ISO/IEC JTC1/SC2/WG2 NXXXX

L2/04-174

Title	**Comments on Chinese-Mongolian joint proposal to encode the HPhags-pa script (N2745)**
Source	Andrew C. West
Document Type	Individual Contribution
Date	27th May 2004

Contents

1. General Remarks

1.1 Background to this Document

A proposal to encode the Phags-pa script was made in document **N2622**. This proposal was initially accepted by the UTC (see **96-C31**), and subsequently approved for addition to PDAM 1 of ISO/IEC 10646 by WG2 (see resolution **M44.4**).

The Chinese National Body raised eight points concerning this proposal in document **N2706** (Project Subdivision Proposal for ISO/IEC 10646: 2003/Amendment 1). Each of these eight points were individually addressed in detail by myself in document **N2719**, and several amendments to the original proposal were suggested in response to some of the points raised.

China and Mongolia have now put forward a joint counter-proposal for encoding the Phags-pa script (**N2745**). Whilst the character repertoire proposed in N2745 is largely the same as that proposed in N2622, the Chinese-Mongolian proposal presents a radically different encoding model.

The current document provides a critique of the Phags-pa proposal given in N2745.

1.2 In Defence of N2622

My proposal to encode the Phags-pa script (N2622) is based on :

1. A close study of over a hundred years of modern scholarship on the Phags-pa script, from Pozdneyev to Junast.
2. An analytical review of all primary sources relating to the Phags-pa script dating from the 13th and 14th centuries.
3. Extensive reading of Phags-pa texts (including manuscripts, printed texts, monumental inscriptions, coins, seals, etc.) in Chinese, Mongolian, Sanskrit, Tibetan and Uighur.

I am actively involved in the study of the Phags-pa script, for example I recently recognised and translated a hitherto unknown fragment of Phags-pa text held at the British Library that was collected by Auriel Stein from the Tangut city of Kharakoto. The main motivation for making my proposal to encode Phags-pa is to facilitate my research into the script, and to enable me to compile an electronic dictionary of Chinese and Mongolian spelled in the Phags-pa script. It is thus a matter of paramount importance to myself that the Phags-pa script is encoded in the best possible way.

On page 4 of N2745-1 the following statement about the current state of Phags-pa studies is made :

> 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.

I agree entirely that over the last twenty years or so a great deal of new Phags-pa material has been uncovered, and that scholars in China and Mongolia have made very important contributions to the study of the script. My encoding proposal is very much based on such modern scholarship, and in particular follows the opinions of Professor Junast 照那斯图 with respect to a number of key issues with which N2745 is in opposition to. Professor Junast, who is a Fellow of the Chinese Academy of Social Sciences, has probably contributed more to Phags-pa studies over the past twenty-five years than all other Chinese scholars combined, and some of the books and monographs written by him that have been consulted are listed in the bibliography at the end of this document. It is also evident from the bibliography that I have consulted a great deal of material published in China over the past three decades, including all articles on Phags-pa published in academic journals such as Minzu Yuwen 民族语文 [Languages and Scripts of Minority Nationalities] and Wen Wu 文物 [Cultural Relics]. Based on my extensive reading of modern Chinese scholarship, I believe that the encoding model proposed in N2622 does closely reflect the opinion of recognised Chinese experts on the Phags-pa script such as Professor Junast.

In contrast, the encoding model proposed in N2745 appears to be largely based on The Mongolian Monuments in HP'AGS-PA Script (1941, English translation 1957), written by the Russian scholar Nicholas Poppe over sixty years ago, and shows little regard for the advances in Phags-pa scholarship made by Professor Junast and others. In particular the proposed treatment of the null consonant as a vowel and the treatment of the digraphs "eo" and "eu" as discrete letters in N2745 both follow Professor Poppe, even though these ideas have been demonstrated to be erroneous by Professor Junast in a series of influential articles published 1987-1989.

As I am competent in Chinese, Tibetan and Mongolian, I have examined numerous Phags-pa texts written in all three of these languages, and have endeavoured to ensure that my proposal is equally suitable for use with any or all of the languages that can be written in the Phags-pa script (Chinese, Mongolian, Tibetan, Sanskrit and Uighur). On the other hand N2745 is very much biased towards the Mongolian usage of Phags-pa, even though the Phags-pa script was intended for use in writing all languages within the Mongolian Empire, and indeed there are more examples of Chinese usage of the Phags-pa script than there are of Mongolian usage. It attempts to artificially correlate the encoding of the Phags-pa script with Mongolian, despite the fact that there is not a one-to-one correspondence between the spelling of Mongolian words in the two scripts (see N2719 for numerous examples of divergence between the two scripts), and Phags-pa owes little to the Uighur-

derived Mongolian script other than its vertical direction of writing.

For me the big difference between the two proposals is that N2622 proposes a simple and very intuitive encoding model based on a one-to-one relationship between graphic units and encoded characters; whereas the encoding model proposed in N2745 has all the Byzantine complexity of the Mongolian encoding model but with none of the benefits. Furthermore, as the encoding model of N2745 is based on phonetic units rather than graphic units, it is not possible to encode a Phags-pa text without reference to a detailed description of the encoding model or without memorising complex shaping rules. On the other hand, the encoding model proposed in N2622 is designed so that anyone with a basic knowledge of the Phags-pa script could encode a Phags-pa text electronically with reference only to the ISO/IEC 10646 or Unicode code charts.

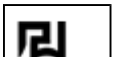
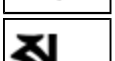
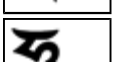
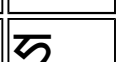
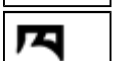
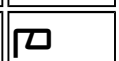
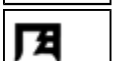
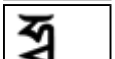
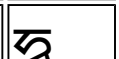
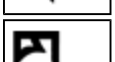
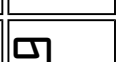
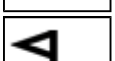
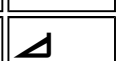
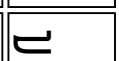
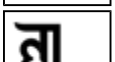
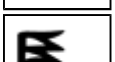
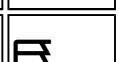
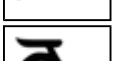
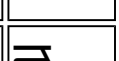
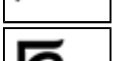
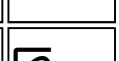
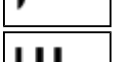
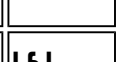
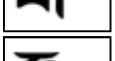
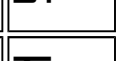
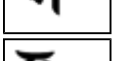
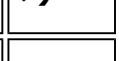
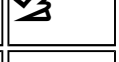
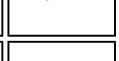
1.3 Comparison between Character Repertoire proposed in N2745 and N2622

The proposed character repertoire for N2745 and N2622/N2719 are summarised in the table below :

Table 1-3-1 : Character Repertoire of N2745 and N2622

N2745			N2622/N2719			Notes
Ref. Glyph	Code Point	Character Name	Ref. Glyph	Code Point	Character Name	
ꡀ	A840	HPHAGS-PA ONE DOT				Clone of 1802
ꡁ	A841	HPHAGS-PA DOUBLE DOT				Clone of 1803
ꡂ	A842	HPHAGS-PA FOUR DOTS				Clone of 1805
ꡃ	A843	HPHAGS-PA CIRCLE				Clone of 3002
ꡄ	A844	HPHAGS-PA JOINER				This is not a character
ꡅ	A845	HPHAGS-PA TIBETAN ANUSVARA	ꠀ	A86F (ABAF)	PHAGS-PA LETTER CANDRABINDU	
ꡐ	A850	HPHAGS-PA LETTER A	ꡐ	A861 (ABA1)	PHAGS-PA LETTER A	
ꡑ	A851	HPHAGS-PA LETTER E	ꡑ	A86E (ABAE)	PHAGS-PA LETTER EE	
ꡒ	A852	HPHAGS-PA LETTER EE	ꡒ	A86C (ABAC)	PHAGS-PA LETTER E	
ꡓ	A853	HPHAGS-PA LETTER I	ꡓ	A86A (ABAA)	PHAGS-PA LETTER I	
ꡔ	A854	HPHAGS-PA LETTER O	ꡔ	A86D (ABAD)	PHAGS-PA LETTER O	
ꡕ	A855	HPHAGS-PA LETTER U	ꡕ	A86B (ABAB)	PHAGS-PA LETTER U	
ꡖ	A856	HPHAGS-PA LETTER OE				This is a precomposed character
ꡗ	A857	HPHAGS-PA LETTER UE				This is a precomposed character

ꡀ	A858	HPHAGS-PA LETTER KA	ꠀ	A840 (AB80)	PHAGS-PA LETTER KA	
ꡁ	A859	HPHAGS-PA LETTER KHA	ꠁ	A841 (AB81)	PHAGS-PA LETTER KHA	
ꡂ	A85A	HPHAGS-PA LETTER GA	ꠂ	A842 (AB82)	PHAGS-PA LETTER GA	
ꡃ	A85B	HPHAGS-PA LETTER NGA	ꠃ	A843 (AB83)	PHAGS-PA LETTER NGA	
ꡄ	A85C	HPHAGS-PA LETTER HAN CA	ꠄ	A844 (AB84)	PHAGS-PA LETTER CA	
ꡅ	A85D	HPHAGS-PA LETTER CHA	ꠅ	A845 (AB85)	PHAGS-PA LETTER CHA	
ꡆ	A85E	HPHAGS-PA LETTER JA	꠆	A846 (AB86)	PHAGS-PA LETTER JA	
ꡇ	A85F	HPHAGS-PA LETTER HAN NYA	ꠇ	A847 (AB87)	PHAGS-PA LETTER NYA	
ꡈ	A860	HPHAGS-PA LETTER TA	ꠈ	A84C (AB8C)	PHAGS-PA LETTER TA	
ꡉ	A861	HPHAGS-PA LETTER THA	ꠉ	A84D (AB8D)	PHAGS-PA LETTER THA	
ꡊ	A862	HPHAGS-PA LETTER DA	ꠊ	A84E (AB8E)	PHAGS-PA LETTER DA	
ꡋ	A863	HPHAGS-PA LETTER NA	ꠋ	A84F (AB8F)	PHAGS-PA LETTER NA	
ꡌ	A864	HPHAGS-PA LETTER PA	ꠌ	A850 (AB90)	PHAGS-PA LETTER PA	
ꡍ	A865	HPHAGS-PA LETTER HAN PHA	ꠍ	A851 (AB91)	PHAGS-PA LETTER PHA	
ꡎ	A866	HPHAGS-PA LETTER BA	ꠎ	A852 (AB92)	PHAGS-PA LETTER BA	
ꡏ	A867	HPHAGS-PA LETTER MA	ꠏ	A853 (AB93)	PHAGS-PA LETTER MA	
ꡐ	A868	HPHAGS-PA LETTER HAN TSA	ꠐ	A854 (AB94)	PHAGS-PA LETTER TSA	
ꡑ	A869	HPHAGS-PA LETTER HAN TSHA	ꠑ	A855 (AB95)	PHAGS-PA LETTER TSHA	
ꡒ	A86A	HPHAGS-PA LETTER HAN DZA	ꠒ	A856 (AB96)	PHAGS-PA LETTER DZA	
ꡓ	A86B	HPHAGS-PA LETTER HAN WA	ꠓ	A857 (AB97)	PHAGS-PA LETTER WA	
ꡔ	A86C	HPHAGS-PA LETTER HAN ZHA	ꠔ	A858 (AB98)	PHAGS-PA LETTER ZHA	
ꡕ	A86D	HPHAGS-PA LETTER ZA	ꠕ	A859 (AB99)	PHAGS-PA LETTER ZA	
ꡖ	A86E	HPHAGS-PA LETTER MINUSCULE A	ꠖ	A85A (AB9A)	PHAGS-PA LETTER -A	
ꡗ	A86F	HPHAGS-PA LETTER YA	ꠗ	A85B (AB9B)	PHAGS-PA LETTER YA	

