order, while the “newly added consonant letters” are arranged in the order of their similar letters and after the consonant letters of the “original alphabet”.

(c) The punctuation marks are put in the preceding line of letters (A840-A84F), as is the regular practice in the UCS.

(9) Styles of Script: The letter figures found in HPhags-pa monuments can be classified under three styles, viz., regular style, decorative style and square style.

We think that it is preferable to adopt a style which has been used in monuments written in HPhags-pa letters and at the same time is to some extent standardized and looks smooth and

beautiful. Hence, we have adopted for our present encoding the very style of HPhags-pa script found in Emperor Khubilai's edicts (1277/1289).

Ⅱ Nominal Glyphs and Their Names in the HPags-pa Script

HPHAGS-PA

	A84	A85	A86	A87	A88
0	 A940	 A850	 A860	 A870	 A880
1	 A841	 A851	 A861	 A871	 A881
2	 A842	 A852	 A862	 A872	 A882
3	◦ A843	 A853	 A863	 A873	A883
4	' A844	 A854	 A864	A874	A884
5	 A845	 A855	 A865	 A875	A885
6	A846	 A856	 A866	 A876	A886
7	A847	A857	 A867	A877	A887
8	A848	 A858	 A868	A878	A888
9	A849	 A859	 A869	A879	A889
A	A84A	 A85A	 A86A	A87A	A88A
B	A84B	A85B	A86B	A87B	A88B
C	A84C	A85C	A86C	A87C	A88C
D	A84D	A85D	A86D	A87D	A88D
E	A84E	A84E	A86E	A87E	A88E
F	A84F	A84F	A86F	A87F	A88F