	A870	HPHAGS-PA LETTER RA		A85C (AB9C)	PHAGS-PA LETTER RA	
	A871	HPHAGS-PA LETTER LA		A85D (AB9D)	PHAGS-PA LETTER LA	
	A872	HPHAGS-PA LETTER SHA		A85E (AB9E)	PHAGS-PA LETTER SHA	
	A873	HPHAGS-PA LETTER SA		A85F (AB9F)	PHAGS-PA LETTER SA	
	A874	HPHAGS-PA LETTER HA		A860 (ABA0)	PHAGS-PA LETTER HA	
	A875	HPHAGS-PA LETTER QHA		A862 (ABA2)	PHAGS-PA LETTER QA	
	A876	HPHAGS-PA LETTER HAN QWA		A863 (ABA3)	PHAGS-PA LETTER XA	
	A877	HPHAGS-PA LETTER HAN FA		A864 (ABA4)	PHAGS-PA LETTER FA	
	A878	HPHAGS-PA LETTER QA		A865 (ABA5)	PHAGS-PA LETTER GGA	
	A879	HPHAGS-PA LETTER HAN HALF U		A866 (ABA6)	PHAGS-PA SUBJOINED LETTER WA	
	A87A	HPHAGS-PA LETTER HAN HALF YA		A867 (ABA7)	PHAGS-PA SUBJOINED LETTER YA	
	A87B	HPHAGS-PA LETTER SANSKRIT TTA		A848 (AB88)	PHAGS-PA LETTER TTA	
	A87C	HPHAGS-PA LETTER SANSKRIT TTHA		A849 (AB89)	PHAGS-PA LETTER TTHA	
	A87D	HPHAGS-PA LETTER SANSKRIT DDA		A84A (AB8A)	PHAGS-PA LETTER DDA	
	A87E	HPHAGS-PA LETTER SANSKRIT NNA		A84B (AB8B)	PHAGS-PA LETTER NNA	
	A87F	HPHAGS-PA LETTER HAN YA				N2622 standardized variant <A85B, FE00>
	A880	HPHAGS-PA LETTER HAN SHA				N2622 standardized variant <A85E, FE00>
	A881	HPHAGS-PA LETTER HAN HA				N2622 standardized variant <A860, FE00>
	A8802	HPHAGS-PA LETTER HAN FHA				N2622 standardized variant <A864, FE00>
				A868 (ABA8)	PHAGS-PA SUBJOINED LETTER RA	N2745 represents this as a variant of the letter RA
				A869 (ABA9)	PHAGS-PA SUPERFIXED LETTER RA	N2745 represents this as a variant of the letter RA
				A870 (ABB0)	PHAGS-PA SINGLE HEAD MARK	
				A871 (ABB1)	PHAGS-PA DOUBLE HEAD MARK	
				A872 (ABB2)	PHAGS-PA MARK SHAD	

			=	A873 (ABB3)	PHAGS-PA MARK DOUBLE SHAD	

N.B. In this document **E** corresponds to N2622/N2719 **A86C** [PHAGS-PA LETTER E] and N2745 **A852** [HPHAGS-PA LETTER EE]; and **Ė** corresponds to N2622/N2719 **A86E** [PHAGS-PA LETTER EE] and N2745 **A851** [HPHAGS-PA LETTER E].

The character repertoire proposed in N2745 is largely the same as that proposed in N2622. Both proposals include the forty-one basic Phags-pa letters recorded in the earliest descriptions of the script, as well as the four additional Sanskrit-usage letters and anusvara/candrabindu sign found in the Juyong Guan inscriptions. The differences in character repertoire between the two proposals are :

- N2745 represents the subjoined letter RA found in the Juyong Guan inscriptions as a variant form of the letter **RA**.
- N2745 represents the superfixed letter RA used for Tibetan as a variant form of the letter **RA**.
- N2745 does not encode any of the head marks and shad marks used in Tibetan Phags-pa texts.
- N2622 does not encode the precomposed letters **OE** and **UE**.
- N2622 does not encode the "joiner" character.
- N2622 does not encode the clones of Mongolian and Chinese punctuation marks.
- The four standardized variants proposed in N2622 are proposed as distinct characters in N2745.

I do not believe that any of the characters proposed in N2745 but omitted in N2622 are necessary or should in fact be encoded. On the other hand, the Tibetan punctuation marks omitted in N2745 are essential for representing Tibetan Phags-pa texts, and the treatment of the subjoined letter RA as a variant form of the letter **RA** is undesirable. The treatment of the four standardized variants proposed on N2622 as distinct characters is also quite problematic. All of these points are elaborated on later in this document.

There is no significant difference between the basic glyph shapes of the reference glyphs provided in the two proposals. The only difference is that the Phags-pa font for N2745 is based on the style of Phags-pa lettering found on monumental inscriptions, whereas the Phags-pa font for N2622 is based on the style of Phags-pa lettering used in printed and manuscript texts. I would say that both fonts styles are acceptable. However, as the letters HPHAGS-PA LETTER HAN YA/SHA/HA/FHA [A87F..A882] proposed in N2745 only occur in the single text Menggu Ziyun, where they are written in a thin line style (see Illustration 2-1-3), the thick inscriptional style reference glyphs for these character given in N2745 does not make for easy identification of the characters.

1.4 Fundamental Flaws with the N2745 Encoding Model

Whilst the differences in character repertoire between the two proposals could perhaps be resolved, the differences in the two encoding models are irreconcilable. I believe that the encoding model proposed in N2745 is fundamentally flawed, and would be completely unworkable in practise. This is because the N2745 encoding model allows for multiple encoding sequences for the representation of certain Phags-pa syllables, and due to the peculiar nature of the encoding model it is logically impossible to define compatibility relationships between any of these sequences. This situation is best explained by means of some examples.

Example 1

This example taken from an imperial edict dated the "year of the dragon" (1328?) shows the Phags-pa spelling of the Mongolian word for "not". Poppe *transliterates* this word as **'eu-lu** and *transcribes* it as **ülu**. In modern scholarship it is normally transcribed as **'ülu** (see Illustration 2-4-2 and Table 2-4-2). This corresponds to the word spelled **ülü** ᠦᠯᠦ in the Mongolian script.

According to my understanding of N2745 this word could be encoded using any one of the following character sequences (in

fact, if we take into account the proposed HPHAGS-PA JOINER [A844] there are many more possible character sequences, but to simplify matters we will ignore the "joiner" character for now) :

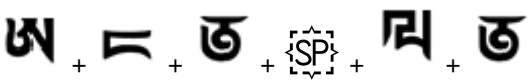
A. <A857, 202F, A871, A857, FE00> = 

B. <A857, 202F, A871, A855> = 

C. <A850, A857, 202F, A871, A857, FE00> = 

D. <A850, A857, 202F, A871, A855> = 

E. <A850, A851, A855, 202F, A871, A857, FE00> = 

F. <A850, A851, A855, 202F, A871, A855> = 

Although the authors of N2745 might claim that only **A** is the correct encoding sequence for this Phags-pa word, all six sequences should produce exactly the same rendered output according to the N2745 encoding model, whilst at the same time it would be impossible to define compatibility mappings between any two sequences. It is worthwhile explaining why there so many different ways of encoding a single word such as this, and why the sequences cannot be defined as canonically equivalent.

A and B

The initial **'ü** of **'ülu** consists of a null consonant followed by the letter **Ė** and the letter **U**. This sequence of three letters is represented as the precomposed character HPHAGS-PA LETTER UE [A857] in N2745.

C and D

Whilst the reference glyph for HPHAGS-PA LETTER UE [A857] comprises three glyph components corresponding to the letters HPHAGS-PA LETTER A [A850], HPAGS-PA LETTER E [A851] and HPHAGS-PA LETTER U [A855], according to N2745 HPHAGS-PA LETTER UE should only be rendered with the initial null consonant glyph component in initial positions within a syllable, and should be rendered like the sequence HPAGS-PA LETTER E plus HPHAGS-PA LETTER U in medial or final positions within a syllable. Thus if HPHAGS-PA LETTER UE is preceded by HPHAGS-PA LETTER A (i.e. the null consonant), as is the case in **C** and **D**, then it loses the null consonant glyph component, and we have the surreal situation that <A850, A857> is identical to A857 by itself. As you obviously cannot decompose A857 to <A850, A857>, and as A857 only corresponds to <A850, A857> initially, it impossible to define a compatibility mapping between A857 and <A850, A857>.

E and F

As A857 is a precomposed character, it could also be encoded using the decomposed sequence of HPHAGS-PA LETTER A [A850] plus HPAGS-PA LETTER E [A851] plus HPHAGS-PA LETTER U [A855], as is the case in **E** and **F**. This would be the logical encoding for anyone following Poppe's system of transliteration for Mongolian Phags-pa (Poppe transliterates this sequence as **'eu**), and corresponds to the character sequence that would be required by the N2622 encoding model. Again, as the precomposed character HPHAGS-PA LETTER UE cannot be decomposed, it would be impossible to canonically equate this sequence with either A857 or <A850, A857>.

A/C/E versus B/D/F

According to my understanding of N2745 (and the document is somewhat ambiguous on this point), as the Phags-pa word **'ülu** corresponds to**ülü** in the Mongolian script, the final **u** of **'ülu** should be considered to be a variant of the character HPHAGS-PA LETTER UE, selected by application of VS-1 (as is the case in **A**, **C** and **E**). However, as Poppe, Junast and all other respected Phags-pa experts transliterate the final letter of **'ülu** as **u** rather than **ü** (see Illustration 2-4-2), there is good reason to suppose that most users would encode the letter using HPHAGS-PA LETTER U rather than HPHAGS-PA LETTER UE plus VS-1 (as is the case in **B**, **D** and **F**), especially as the unintuitive variation sequence could not be guessed from the code charts alone.

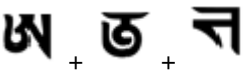
Note that in N2622 the word **'ülu** is unambiguously encoded by the single character sequence <A861, A86E, A86B, 202F, A85D, A86B> (though of course NNBSP [202F] is only the suggested space character to use between syllables within a word -- there is no reason why any space character should not be appropriate).

Example 2

This example taken from the Phags-pa version of the "Hundred Chinese Family Names" is the Phags-pa spelling of the Chinese ideograph 危 wēi, which is normally transliterated as **'uė** or **'ue**.

This Phags-pa syllable could be encoded using either of the following character sequences :

A. <A855, FE00, A852> = 

B. <A850, A855, A852> = 

The preferred encoding according to N2745 is to use a variation selector (i.e. sequence **A**). The Phags-pa letter **U** does not need to be preceded by a null consonant [A850] initially when it is a pure vowel, but when it is part of a diphthong, as is the case here, it is always preceded by a null consonant. The sequence of null consonant plus letter **U** is considered as variant of the initial form of the letter **U** in N2745, and is selected by means of VS-1. From a Chinese perspective this is ridiculous, as there is nothing for it to be a variant of – <A855, A852> (i.e. initial form of **U** followed by the letter **E**) does not occur in Chinese. Moreover, the treatment of the null consonant as part of the initial form of the letter **U** runs completely contrary to scholarly practise of transliterating the null consonant with an apostrophe.

As I do not think that two multi-character sequences can be canonically equivalent, <A855, FE00> cannot have a compatibility mapping to <A850, A855>, and so the two sequences cannot be considered as equivalent even though they should produce exactly the same rendered output. Given the unintuitive nature of the preferred encoding sequence (using a variation sequence to select a standard sequence of letters) it may be presumed than many users will represent Chinese Phags-pa texts using the second encoding sequence.

Note that in N2622 the word **'ue** is unambiguously encoded by the single character sequence <A861, A86B, A86C>.

Example 3

This example, taken from a modern book on Tibetan calligraphy, shows the syllable **'om**, corresponding to the Tibetan oṁ ཨོཾ (in modern Tibetan Phags-pa the anusvara is represented by the letter **MA**).

This Phags-pa syllable could be encoded using either of the following character sequences :

A. <A854, FE00, A867> = 

B. <A850, A854, A867> = 

Unlike Mongolian and Chinese Phags-pa usage, Tibetan Phags-pa usage requires the null consonant before all vowels initially. Compare this Tibetan spelling of oṁ (with initial null consonant) with the spelling **om** (initial form of letter **O** followed by the letter **MA**) found in the Mongolian Phags-pa inscriptions at Juyong Guan.

As with Example 2, the null consonant before a vowel letter is not recognised as a separate character according to the encoding model proposed in N2745. Instead, this sequence of two letters is encoded as a variation of the vowel letter, i.e. HPHAGS-PA LETTER O [A854] plus VS-1 (sequence **A**). Again, from a Tibetan perspective this makes no sense whatsoever, as in the Tibetan script the null consonant is always a distinct character. It is highly unlikely that any Tibetan would ever encode a syllable such as this by means of vowel plus variation selector following the N2745 encoding model; but would almost certainly encode the syllable as <A850, A854, A867>, following the Tibetan encoding model. Furthermore, the use of a variation selector for syllables like this simply puts an unnecessary obstacle in the way of transcoding between the Tibetan and Phags-pa scripts.

A point that is worthwhile making is that if the authors of N2745 claim that encoding a null consonant plus a vowel (or the semi-vowel HPHAGS-PA LETTER HAN HALF [A879] in Chinese Phags-pa) instead of vowel (or semi-vowel) plus variation selector should be considered illegal (and I believe that this would be an untenable position), then the Tibetan Phags-pa syllable **'i** (see N2622 Example 5 top of right column for an example) would be unrepresentable as N2745 does not define a "null consonant plus i" variation of HPHAGS-PA LETTER I [A853]. On the other hand, as null consonant plus the letter **I** must of necessity be encoded as <A850, A853>, then why should other sequences of null consonant plus a vowel letter not also be encoded in this manner rather than by means of an unintuitive variation selector ?

From these examples I think that it is evident that not only are there multiple possible representations for the same Phags-pa syllables, but that there is also a very high likelihood that some users will not encode Phags-pa syllables according to the model proposed in N2745, whether deliberately or from ignorance (and it should be remembered that the only guidance that many users will have are the ISO/IEC 10646 or Unicode code charts). This will mean that the same Phags-pa text may be represented by multiple character sequences, and as these character sequences cannot be defined as canonically equivalent (due to the peculiarity of the encoding model, as explained above), this will cause complete chaos for data processing operations such as searching and collation.

Unfortunately, the scope for confusion and multiple encodings of the same Phags-pa syllable is not restricted to the problem of the null consonant, but is further compounded by the proposed HPHAGS-PA JOINER [A844]. A single example should suffice to demonstrate the confusion that encoding such a character would engender.

Example 4

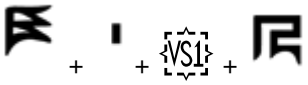
These are three variant representations of the Phags-pa syllable **tth-a** (i.e. HPHAGS-PA LETTER SANSKRIT TTHA [A87C] followed by HPHAGS-PA LETTER MINUSCULE A (**-A**) [A86E]) taken from the Buddhist inscriptions on the east wall of Juyong Guan), and all represent the Sanskrit syllable **ṭhā**. Normally Phags-pa letters within the same syllable are ligated together along the right-hand side, but after the reversed Sanskrit-usage letters **TTA**, **TTHA**, **DDA** and **NNA** Phags-pa letters

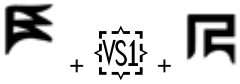
are usually reversed, and the ligature is made on the left-hand side. However, as a reversed letter **-A** is identical to the letter **SHA**, it normally does not reverse, and the ligature may be made on the left or right. Out of the six examples of **TTHA plus -A** in the Juyong Guan inscriptions, four have unreversed **-A** with ligature on the left (above left), one has unreversed **-A** with ligature on the right (above middle), and one has reversed **-A** with ligature on the left (above right). These are fundamentally the same character sequence, and all should be matched in a search for <A87C, A86E>.

N2745 is contradictory about how the ligature between adjacent Phags-pa letters should be achieved. It seems to suggest that the physical extender that joins letters together should be hard-coded using the proposed HPHAGS-PA JOINER [A844] character; but it also seems to suggest that "letter plus joiner" combinations should be considered as presentation forms for the particular letter in initial and medial positions within a syllable, in which case the presentation form with the "joiner" would be selected automatically by the rendering system without any need for the user to hard-code A844 in the character sequence. The implication is perhaps that users with basic rendering systems would be expected to hard-code A844 in the character stream in order to manually produce the ligature between adjacent letters; whereas users with sophisticated rendering systems would let the rendering system select the appropriate presentation form without any need for user intervention. The upshot is that almost any Phags-pa syllable could be encoded either with or without the insertion of A844 in the character stream.

As, according to N2745, the application of VS-1 can be used to reverse the default orientation of both HPHAGS-PA JOINER [A844] and HPHAGS-PA LETTER MINUSCULE A (**-A**) [A86E], and also to change the side that the initial form of HPHAGS-PA LETTER SANSKRIT TTHA [A87C] ligates on, it is possible to encode the three variant forms of the syllable **tth-a** shown above in various ways, as shown below :

A. <A87C, A86E> = 

B. <A87C, A844, FE00, A86E> = 

C. <A87C, FE00, A86E> = 

D. <A87C, A844, A86E> = 

E. <A87C, A86E, FE00> = 

F. <A87C, A844, FE00, A86E, FE00> = 

A or **B** represent possible character sequences for the left-hand example of **tth-a**.

C or **D** represent possible character sequences for the middle example of **tth-a**.

E or **F** represent possible character sequences for the right-hand example of **tth-a**.

The variation selector VS-1 could be stripped out or ignored by processes if so required, but even so there are still two possible encodings of **tth-a**, with a "joiner" character (<A87C, A844, A86E>) and without (<A87C, A86E>). Whilst this example is more complex than the typical case, the ambiguity over whether to encode a "joiner" character or not is common to all sequences of two adjacent Phags-pa letters other than those involving the letter **O** (which ligates down the middle).

1.5 Summary of Specific Comments on N2745

The specific comments on N2745 are very detailed, and so for convenience a summary of these comments is provided.

1. Code Point Order

The order of characters proposed in N2745 is generally acceptable, with the exception of the proposed "vowel" characters A850..A857, which artificially follow the Mongolian vowel order. A850 is not a vowel at all, but a "null consonant", and should be placed after A874 as the thirtieth consonant letter. The order of vowel letters should follow the Tibetan vowel order as given in all 14th-century descriptions of the Phags-pa script, that is A853, A855, A852, A854, A851 (A856 and A857 should not be encoded, as discussed in Section 2-3).

2. HPHAGS-PA LETTER A [A850]

In N2745 the null consonant initial, HPHAGS-PA LETTER A [A850], is not directly encoded when it occurs before a vowel or semi-vowel letter, but is taken to be part of the initial-position glyph form of the letter it precedes. This is both wrong from a script perspective, and unworkable from a character encoding perspective.

3. HPHAGS-PA LETTER OE [A856] and HPHAGS-PA LETTER UE [A857]

N2745 proposes the encoding of two vowel letters not included in N2622, that is the letters HPHAGS-PA LETTER OE [A856] and HPHAGS-PA LETTER UE [A857]. It is my belief that the encoding of these two precomposed letters is neither justified from a script perspective nor desirable from a character encoding perspective.

4. HPHAGS-PA LETTER OE and UE Second Medial/Final Forms

N2745 proposes to treat occurrences of the letters **O** and **U** as glyph variants of the letters **OE** and **UE** in "feminine" words, following the treatment of the letters **OE** [1825] and **UE** [1826] in the Mongolian encoding model. This is contrary to modern scholarly practice, under which the Phags-pa letters **O** and **U** are never transcribed as [ö] and [ü], even in "feminine" words. To accept this encoding strategy would just cause confusion for end-users, and would probably result in the same Mongolian Phags-pa word being encoded differently by different people.

5. Simple Glyph Variants for A852, A86B and A87A

N2745 proposes the encoding of simple glyph variants for A852 [HPHAGS-PA LETTER EE], A86B [HPHAGS-PA LETTER HAN WA] and A87A [HPHAGS-PA LETTER HAN HALF YA] by means of variation selectors. This is totally contrary to the fundamental principle that ISO/IEC 10646 encodes characters not glyphs.

6. HPHAGS-PA JOINER [A844]

The proposed "joiner" character A844 is categorically unnecessary, as the ligature between adjacent Phags-pa letters should be achieved automatically by the rendering system in conjunction with "smart" font technology, as is the case with the Mongolian script. If such a character were to be accepted, then some users (with basic rendering technology and fonts) might insert A844 into the code sequence after almost every non-terminal Phags-pa letter in a syllable unit, whilst other users (with sophisticated rendering technology and fonts) should never have occasion use it, resulting in the same text being encoded differently by different users.

7. Subjoined Letter RA and Superfixed Letter RA

The subjoined and superfixed forms of the letter RA proposed for encoding as distinct characters in N2622 are treated as positional or contextual forms of the letter RA in N2745, and in non-medial contexts are selected using variation selectors. I believe that for consistency with the subjoined letters WA [A879] and YA [187A] that are encoded as distinct characters in N2745, and for ease of transcoding between Phags-pa and Tibetan, encoding Subjoined Letter RA as a distinct character would be a better solution. It is not as important to encode Superfixed Letter RA as a separate character, rather than use a variation selector to select it, but I think that to do so would simplify matters for the end user.

8. HPHAGS-PA LETTERS HAN YA/SHA/HA/FHA [A87F..A882]

N2622 proposes that four variant forms of the letters YA, SHA, HA and FA used exclusively in the Chinese Phags-pa rhyming dictionary Menggu Ziyun be encoded as standardized variants. N2745 proposes that these four variant letters be encoded as distinct characters. I believe that as the variant forms of the letters are graphically very similar to the standard forms of the letters, encoding them as distinct characters will cause confusion amongst users as to which character to use in order to represent the letters YA, SHA, HA and FA in ordinary usage. Encoding these special variant letters as standardized variants will prevent accidental misuse of them to encode texts other than Menggu Ziyun. Encoding as standardized variants will also have the advantage of allowing the variation selector to be stripped out or ignored in contexts where there is no need to distinguish between the standard and variant forms of the letters.