A840 **HPHAGS-PA ONE DOT**

→ 1802 mongolian comma

A841 **HPHAGS-PA DOUBLE DOTS**

→ 1803 mongolian full stop

A842 **HPHAGS-PA FOUR DOTS**

→ 1803 mongolian four dots

A843 **HPHAGS-PA CIRCLE**

→ 3002 ideographic full stop

A844 **HPHAGS-PA JOINER**

→ 189A mongolian nirugu

A845 **HPHAGS-P A TIBETAN ANUSVARA**

→ 1880 mongolian letter ag anusvara one

→ 0901 devanagari sign candrabindu

→ 0F83 tibetan sign sna ldan

A846 **(This position shall not be used)**

A847 **(This position shall not be used)**

A848 **(This position shall not be used)**

A849 **(This position shall not be used)**

A84A **(This position shall not be used)**

A84B **(This position shall not be used)**

A84C **(This position shall not be used)**

A84D **(This position shall not be used)**

A84E **(This position shall not be used)**

A84F **(This position shall not be used)**

A850 **HPHAGS-PA LETTER A**

→ 0F68 tibetan letter a

A851 **HPHAGS-PA LETTER E**

A852 **HPHAGS-PA LETTER EE**

≡ 090F devanagari letter e

A853 **HPHAGS-PA LETTER I**

→ 0F72 tibetan vowel sign I

A854 **HPHAGS-PA LETTER O**

→ 0F7C tibetan vowel sign o

A855 **HPHAGS-PA LETTER U**

→ 0F74 tibetan vowel sign u

A856 **HPHAGS-PA LETTER OE**

≡ 0911 devanagari candra o

A857 **HPHAGS-PA LETTER UE**

A858 **HPHAGS-PA LETTER KA**

→ 0F40 tibetan letter ka

A859 **HPHAGS-PA LETTER GA**

A85A **HPHAGS-PA LETTER KHA**

A85B **HPHAGS-PA LETTER NGA**

A85C **HPHAGS-PA LETTER HAN CA**

→ 0F45 tibetan letter ca

A85D **HPHAGS-PA LETTER CHA**

→ 0F46 tibetan letter cha

A85E **HPHAGS-PA LETTER JA**

→ 0F47 tibetan letter ja

A85F **HPHAGS-PA LETTER HAN NYA**

→ 0F49 tibetan letter nya

A860 **HPHAGS-PA LETTER TA**

A861 **HPHAGS-PA LETTER THA**

→ 0F50 tibetan letter tha

A862 **HPHAGS-PA LETTER DA**

→ 0F51 tibetan letter da

A863 **HPHAGS-PA LETTER NA**

→ 0f53 tibetan letter na

A864 **HPHAGS-PA LETTER BA**

→ 0F56 tibetan letter ba

A865 **HPHAGS-PA LETTER HAN PHA**

A866 **HPHAGS-PA LETTER PA**

→ 0F54 tibetan letter pa

A867 **HPHAGS-PA LETTER MA**

→ 0F58 tibetan letter ma

A868 **HPHAGS-PA LETTER HAN TSA**

→ 091E devanagari letter nya

A869 HPHAGS-PA LETTER HAN TSHA
→ 091C ज devanagari letter ja

A86A HPHAGS-PA LETTER HAN DZA

A86B HPHAGS-PA LETTER HAN WA

A86C HPHAGS-PA LETTER HAN ZHA
→ 0F5E ཞ tibetan letter zha

A86D HPHAGS-PA LETTER ZA
→ 0F5F ཟ tibetan letter za

A86E HPHAGS-PA LETTER MINUSCULE A
→ 0F60 འ tibetan letter -a

A86F HPHAGS-PA LETTER YA
→ 0F61 ཡ tibetan letter ya

A870 HPHAGS-PA LETTER RA
→ 0F62 ར tibetan letter ra

A871 HPHAGS-PA LETTER LA
→ 0F63 ལ tibetan letter la

A872 HPHAGS-PA LETTER SHA
→ 0F64 ཤ tibetan letter sha

A873 HPHAGS-PA LETTER SA
→ 0F66 ས tibetan letter sa

A874 HPHAGS-PA LETTER HA
→ 0F67 ཧ tibetan letter ha

A875 HPHAGS-PA LETTER QHA
→ 0F41 ཁ tibetan letter kha

A876 HPHAGS-PA LETTER HAN QWA

A877 HPHAGS-PA LETTER HAN FA

A878 HPHAGS-PA LETTER QA
→ 0F42 ག tibetan letter ga

A879 HPHAGS-PA LETTER HAN HALF U
→ 18A6 ᢦ mongolian letter ag half u
→ 0FAD ྭ tibetan subjoined letter wa

A87A HPHAGS-PA LETTER HAN HALF YA
→ 18A7 ᢧ mongolian letter ag half ya
→ 0FB1 ྱ tibetan subjoined letter ya

A87B HPHAGS-PA LETTER SANSKRIT TTA

A87C HPHAGS-PA LETTER SANSKRIT TTHA
→ 0F4B ཋ tibetan letter ttha

A87D HPHAGS-PA LETTER SANSKRIT DDA
→ 0F4C ཌ tibetan letter dda

A87E HPHAGS-PA LETTER SANSKRIT NNA
→ 0F4E ཎ tibetan letter nna

A87F HPHAGS-PA LETTER HAN YA
≡ 0F61 ཡ tibetan letter ya

A880 HPHAGS-PA LETTER HAN SHA
→ 0F64 ཤ tibetan letter sha

A881 HPHAGS-PA LETTER HAN HA
≡ 0F67 ཧ tibetan letter ha

A882 HPHAGS-PA LETTER HAN FHA

A883 (This position shall not be used)

A884 (This position shall not be used)

A885 (This position shall not be used)

A886 (This position shall not be used)

A887 (This position shall not be used)

A888 (This position shall not be used)

A889 (This position shall not be used)

A88A (This position shall not be used)

A88B (This position shall not be used)

A88C (This position shall not be used)

A88D(This position shall not be used)

A88E(This position shall not be used)

A88F(This position shall not be used)