9. Presentation Forms

The list of ninety-one presentation forms given in N2745 is analysed, and the relationship between these presentation forms and the rendering model proposed in N2622 is discussed.

10. Variation Selectors

N2745 proposes to use three variation selectors, VS-1..VS-2 [FE00.FE02], for various purposes. I believe that most of these uses are inappropriate, and that the only valid use of variation selectors can be met with one variation selector, not three. Furthermore, I believe that it would be more appropriate to use a Phags-pa specific Free Variation Selector (FVS), as suggested in N2719, rather than an ordinary variation selector [FE00].

11. Punctuation Marks [A840..A843]

Phags-pa texts do not normally use punctuation marks. Those that do simply borrow Chinese or Mongolian punctuation marks. N2745 proposes to encode Phags-pa clones of 3002 [IDEOGRAPHIC FULL STOP], 1802 [MONGOLIAN COMMA], 1803 [MONGOLIAN FULL STOP] and 1805 [MONGOLIAN FOUR DOTS]. I do not believe that there is any need to duplicate the encoding of these characters, as their occasional use in Phags-pa texts can be represented quite appropriately by the existing CJK and Mongolian characters.

12. Tibetan Phags-pa Characters

N2745 does not include any of the special Tibetan-usage Phags-pa punctuation marks proposed in N2622. These marks are essential for writing Tibetan-style Phags-pa, and if they were not to be encoded then the only living usage of the Phags-pa script would be excluded from Unicode representation.

13. Character Names

Many of the proposed names include an unnecessary linguistic qualifier, contrary to Rule 9 of the "Character-naming guidelines" (N2652R Annex L). Linguistic usage should be indicated in the code chart notes, not the character names.

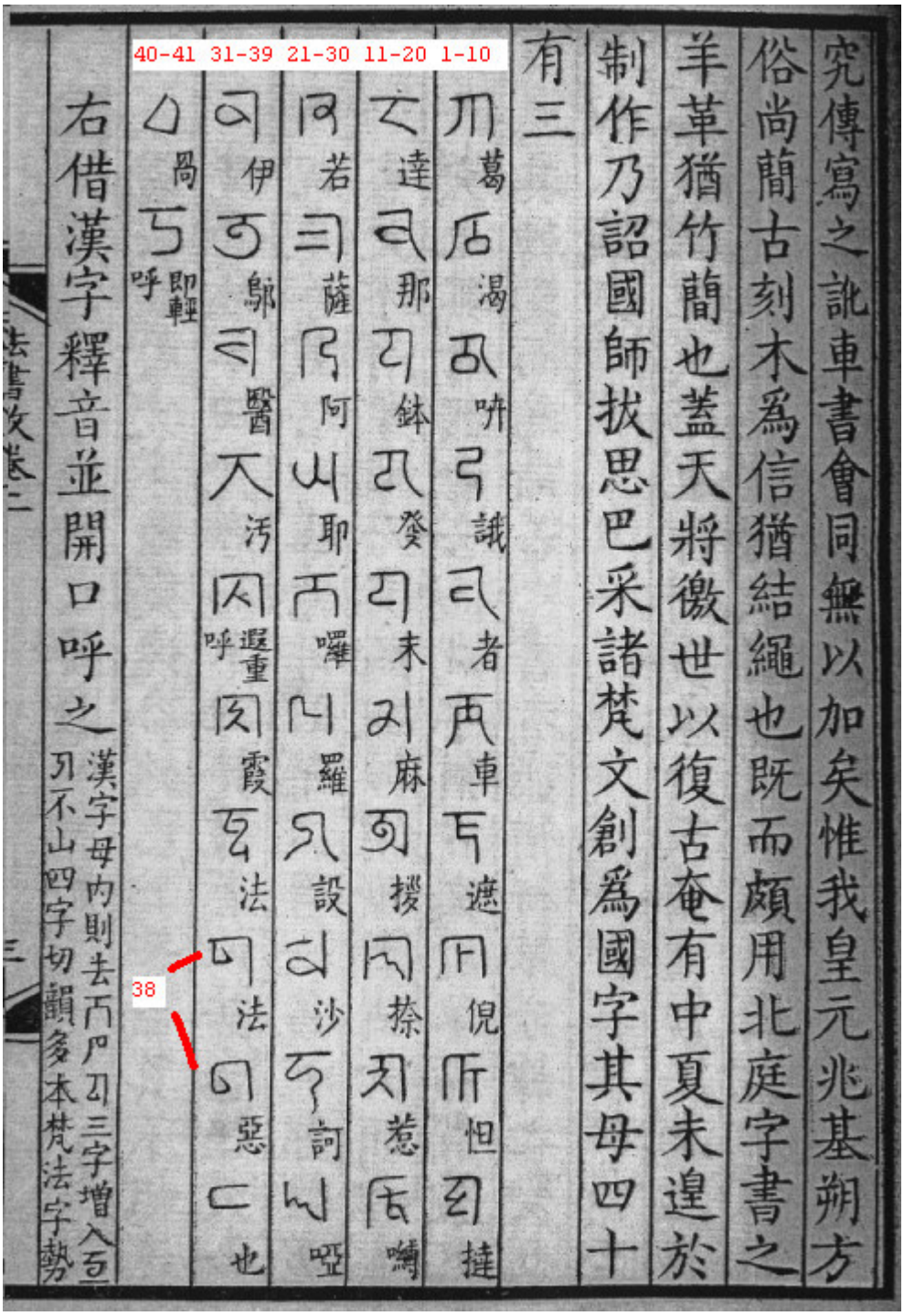
14. Reference Glyphs

The font used in N2745 is beautiful, but I believe that the font used in N2622 is better suited to the purpose of facilitating character lookup in the ISO/IEC 10646 and Unicode code charts.

2. Specific Comments on N2745

2.1 Code Point Order

Illustration 2-1-1 : Fashu Kao (1334)



Source : Lianting Shier Zhong 楝亭十二種 vol.2 folio 3a.

Illustration 2-1-2 : Shushi Huiyao (1376)

Source : Shushi Huiyao 書史會要 vol. 7 folio 22a.

N2745-1 page 7 states :

> 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.

Whilst this is true to a certain extent, it is also true that in the earliest Chinese descriptions of the Phags-pa script the forty-one basic letters are ordered according to the traditional order of the Tibetan script; and given that the creator of the script, the 'Phags-pa Lama, was himself Tibetan, there can be little reason to doubt that the original order of the letters followed the corresponding Tibetan order.

Illustration 2-1-1 and 2-1-2 show tables of the forty-one basic Phags-pa letters that are given in two fourteenth-century

Chinese works on calligraphy, Fashu Kao 法書考 (a work on calligraphy composed by the Yuan dynasty Uighur official Sheng Ximing 盛熙明, first published in 1334), and Shushi Huiyao 書史會要 (a work on the history of calligraphy by the late Yuan / early Ming author Tao Zongyi 陶宗儀, first published in 1376, eight years after the fall of the Mongolian Yuan dynasty).

Note that the forms of the Phags-pa letters are quite corrupt in these two works, and in Fashu Kao the letter GGA (No.38) has been duplicated. Nonetheless, their early date makes them very important source materials, and in all probability they are derived from an earlier official document describing the new script.

The order of the forty-one basic Phags-pa letters given in these two sources is presented in Table 2-1-1, which also shows the proposed code points for these characters in N2622 and N2745, as well as the corresponding Tibetan characters.

Table 2-1-1 : The Order of the Forty-One Basic Phags-pa Letters

No.	'Phags-pa Letter	Short Name [N2622]	Corresponding Tibetan Character	Code Point [N2622]	Code Point [N2745]
1		KA	0F40 TIBETAN LETTER KA ཀ	A840	A858
2		KHA	0F41 TIBETAN LETTER KHA ཁ	A841	A859
3		GA	0F42 TIBETAN LETTER GA ག	A842	A85A
4		NGA	0F44 TIBETAN LETTER NGA ང	A843	A85B
5		CA	0F45 TIBETAN LETTER CA ཅ	A844	A85C
6		CHA	0F46 TIBETAN LETTER CHA ཆ	A845	A85D
7		JA	0F47 TIBETAN LETTER JA ཇ	A846	A85E
8		NYA	0F49 TIBETAN LETTER NYA ཉ	A847	A85F
9		TA	0F4F TIBETAN LETTER TA ཏ	A84C	A860
10		THA	0F50 TIBETAN LETTER THA ཐ	A84D	A861
11		DA	0F51 TIBETAN LETTER DA ད	A84E	A862
12		NA	0F53 TIBETAN LETTER NA ན	A84F	A863
13		PA	0F54 TIBETAN LETTER PA པ	A850	A864
14		PHA	0F55 TIBETAN LETTER PHA ཕ	A851	A865
15		BA	0F56 TIBETAN LETTER BA བ	A852	A866
16		MA	0F58 TIBETAN LETTER MA མ	A853	A867
17		TSA	0F59 TIBETAN LETTER TSA ཙ	A854	A868
18		TSHA	0F5A TIBETAN LETTER TSHA ཚ	A855	A869
19		DZA	0F5B TIBETAN LETTER DZA ཛ	A856	A86A
20		WA	0F5D TIBETAN LETTER WA ཝ	A857	A86B

21	ཞ	ZHA	0F5E TIBETAN LETTER ZHA ཞ	A858	A86C
22	ཟ	ZA	0F5F TIBETAN LETTER ZA ཟ	A859	A86D
23	འ	-A	0F60 TIBETAN LETTER -A འ	A85A	A85E
24	ཡ	YA	0F61 TIBETAN LETTER YA ཡ	A85B	A86F
25	ར	RA	0F62 TIBETAN LETTER RA ར	A85C	A870
26	ལ	LA	0F63 TIBETAN LETTER LA ལ	A85D	A871
27	ཤ	SHA	0F64 TIBETAN LETTER SHA ཤ	A85E	A872
28	ས	SA	0F66 TIBETAN LETTER SA ས	A85F	A873
29	ཧ	HA	0F67 TIBETAN LETTER HA ཧ	A860	A874
30	ཨ	A	0F68 TIBETAN LETTER A ཨ	A861	A850
31	ི	I	0F72 TIBETAN VOWEL SIGN I ◌ི	A86A	A853
32	ུ	U	0F74 TIBETAN VOWEL SIGN U ◌ུ	A86B	A855
33	ེ	E	0F7A TIBETAN VOWEL SIGN E ◌ེ	A86C	A852
34	ོ	O	0F7C TIBETAN VOWEL SIGN O ◌ོ	A86D	A854
35	ཀ	QA		A862	A875
36	ཁ	XA		A863	A876
37	ཋ	FA		A864	A877
38	ག	GGA		A865	A878
39	ཱ	EE		A86E	A851
40	ྭ	Subjoined WA	0FAD TIBETAN SUBJOINED LETTER WA ◌ྭ	A866	A879
41	ྱ	Subjoined YA	0FB1 TIBETAN SUBJOINED LETTER YA ◌ྱ	A867	A87A

These forty-one letters are grouped into four groups of consonants, vowels and semi-vowels :

- Letters 1-30 correspond to the thirty basic Tibetan consonants (KA through A), listed in the Tibetan dictionary order.
- Letters 31-34 correspond to the four primary Tibetan vowels (I, U, E and O), also in Tibetan dictionary order.
- Letters 35-38 are additional consonants representing sounds that do not occur in Tibetan.
- Letter 39 is an additional vowel, representing a vowel quality not found in Tibetan, but used in writing Mongolian and Chinese.
- Letters 40-41 are semi-vowels corresponding to Letters 20 and 24 respectively, and can only be attached to a preceding base consonant (including the null consonant A). They are used in Chinese and Tibetan, but not in Mongolian (except where transliterating Chinese proper names).

N2622 follows the order of letters within each of these four groups, but moves the vowel letters (letters 31-34 and 39) to after the semi-vowels, and interpolates the extended Phags-pa letters into the sequence as appropriate.

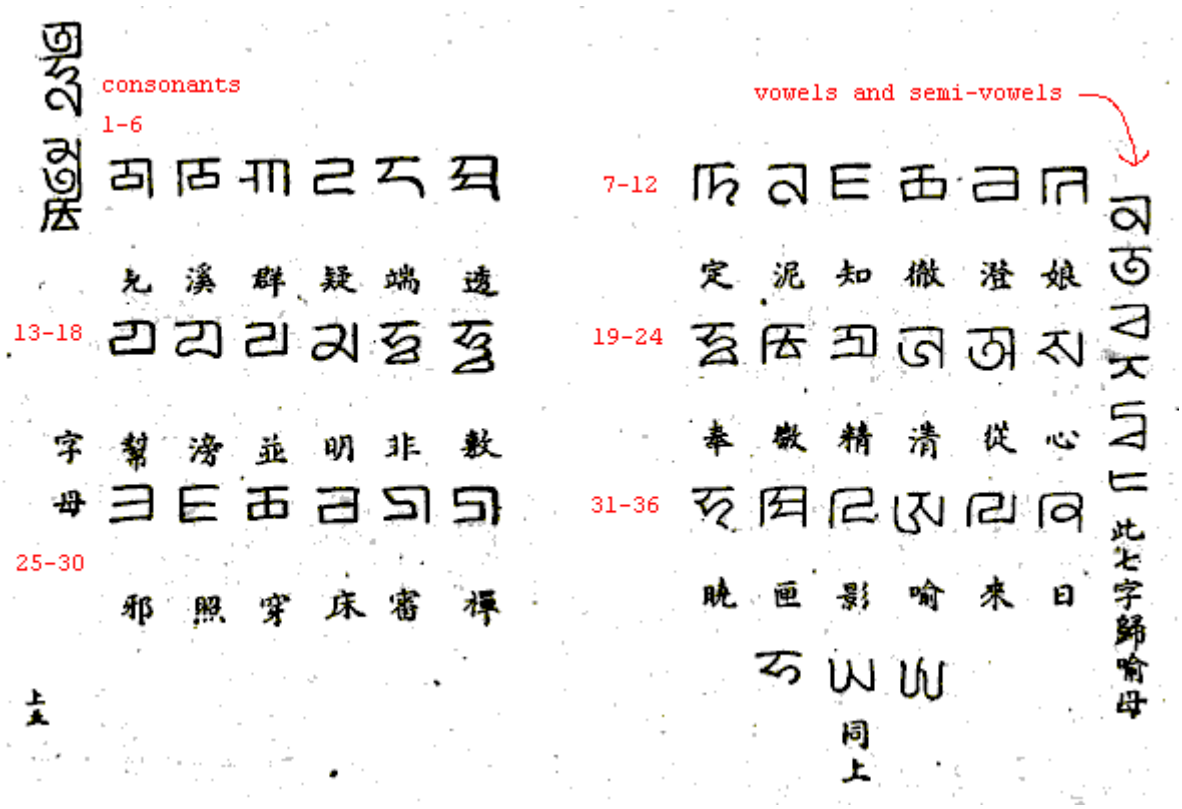
N2745 places the null consonant letter **A** (normally transliterated as **'a**) at A850, followed by the five vowels plus OE and UE (but following the Mongolian dictionary order for vowels, not the Tibetan dictionary order), followed by the other consonants (letters 1-29) and the semi-vowels (letters 40-41). The extended Phags-pa letters are appended at the end.

Placing the vowels before the consonants, and appending the extended Phags-pa letters after the forty-one basic letters is an acceptable alternative to the ordering proposed in N2622. However, I have two objections to the ordering proposed in N2745 :

1. The null consonant letter **A** should not be placed with the vowels, as it is not a vowel (see further below), but should be placed as the thirtieth consonant letter, following the Tibetan order of letters, and the ordering of letters given in Fashu Kao and Shushi Huiyao.
2. The vowel letters should be ordered according to Tibetan dictionary order (**I**, **U**, **E**, **O**), following Fashu Kao and Shushi Huiyao, rather than Mongolian dictionary order (**A**, **E**, **I**, **O**, **U**, **OE**, **UE**). Ordering according to Mongolian order may be more convenient for Mongolian, but the Phags-pa script was used to write Chinese and other languages (indeed there are more Chinese language Phags-pa texts surviving than Mongolian Phags-pa texts), and an ordering biased towards Mongolian simply makes things more difficult for Chinese and other languages, especially Tibetan, which is the only language that is still written using the Phags-pa script (albeit for decorative purposes only).

These two objections are strengthened by the only other early source that enumerates the individual Phags-pa letters, that is Menggu Ziyun 蒙古字韻 (a rhyming dictionary of Chinese with Phags-pa spellings, revised and edited by Zhu Zongwen 朱宗文 in 1308) :

Illustration 2-1-3 : Table of 36 Initials in Menggu Ziyun



Source : Menggu Ziyun 蒙古字韻 1st fascicle folio 5.

Although in this dictionary the consonants are ordered according to the traditional order of consonant sounds used in Chinese rhyming dictionaries (the so-called "36 initials") rather than the Tibetan dictionary order of consonants, the null consonant **A** is nevertheless treated as a consonant not a vowel, as it is listed as the 34th consonant. Indeed, in Menggu Ziyun Chinese words spelled with Phags-pa letter **A** initially correspond to Ancient Chinese [j] or [ŋ] initials, whereas it is words spelled with Phags-pa letter **-A** [A86E] initially that correspond to Ancient Chinese [Ø].

This dictionary thus provides further proof that the Phags-pa letter **A** was considered to be a consonant during the period during which the Phags-pa script was actively used for writing Mongolian and Chinese (1269-1368), and not a vowel.

The mistaken interpretation of A850 as a vowel rather than a null consonant in N2745 follows Poppe's influential treatise on the corpus of Mongolian Phags-pa inscriptions (original Russian edition published in 1941, English translation published in 1957), where he includes the letter A [A850] within the table of Phags-pa vowels (see Illustration 2-2-1 below).

However, as is well-known, Tibetan and other Brahmic scripts, do not have an explicit sign for the vowel [a], but each base consonant in a syllable unit has an inherent [a] sound, which is only overridden by an explicit sign representing some other vowel sound. Thus the letter KA by itself represents the syllable [ka], but the letter KA followed by the letter I represents the syllable [ki]. In order to represent a vowel initially Brahmic scripts need a null consonant which either carries the inherent [a] sound or to which an explicit vowel letter may be attached. The Phags-pa script inherits a null consonant letter from Tibetan, which is used to allow the representation of an initial [a] in a word, or to act as a base for semi-vowels (subjoined or half-form letter WA), digraphs (**ĖU** and **ĖO**), diphthongs (**UE**) and rarely pure vowels (**E**). The only difference between Phags-pa and Tibetan usage is that the vowels **I**, **U**, **E** and **O** are independent letters in Phags-pa, not dependent signs as in Tibetan, and normally occur initially without the null consonant in Mongolian and Chinese Phags-pa texts, except when immediately followed by another vowel.

The interpretation of the letter A [A850] as a vowel letter per se is immediately seen as wrong by anyone who has any understanding of Tibetan and other related Brahmic scripts. Moreover, China's foremost authority on the Phags-pa script, Professor Junast, has written a monograph on the Phags-pa null consonant, where he emphatically declares that this letter is in no shape or form a vowel letter :

> As to the letter 𰸏 (’), according to the traditional explanation given in the past this is a special letter representing the vowel **a**. The representative proponent of this point of view is Professor Nicholas Poppe. Although he correctly interprets the occurrence of the letter “ ’ ” before a vowel letter as a "letter head", i.e. a null consonant letter, in other circumstances where “ ’ ” is used he regards the letter as specifically representing the vowel **a**. In other words, as far as Professor Poppe is concerned, the single letter “ ’ ” has two functions. This point of view has also exerted quite widespread influence within the academic study of the Phags-pa script within our own country [China]. The question is, does the letter “ ’ ” of itself really represent the vowel **a** ? Our answer is "No" : under all circumstances the letter “ ’ ” represents a null consonant, and it never represents the vowel **a**.
>
> Junast, "Basibazi zhong de ling shengmu fuhao" [The Phags-pa null consonant sign]; in Minzu Yuwen 1989.2 : 29-36.
>
> 关于𰸏（’）：按过去传统的说法，这个字母是表示词首元音 **a** 的专门字母。持这种观点的代表人物是尼古拉·波普教授。他虽然对于处在元音字母前的“’”作出了正确的解释，认为是字冠即零声母符号。但对于另一种条件下使用的“’”，他却看作是元音 a 的专门字母。换句话说，在波普教授看来，“’”一身兼有两种功能。这种观点，在我国八思巴字学术界也有相当广泛的影响。但问题是“’”真的是元音 **a** 的专门字母吗？我们的回答是否定的。“’”在任何条件下所表达的都是零声母，从来不表示 a。
>
> 照那斯图，《八思巴字中的零声母符号》。载《民族语文》 1989.2 : 29-36。

This quotation further demonstrates how inappropriate it is to place the letter **A** with the vowel letters, rather than with the consonant letters. Whilst the code point position of a single letter may seem relatively insignificant, the mistaken treatment of this letter as a vowel rather than a null consonant is the most important flaw in N2745, and this misplacement of the letter's code point is a reflection of the fundamental misunderstanding of the nature of the letter.

In addition to demonstrating that the letter **A** should be treated as a consonant, Illustration 2-1-3 also demonstrates that the correct ordering of the vowel letters should be according to the Tibetan dictionary order, as it shows the five vowels and two semi-vowels appended at the far right of the page, in the order **I**, **U**, **E**, **O**, **Ė**, **WA** and **YA** (**Ė** and **WA** are ligated together due to textual corruption). This is the same order as given in Fashu Kao and Shushi Huiyao, as well as the order chosen in N2622. In contrast, there is no pre-modern source that indicates that the Phags-pa vowels should be ordered according to Mongolian dictionary order, as proposed in N2745.

2.2 HPHAGS-PA LETTER A [A850]

As discussed in detail above (see Section 2-1), the Phags-pa letter **A**  [A850] (normally transliterated with an apostrophe ’, and known as the letter ’**A**) is a null consonant used at the start of a word either to support an inherent [a] vowel sound or to support the letters **Ė** or subjoined/half-form **WA** (or rarely the vowel **E**).