Ⅲ Variant Presentation Glyphs and Their Names in HPags-pa Script



0000 ' hpha[1]. left joiner
0001  hphal. a initial form
0002  hphal. e medial form
0003  hphal. ee second isolate form
0004  hphal. ee first initial form
0005  hphal. ee second initial form
0006  hphal. ee third initial form
0007  hphal. ee first medial form
0008 hphal. ee second medial form
0009  hphal. ee first final form
000A  hphal. ee second final form
000B  hphal. ee first reversed final form
000C  hphal. ee second reversed final form
000D  hphal. i initial form
000E  hphal. i medial form
000F hphal. i final form
0010  hphal. i reversed final form
0011  hphal. o first initial form
0012  hphal. o second initial form
0013  hphal. o medial form
0014 hphal. o final form
0015  hphal. u first initial form
0016 hphal. u second initial form
0017  hphal. u third initial form
0018  hphal. u medial form
0019  hphal. u final form
001A  hphal. u reversed final form
001B hphal. oe initial form
001C  hphal. oe first medial form
001D  hphal. oe first final form
001E  hphal. ue initial form
001F  hphal. ue first medial form
0020  hphal. ue first final form
0021  hphal. ka initial-medial form
0022  hphal. kha initial-medial form
0023  hphal. ga initial-medial form
0024 hphal. nga initial-medial form
0025  hphal. ca initial-medial form
0026 hphal. cha initial-medial form
0027 hphal. ja initial-medial form
0028 hphal. nya initial-medial form
0029 hphal. ta first initial-medial form
002A hphal. tha initial-medial form
002B hphal. da initial-medial form
002C hphal. na initial-medial form
002D hphal. pa initial-medial form
002E hphal. pha initial-medial form
002F hphal. ba initial-medial form
0030 hphal. ma initial-media form
0031 hphal. han tsa initial-media form
0032 hphal. han tsha initial-media form
0033 hphal. han dza first initial-medial form
0034 hphal. han wa second isolate-final form
0035 hphal. han wa first initial-medial form
0036 hphal. han wa second initial-medial form
0037 hphal. han nya second initial-medial form
0038 hphal. han za second initial-medial form

0039  hphal. minuscule a initial-medial form

003A  hphal. minuscule a reversed initial-medial form

003B  hphal. ya initial-medial form

003C  hphal. ra first initial-medial form

003D  hphal. ra second initial form

003E  hphal. ra second medial form

003F  hphal. ra second final form

0040  hphal. la initial-medial form

0041  hphal. sha initial-medial form

0042  hphal. sa initial-medial form

0043  hphal. ha initial-medial form

0044  hphal. ha reversed medial form

0045  hphal. ha reversed final form

0046 hphal. qha initial-medial form

0047  hphal. han qwa initial-medial form

0048  hphal. han fa initial-medial form

0049  hphal. qa initial-medial form

004A  hphal. han half u initial form

004B  hphal. han half u medial form

004C  hphal. han half u final form

004D hphal. han half u reversed final form

004E  hphal. han half ya first medial form

004F  hphal. han half ya second medial form

0050  hphal. han half ya second final form

0051  hphal. han half ya reversed final form

0052 hphal. sanskrit tta initial-medial form

0053 hphal. sanakrit ttha first initial-medial form

0054 hphal. sanakrit ttha second initial-medial form

0055 hphal. sanakrit dda initial-medial form

0056 hphal. sanakrit nna initial-medial form

0057 hphal. han ya initial-medial form

0058 hphal. han sha initial-medial form

0059 hphal. han ha initial-medial form

005A hphal. han fha initial-medial form

005B (This position shall not be used)

005C (This position shall not be used)

005D (This position shall not be used)

005E (This position shall not be used)

005F (This position shall not be used)

hpha. = hphags-pa

hphal. = hphags-pa letter

Ⅳ. The Reference Table for HPhags-pa Script

Explanations:

The present Reference Table is prepared in order that our users may use the Control Symbols in one and the same way and are provided with detailed regulations for the recording of individual variant presentation glyph.