However, in N2745 the occurrence of the null consonant letter **A** before the letters **E**, **O**, **U**, **OE** (i.e. **ĖO**), **UE** (i.e. **ĖU**) and subjoined/half-form **WA** is not treated as a letter at all, but merely as part of the glyph variant of the particular letter in the initial position (see N2745-3 Section III "Variant Presentation Glyphs" nos. 0006, 0012, 0016, 001B, 001E and 004A). These null consonant glyph variants are shown in Table 2-2-1, together with the corresponding glyphs for the same letter in medial or final positions.

Table 2-2-1 : N2745 Presentation Forms of Letters with a Null Consonant Initial

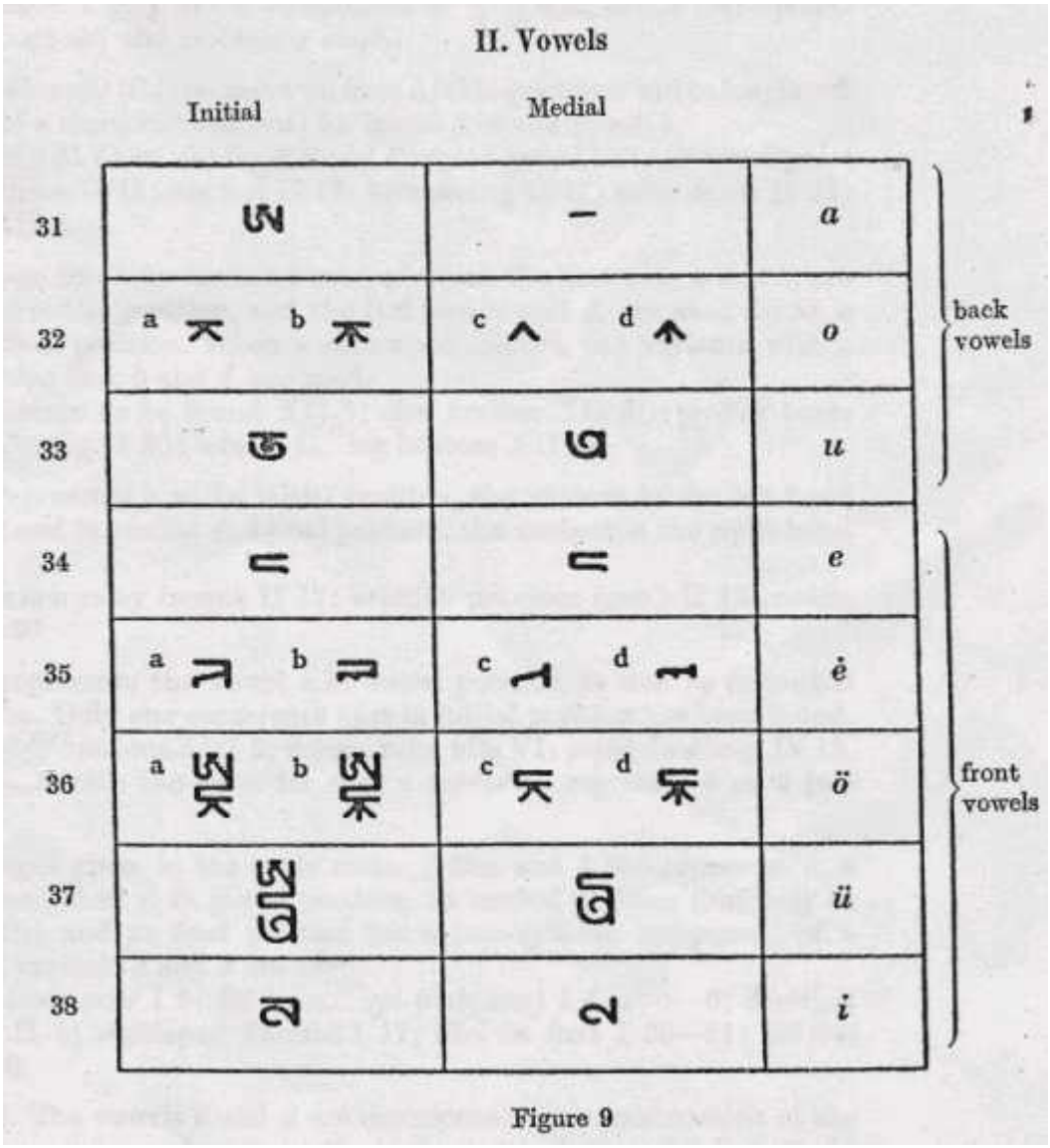
Letter	Glyph ID	Glyph	Encoding	Corresponding Glyph(s) in Medial/Final Positions
E	0006		A852, FE01 (VS2)	
O	0012		A854, FE00 (VS1)	
U	0016		A855, FE00 (VS1)	
OE	001B		A856	
UE	001E		A857	
Half U	004A		A879	

For the proposed letters **OE** [A856], **UE** [A857] and **Half U** (subjoined WA) [A879] the glyph form with the null consonant is given as the default glyph for the letter in an initial position, whereas for the letters **E** [A852], **O** [A854] and **U** [A855] the glyph form with the null consonant is selected by VS-1 or VS-2, as the normal initial form of these letters does not need to be prefixed by a null consonant.

Clearly, some might even say self-evidently, the null consonant is a separate letter, not simply part of the glyph form of certain letters in an initial position, and needs to be encoded as such. I can only guess that the reason why the null consonant has been airbrushed out of the encoding picture in N2745, is that it is not considered necessary or useful to encode a silent letter, and that ignoring the null consonant somehow helps in the collation of Mongolian words written in Phags-pa and Mongolian scripts. Presumably, by the same reasoning the initial "kn" in English words such as "knight" should be encoded as a glyph variant of the letter "n" (e.g. <006E, FE00, 0069, 0067, 0068, 0074>) in order to facilitate phonetic sorting in English. This approach to character encoding is just plain wrong.

The source of this misunderstanding of the nature of the null consonant is almost certainly the extremely influential The Mongolian Monuments in HP'AGS-PA Script (1941, English translation 1957), written by the Russian scholar Nicholas Poppe over sixty years ago, in which he shows the initial forms of the letters "ö" and "ü" with the null consonant attached :

Illustration 2-2-1 : Poppe's Table of Phags-pa Vowels



Source : The Mongolian Monuments in HP'AGS-PA Script Figure 9.

However, this table should not be taken out of context (and unfortunately, for some Chinese scholars who do not read Russian or English, it often is), as it is clear from elsewhere in his book that Poppe appreciated that the null consonant was a separate letter, and that it was only omitted from his transcriptions of Phags-pa words in order to more clearly indicate the phonetic correspondence between Mongolian words spelled in the Phags-pa and Mongolian scripts :

> The text of each monument is reproduced in transliteration and transcription in Latin letters. ... With regard to the transliteration and transcription of the texts, it should be pointed out that in the transliteration I have endeavoured to reproduce every sign of the hP'ags-pa alphabet with the greatest accuracy. For this reason, those hP'ags-pa letters which are made up of several signs are given in like manner. Consequently, §36c and §37b are given as *eo* and *eu*. ... §31 is transliterated by ', and §36a or §37a are given as *'eo* and *'eu* respectively. ... In the transcription, I have striven to reproduce not so much those elements which make up the signs of the hP'ags-pa alphabet as the pronunciation of the words.
>
> Nicholas Poppe, The Mongolian Monuments in HP'AGS-PA Script (1957) page 45.

In more recent transcriptions of Phags-pa texts, in both China and elsewhere, the null consonant is always transcribed with an apostrophe. See for example Illustration 2-4-2 (Beijing, 1994), which includes the following transcriptions of Phags-pa words with null consonant initials :

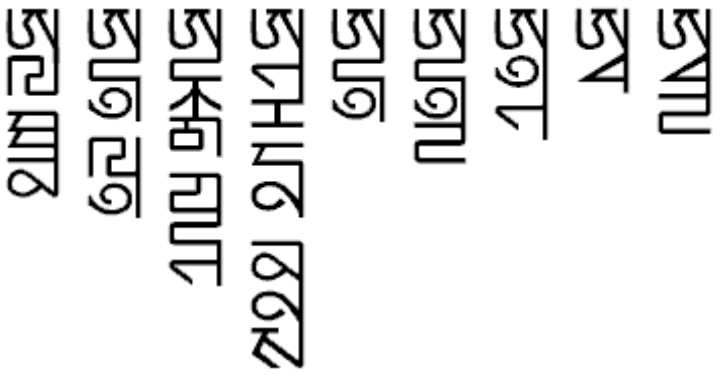
- ’ök‘ödeė
- ’öl̆jeėt‘u
- ’ali
- ’alba
- ’ülu
- ’ül̆jen
- ’ögun

The fact that the null consonant is transcribed by an apostrophe in scholarly transcriptions of Phags-pa texts such as this clearly indicates that it is considered to be a distinct letter in its own right. One of the stated aims of N2745 is the "automatic transliteration of HPhags-pa text into Latin alphabet" (N2745-1 p.5), but obviously the treatment of the null consonant in N2745 merely hinders this goal.

Examples of various Mongolian words (Nos.1-4) and Chinese syllables (Nos.5-9) that are spelled with an initial null consonant are given in Table 2-2-2. This table contrasts the different encoding of these words following N2622 on the one hand, and N2745 on the other. It should be noted that N2622 encodes each discrete letter of the word separately, whereas N2745 does not recognise the null consonant unless it is supporting an inherent vowel [a].

Table 2-2-2 : Examples of Mongolian and Chinese Words with Null Consonant Initials

1. 2. 3. 4. 5. 6. 7. 8. 9.



No.	Meaning	Transliteration	Encoding [N2622]	Encoding [N2745]
1.	"golden"	**'al** than	**A861 A85D** 202F A84D A84F	**A850 A871** 202F A861 A863
2.	"not"	**'ėu** lu	**A861 A86E A86B** 202F A86B	**A857** 202F A871 A855
3.	"gave"	**'ėog** bėe	**A861 A86E A86D A842** 202F A852 A86E A86C	**A856 A85A** 202F A866 A851 A852
4.	"jewels"	**'er** di nis	**A861 A86C A85C** 202F A84E A86A 202F A84F A86A A85F	**A852 FE01 A870** 202F A862 A853 202F A863 A853 A873
5.	yu 魚	**'ėu**	**A861 A86E A86B**	**A857**
6.	yong 顒	**'ėung**	**A861 A86E A86B A843**	**A857 A85B**
7.	wei 危	**'ue**	**A861 A86B A86C**	**A855 FE00 A852**
8.	wa 瓦	**'wa**	**A861 A866**	**A879**
9.	wang 王	**'wang**	**A861 A866 A843**	**A879 A85B**

Note that one undesired consequence of the encoding approach proposed in N2745 is that common Chinese words such as wei **危** "dangerous" (Table 2-2-2 no.7) have to be encoded using a variation selector, simply because Chinese words are spelled with a null consonant before the letter **U** when part of a diphthong, whereas the letter **U** never occurs after a null consonant in Mongolian.