Theoretically, each HPhags-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 HPhags-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 38 letters are “regular letters”: ꡀ ꡁ ꡂ ꡃ ꡄ ꡅ ꡆ ꡇ ꡈ ꡉ ꡊ ꡋ ꡌ ꡍ ꡎ ꡏ ꡐ ꡑ ꡒ ꡓ ꡔ ꡕ ꡖ ꡗ ꡘ ꡙ ꡚ ꡛ ꡜ ꡝ ꡞ ꡟ ꡠ ꡡ ꡢ ꡣ ꡤ ꡥ. According to the *UCS* regulations, it suffices to use the [ZWJ] to record individual variant presentation glyph of such letters.

By “multi-variant letters” are meant letters which differ more clearly in their figures and have more than three “positional variants” or have “free variants”, and require the use of the “control symbol” letters to get recorded in a regular way. In the following table such “multi-variant letters” are clearly explained. The table includes two parts: nominal glyphs and variant presentation glyphs. The part of “nominal glyphs” repeats the part of HPhags-pa script in the *UCS*. In the part of “variant presentation glyphs” are found the number of each variant presentation glyph of a letter; figure of each variant presentation glyph and the name of each variant presentation glyph (such names are printed in minuscule so as to distinguish them from the names of the nominal glyphs in the *UCS*); and the method of how to record individual glyph of that variant presentation glyph.

⑴ The number of each variant presentation glyph indicates the order of that glyph in that letter;

⑵ The name in minuscule in the “NAME” column is that of the variant presentation glyph, while in the brackets ( ) are nominal glyphs used as variant presentation glyphs.

⑶ In the column “PRESENTATION RULE” for the variant presentation glyphs is shown how to record individual glyph of the variant presentation glyphs. For regular presentation (writing of words or syllables) we may just remove the zero-break joiner [ZWJ] but retain the variant selectors [VS1], [VS2] and [VS3]. The “variant selectors” in the brackets ( ) [VS1], [VS2] and [VS3] are used only to write individual glyphs and we should remove the “variant selectors” in the brackets during regular presentation.

Reference Table

BASIC CHARACTERS			PRESENTATION CHARACTERS			
№	GRAPHIC SYMBOL	NAME	№	GRAPHIC SYMBOL	NAME	PRESENTATION RULE
A844	'	**HPHA. JOINER**	000		(right joiner)	'
			001	'	hpha. left joiner	' VS1
A852	ನಾ	**HPHAL. EE**	000		(first isolate form)	ನಾ
			001	ನಾ	hphal. ee second isolate form	ನಾ VS1
			002	ನಾ	hphal. ee first initial form	ನಾ ZWJ
			003	ನಾ	hphal. ee second initial form	ನಾ ZWJ VS1
			004	ನಾ	hphal. ee third initial form	ನಾ ZWJ VS2
			005	ನಾ	hphal. ee first medial form	ZWJ ನಾ ZWJ
			006	ನಾ	hphal. ee second medial form	ZWJ ನಾ ZWJ VS1
			007	ನಾ	hphal. ee first final form	ZWJ ನಾ
			008	ನಾ	hphal. ee second final form	ZWJ ನಾ VS1
			009	ನಾ	hphal. ee first reversed final form	ZWJ ನಾ (VS2)
			00A	ನಾ	hphal. ee second reversed final form	ZWJ ನಾ (VS3)
A853	ನಿ	**HPHAL. I**	000		(isolate form)	ನಿ
			001	ನಿ	hphal. i initial form	ನಿ ZWJ
			002	ನಿ	hphal. i medial form	ZWJ ನಿ ZWJ
			003	ನಿ	hphal. I final form	ZWJ ನಿ
			004	ನಿ	hphal. i reversed final form	ZWJ ನಿ (VS1)
A854	ನೊ	**HPHAL. O**	000		(first isolate form)	ನೊ
			001	ನೊ	hphal. o first initial form	ನೊ ZWJ
			002	ನೊ	hphal. o second initial form	ನೊ ZWJ VS1

BASIC CHARACTERS			PRESENTATION CHARACTERS			
№	GRAPHIC SYMBOL	NAME	№	GRAPHIC SYMBOL	NAME	PRESENTATION RULE
			003	৸	hphal. o medial form	ZWJ ৸ ZWJ
			004	৲	hphal. o final form	ZWJ ৲
A855	ভ	**HPHAL. U**	000		(first isolate form)	ভ
			002	ভ	hphal. u first initial form	ভ ZWJ
			003	ভ	hphal. u second initial form	ভ ZWJ VS1
			004	ভ	hphal. u third initial form	ভ ZWJ VS2
			005	ভ	hphal. u medial form	ZWJ ভ ZWJ
			005	ভ	hphal. u final form	ZWJ ভ
			006	ভ	hphal. u reversed final form	ZWJ ভ (VS1)
A856	ভ্ৰ	**HPHAL. OE**	000		(isolate form)	ভ্ৰ
			001	ভ্ৰ	hphal. oe initial form	ভ্ৰ ZWJ
			002	ভ্ৰ	hphal. oe first medial form	ZWJ ভ্ৰ ZWJ
			003	ভ্ৰ	hphal. oe second medial form	ZWJ ভ্ৰ ZWJ VS1
			004	ভ্ৰ	hphal. oe first final form	ZWJ ভ্ৰ
			005	ভ্ৰ	hphal. oe second final form	ZWJ ভ্ৰ VS1
A857	ভ্ৱ	**HPHAL. UE**	000		(isolate form)	ভ্ৱ
			001	ভ্ৱ	hphal. ue initial form	ভ্ৱ ZWJ
			002	ভ্ৱ	hphal. ue first medial form	ZWJ ভ্ৱ ZWJ
			003	ভ্ৱ	hphal. ue second medial form	ZWJ ভ্ৱ ZWJ VS1
			004	ভ্ৱ	hphal. ue first final form	ZWJ ভ্ৱ

BASIC CHARACTERS			PRESENTATION CHARACTERS			
№	GRAPHIC SYMBOL	NAME	№	GRAPHIC SYMBOL	NAME	PRESENTATION RULE
			005		hphal. ue second final form	ZWJ VS1
A86E		**HPHAL. MINUSCUL A**	000		(isolate-final form)	
			001		hphal. minuscul a initial-medial form	ZWJ
			002		hphal. minuscul a reversed initial-medial form	ZWJ (VS1)
A86B		**HPHAL. HAN WA**	000		(first isolate-final form)	
			001		hphal. han wa second isolate-final form	VS1
			002		hphal. han wa first initial-medial form	ZWJ
			003		hphal. han wa second initial-medial form	ZWJ VS1
A870		**HPHAL. RA**	000		(first isolate-final form)	
			001		hphal. ra first initial-medial form	ZWJ
			002		hphal. ra second initial form	ZWJ VS1
			003		hphal. ra second medial form	ZWJ
			004		hphal. ra second final form	ZWJ VS1
A874		**HPHAL. HA**	000		（ isolate-final form）	
			001		hphal. ha initial-medial form	ZWJ
			002		hphal. ha reversed medial form	ZWJ ZWJ (VS1)
			003		hphal. ha reversed final form	ZWJ (VS1)

BASIC CHARACTERS			PRESENTATION CHARACTERS			
№	GRAPHIC SYMBOL	NAME	№	GRAPHIC SYMBOL	NAME	PRESENTATION RULE
A879		**HPHAL. HAN HALF U**	000		(final form)	
			001		hphal. han half u initial form	
			002		hphal. han half u medial form	
			003		hphal. han half u final form	
			004		hphal. han half u reversed final form	(VS1)
A87A		**HPHAL. HAN HALF YA**	000		（first final form）	
			001		hphal. han half ya first medial form	  
			002		hphal. han half ya second medial form	(VS1)
			003		hphal. han half ya second final form	(VS1)
			004		hphal. han half ya reversed final form	(VS2)
A87C		**HPHAL. SANSKRIT TTHA**	000		（isolate-final form）	
			001		hphal. sanskrit ttha first initial-medial form	 
			002		hphal. sanskrit ttha second initial-medial form	VS1