The encoding model proposed in N2745 also causes problems for the encoding of Tibetan Phags-pa texts, which have different orthographic rules to Mongolian and Chinese. The main difference is that in Tibetan Phags-pa texts an initial null consonant is usually required in front of all vowel letters. See Example 5 in N2622, where the right-hand column of Phags-pa

text reads **'i 'u 'e 'om**, with the null consonant letter **A** preceding the letters **I**, **U**, **E** and **O**. Following N2745 **'u**, **'e** and **'om** would have to be encoded using variation selectors as <A855, FE00>, <A852, FE01>, and <A854, FE00 A867> respectively. As N2745 does not define a presentation form for **'i**, it is not clear how this Tibetan syllable should be encoded following this encoding model.

Not only does the N2745 encoding model force the use of variation selectors for writing normal Tibetan words in the Phags-pa script, but it destroys the general one-to-one correspondence between Phags-pa letters and Tibetan letters, which makes automatic script conversion for Tibetan more difficult. As Tibetan usage of the Phags-pa script is the only living usage of the script, any encoding model for Phags-pa should not hinder the encoding of Tibetan Phags-pa texts. The encoding model proposed in N2622 was arrived at after a thorough examination of Phags-pa script usage in Phags-pa texts written in the Mongolian, Chinese, Uighur, Sanskrit and Tibetan languages; and in consequence the encoding model is equally suited to all these languages. On the other hand, the encoding model proposed in N2745 is strongly biased towards Mongolian usage, and attempts to artificially distort the Phags-pa script so that it more closely mirrors the Mongolian encoding model, when in fact much of the Mongolian encoding model is not appropriate to the Phags-pa script.

The treatment of the null consonant in N2745 is not only wrong from a script perspective, but I believe that it is also unworkable from an encoding perspective. The problem is that although N2745 defines presentation forms for null consonant followed by initial vowels **E**, **O**, **U**, **OE** and **UE** and the semi-vowel **Half U** (subjoined WA), there is nothing to stop anyone from encoding words with an initial null consonant with the letter **A** [A850] followed by the appropriate vowel or semi-vowel letter. As the proposed medial and final form glyphs for the vowel letters and semi-vowel **Half U** do not have a null consonant component, A850 plus vowel/semi-vowel would render identically to the initial form glyph of the vowel/semi-vowel proposed in N2745. Thus, for the example words in Table 2-2-2, each syllable could be encoded in two different ways to produce exactly the same rendering, as shown in Table 2-2-3.

Table 2-2-3 : Different Encoding Strategies for Words with Null Consonant Initials

No.	Syllable	Encoding without A850	Encoding with A850
2.	'ėu	A857	A850 A857
3.	'ėog	A856 A85A	A850 A856 A85A
4.	'er	A852 FE01 A870	A850 A852 A870
5.	'ėu	A857	A850 A857
6.	'ėung	A857 A85B	A850 A857 A85B
7.	'ue	A855 FE00 A852	A850 A855 A852
8.	'wa	A879	A850 A879
9.	'wang	A879 A85B	A850 A879 A85B

Whilst the authors of N2745 may claim that the encoding given in the last column of Table 2-2-3 is wrong (according to their encoding model), I cannot think of any way of explicitly outlawing such variant encoding strategies. Moreover, there is every reason to believe that many users would, either deliberately or by accident, encode words with a null consonant in the manner shown in column 4 of Table 2-2-3. In fact, unless the user had read the "User's Agreement" (N2745-3) he would not know that A850 was not intended to be used before vowel and semi-vowel letters, and so would naturally use A850 to represent a null consonant in every circumstance that it occurred. It may be assumed that many users will indeed rely solely on the published Unicode or ISO/IEC 10646 code charts, and thus not follow the encoding model proposed in N2745. In addition to such accidental deviation from the N2745 encoding model, there may well be many users who would deliberately follow the alternate encoding strategy of always encoding the null consonant with A850, especially as this avoids the necessity of using variation selectors (which would be needed for many words in both Chinese and Tibetan following the N2745 encoding model). This alternative encoding strategy will almost certainly be widely used by those encoding Tibetan Phags-pa texts, as the null consonant is always encoded separately in the Tibetan script. Indeed, there would be no alternative for the syllable **'i** as N2745 does not propose a presentation form for null consonant followed by the letter **I**.

The result is that different people are going to encode the same Phags-pa syllables differently in order to achieve exactly the same glyph rendering. This is a recipe for chaos. The only way that these two different encoding strategies could possibly be reconciled is if the sequences shown in columns 3 and 4 of Table 2-2-3 were to be made canonically equivalent. However, it is impossible to decompose the sequences shown in column 3 into the sequences shown in column 4 as the null consonant component only occurs in the initial position of a word. For example, **A879** (**Half U**) cannot be canonically equivalent to the sequence **A850 A879** (**'wa**) as :

1. A879 is only equivalent to A850 A879 in initial positions, medially and finally it is equivalent to itself (A879)
2. A879 would end up being recursively decomposed to A850 A879 ad infinitum

I suggest that the logical impossibility of canonically equating two different encoding sequences that represent exactly the same glyph sequence makes the null consonant encoding model proposed in N2745 completely unviable.

2.3 HPHAGS-PA LETTER OE [A856] and HPHAGS-PA LETTER UE [A857]

The Phags-pa script has five vowel letters, generally transcribed as **i**, **u**, **ė** (or **e**), **o** and **e** (or **ė**). All sources dating from the 14th-century, when Phags-pa was the official script of the Mongolian empire, agree upon these five vowel letters (see Illustrations 2-1-1, 2-1-2 and 2-1-3 above).

Whilst the Phags-pa script only has five vowels letters, the languages that it is used to write have more than five vowels sounds. Diphthongs are represented by a combination of a vowel letter followed by the letter **WA** (e.g. **aw**, **ew**, **ėw**, **iw**, **uw**, **hiw** and **ow** in Chinese), **YA** (e.g. **ay**, **iy** and **hiy** in Chinese) or **E** (e.g. **ue** and **ėue** in Chinese, and **ėe** in Mongolian). In addition to such diphthongs, a number of pure vowel sounds are also represented by digraphs :

- The letter sequence **ėu**  is used to represent [ü] in Mongolian and Chinese
- The letter sequence **ėo**  is used to represent [ö] in Mongolian and Chinese
- The letter sequence **ėi**  is used to represent an uncertain vowel quality in Chinese only
- The letter sequence **hi**  is used to represent [ŋ] in Chinese only

The Phags-pa letter sequences **ėu** and **ėo** are often transcribed as **ü** and **ö** respectively (see Illustration 2-4-2), and likewise the letter sequence **hi** is sometimes transcribed as **ï**. However this is simply a phonetic transcription, and no more means that the sounds [ü] and [ö] can be considered to correspond to a single Phags-pa letter than the fact that English "ng" is represented as [ŋ] in IPA means that "ng" should be considered as a single letter; and there is no more reason to encode **ėu** and **ėo** as precomposed characters than there is to encode **ng** as a precomposed character.

As with the treatment of the null consonant, the proposed encoding of **ėu** and **ėo** as separate characters is ultimately based on Poppe's The Mongolian Monuments in HP'AGS-PA Script, where he lists the letter combinations representing the vowels [ü] and [ö] as if they were distinct letters (see Illustration 2-2-1). However, as Poppe himself states (see quotation given in Section 2-2), his phonetic transcription of **ėu** and **ėo** as **ü** and **ö** deliberately differs from his strict transliteration of Phags-pa texts, where he always transliterates the letter sequences **ėu** and **ėo** as two letters.

That the Phags-pa letter sequences **ėu** and **ėo** should not be considered to be discrete entities is recognised by authoritative modern scholars of the Phags-pa script, such as Professor Junast, who states :

> The Phags-pa script has five vowels :  [i] <A86A>, [u] <A86B>, [ė] <A86C>, [o] <A86D> and [e] <A86E>. … In Phags-pa orthography certain sounds are represented by means of digraphs, including both vowels and consonants. For example, the Mongolian "feminine" vowels ö and ü are represented by the combination sequences <A86E> and <A86D> (i.e. eo) and <A86E> and <A86B> (i.e. eu) respectively.
>
> Junast, "Basibawen yuanyin zimu zixing wenti shang de liangzhong tixi"; in Minzu Yuwen 1987.4.
>
> 八思巴文有五个元音子母： [i] <A86A>, [u] <A86B>, [ė] <A86C>, [o] <A86D>, [e] <A86E>。… 在八思巴文的书写法中也用双子母表示某些语音，其中包括元音和辅音，例如蒙古语的阴性元音 ö 和 ü，分别用 <A86E> 与 <A86D> 的组合即 eo 和 <A86E> 与 <A86B> 的组合即 eu 表示。
>
> 照那斯图，《八思巴文元音字母字形问题上的两种体系》。载《民族语文》1987.4。

The fact that neither 14th-century sources nor modern experts such as Professor Junast (nor even Professor Poppe) consider **ėo** and **ėu** to be discrete letters should be reason enough not to encode these two digraphs separately (as A856

and A857 respectively in N2745). However, even though the authors of N2745 admit that these are precomposed characters (see N2745-1 pp.4-5 : "Although their variant presentation glyphs are "compound letters" each consisting of two or three "elements" ..."), they seem to be suggesting that these precomposed sequences should nevertheless be encoded separately in order to facilitate sorting and other processing of Mongolian Phags-pa data, as well as the "automatic transliteration of HPhags-pa text into Latin alphabet" (N2745-1 p.5). I do not believe that such considerations should affect the determination of character repertoire. Nor do I believe that the failure to encode precomposed characters representing the letter sequences **ėo** and **ėu** in any way hinders such objectives as sorting, collation and automatic transliteration. Therefore I strongly suggest that the precomposed characters A856 [HPHAGS-PA LETTER OE] and A857 [HPHAGS-PA LETTER UE] are not accepted for encoding.

A more convincing argument for the separate encoding of precomposed letters **ėo** (OE) and **ėu** (UE) could perhaps have been made if there was otherwise a one-to-one mapping between Mongolian words spelled in the Phags-pa script on the one hand and in the Uighur-derived Mongolian script on the other, but as has been demonstrated in N2719 with the help of numerous examples, the two scripts have different orthographic rules for writing Mongolian, with the result that no simple mapping rules between the two scripts can be established for Mongolian words. Thus, the encoding of A856 [HPHAGS-PA LETTER OE] and A857 [HPHAGS-PA LETTER UE] for "compatibility with the Mongolian script" would not significantly facilitate the collation of Mongolian words written in the two scripts or the automatic conversion of Mongolian words from Phags-pa to Mongolian or vice versa.

N2652 Annex G "Formal criteria for coding precomposed characters" lists the criteria that are used to determine whether a precomposed character should be accepted for encoding in ISO/IEC 10646 or not :

- Positive:
  - Existence in another character encoding standard (for the purpose of 1:1 character conversion)
  - Existence of a precomposed letter in a well-established or official alphabet.
- Negative:
  - If it were to introduce multiple spellings (encodings) for a script where NO multiple spellings existed previously.
  - If combining character sequences can be shown to meet the stated information processing needs (e.g. archival use)
  - If solely intended to overcome short-term deficiency of rendering technology.
  - If the intended use of the character is solely for transliteration purposes.

Neither of the two positive criteria can be claimed for the encoding of the precomposed characters A856 [HPHAGS-PA LETTER OE] and A857 [HPHAGS-PA LETTER UE]. On the other hand, encoding these two precomposed characters would introduce multiple spellings, and the intended use of the two characters seems to be solely for compatibility with the phonetic transliteration of these two letter sequences as "ö" and "ü" and to facilitate transliteration between the Mongolian and Phags-pa scripts. There is also absolutely no reason why combining sequences cannot meet all information processing needs.

Nevertheless, if it is really felt necessary to encode these precomposed characters, then they should have a canonical decomposition to <A851, A854> and <A851, A855> respectively. Unfortunately, due to the fact that in initial positions A856/A857 is equivalent to <A850, A851, A854/A855> and in medial and final positions it is equivalent to <A851, A854/A855>, it is impossible to define any canonical compatibility between the precomposed characters A856 or A857 and their constituent elements.

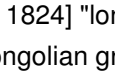
2.4 HPHAGS-PA LETTER OE and UE Second Medial/Final Forms

The "Reference Table for HPhags-pa Script" in N2745-3 lists the following variant forms of the letters **OE** and **UE** in the medial and final positions. Note that these variant medial and final forms of the letters OE and UE are not listed in "Variant Presentation Glyphs and Their Names in HPags-pa Script" in the same document, so there may perhaps be some doubt as to whether N2745 actually proposes these variant forms or not.

Table 2-4-1 : Second Medial/Final Forms of the Letters OE and UE

Basic Characters			Presentation Characters			
No.	**Graphic Symbol**	**Name**	**No.**	**Graphic Symbol**	**Name**	**Presentation Rule**
A856		HPHAL. OE	003		hphal. oe second medial form	ZWJ-  -ZWJ-VS1
			005		hphal. oe second final form	ZWJ-  -VS1
A857		HPHAL. UE	003		hphal. ue second medial form	ZWJ-  -ZWJ-VS1
			005		hphal. ue second final form	ZWJ-  -VS1

The second medial/final forms of the letters **OE** and **UE** are in fact the medial/final forms of the letters **O** and **U** respectively. N2745 does not explain the usage of these variant forms, but it is obvious that the reason why the medial/final forms of the letters **O** and **U** are proposed as variant forms of the letters **OE** and **UE** in certain circumstances is for compatibility with the encoding model for the Mongolian script, which is briefly outlined below.

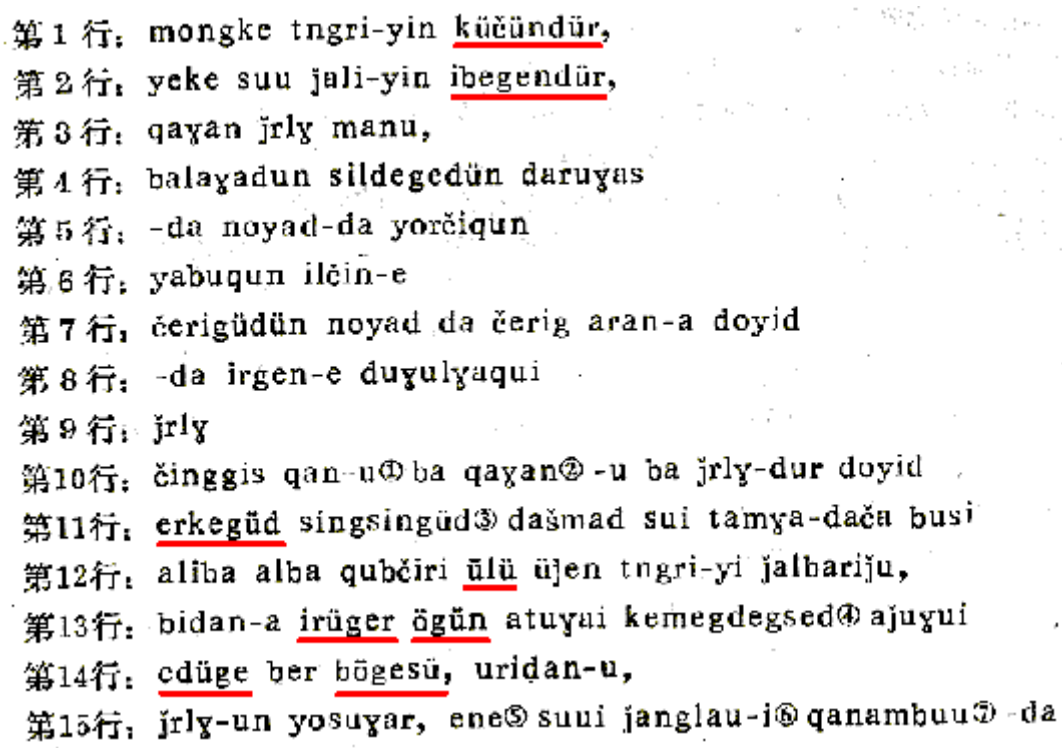
From a graphic point of view the Mongolian script really only has three distinct vowel letters (**A**/**E**, **I** and **O**/**U**). In particular the phonetic quality of the letter **O**/**U** cannot normally be determined from the glyph shape of the letter, but can only be determined by semantic context, so that, for example, **urtu**  [1824 1837 1832 1824] "long" and **ordu** [1823 1837 1833 1824] "palace, camp, horde" are written identically. However, modern Mongolian grammarians have a phonetic concept of letters, so that the vowels [o] and [u] are considered to be represented by separate letters even though they normally share exactly the same glyph. Likewise, modern Mongolian grammarians have the concept of the letters [ö] and [ü], even though these vowel sounds are represented by the digraph **O**/**U** + **I** in the first syllable of a word, and simply by the letter **O**/**U** elsewhere. Although perhaps the correct analysis is that the letter O/U is pronounced [o] or [u] in "masculine" words (i.e. words with back vowel harmony) and pronounced [ö] or [ü] in "feminine" words (i.e. words with front vowel harmony), and that the digraph **O**/**U** + **I** is used in the first syllable of a word in order to signal the fact that a word is "feminine", modern Mongolian grammarians recognise the letters **OE** and **UE** as distinct letters with two glyph forms, one glyph form that looks like the letter **O**/**U** plus the letter **I** (the glyph form that occurs in the initial syllable of a word), and one glyph form that is identical to the letter **O**/**U** (the glyph form that occurs in post-initial syllable of a word). From a character encoding perspective this phonetic-based concept of letters (**A**, **E**, **I**, **O**, **U**, **OE** and **UE**) is probably wrong, but for better or worse it was the model accepted for the encoding of the Mongolian script into ISO/IEC 10646 and Unicode. The upshot of this is that the second medial and first final forms of MONGOLIAN LETTER OE [1825] and MONGOLIAN LETTER UE [1826] are identical to the first medial and first final forms of MONGOLIAN LETTER O [1823] and MONGOLIAN LETTER U [1824] (and furthermore the respective forms of MONGOLIAN LETTER O and MONGOLIAN LETTER U are identical to each other).

By defining variant forms of the Phags-pa letters **OE** and **UE** that are identical to the Phags-pa letters **O** and **U**, N2745 is implicitly applying this aspect of the Mongolian encoding model to the Phags-pa script. However, for a number of reasons, I believe that this application of the Mongolian encoding model is not appropriate for the Phags-pa script.

Foremost, there is no tradition of ever treating the letters **O** and **U** as glyph variants of the letters **OE** and **UE** in modern Phags-pa scholarship. Even Poppe does not treat the letters **O** and **U** in post-initial syllables as the letters **Ö** and **Ü**; so, for example, his phonetic transcription includes "feminine" words such as **ülu** and **ölǰeėt‘u**, where according to N2745 the letter **u** should be encoded as a medial/final form of the letter **UE**.

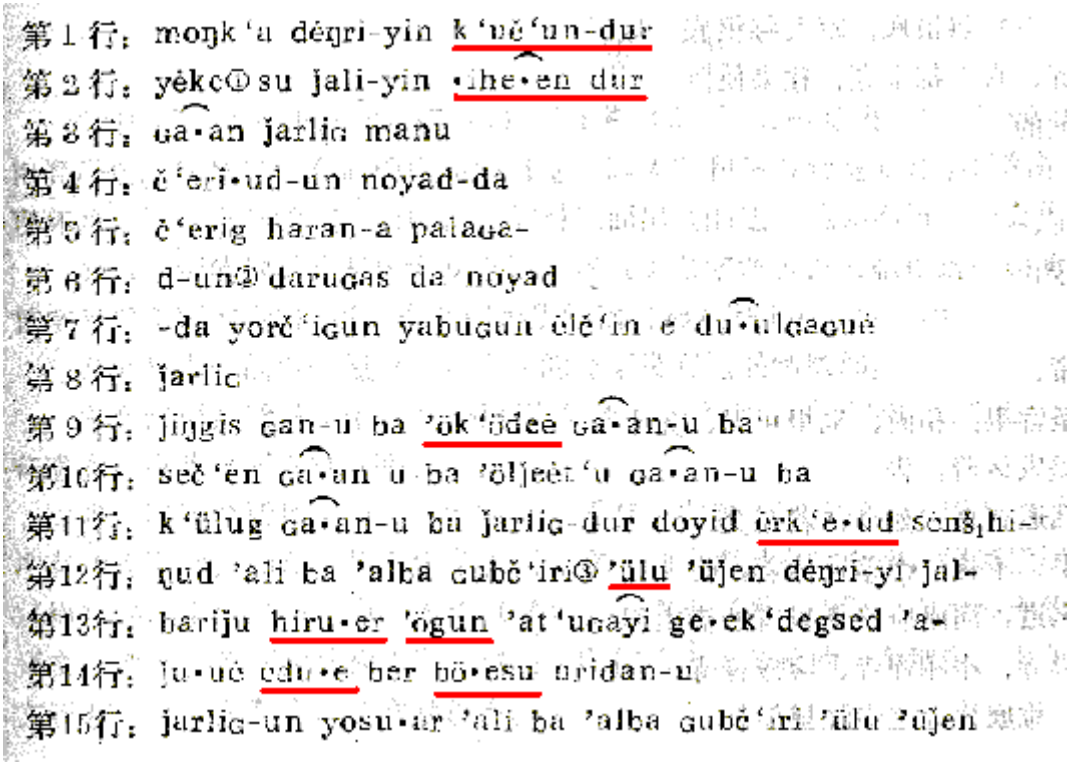
Illustrations 2-4-1 and 2-4-2 show recent scholarly transcriptions of two inscriptions of similarly-worded imperial edicts, the first in the Uighur-derived Mongolian script, and the second in the Phags-pa script.

Illustration 2-4-1 : Transcription of an Imperial Edict dated 1268 (Mongolian script)



Source : Minzu Yuwen 民族语文 1994.1 page 32.

Illustration 2-4-2 : Transcription of an Imperial Edict dated 1312 (Phags-pa script)



Source : Minzu Yuwen 民族语文 1994.1 page 37.

In these two inscriptions "feminine" words with letters **OE** or **UE** have been underlined. These are summarised in Table 2-4-2 below.

Table 2-4-2 : Comparison between Feminine Words in Mongolian and Phags-pa Scripts

Line	Mongolian	Phags-pa
1	kŭčündür	k‘uč‘un-dur
2	ibegendür	·ihe·en-dur
11	erkegüd	ėrk‘e·ud
12	ülü	’ülu
13	irüger	hiru·er
13	ögün	’ögun
14	edüge	ėdu·e
14	bögesü	bö·esu

For the Mongolian script inscription, each **U**/**O** or **UI**/**OI** glyph form in a "feminine" word is transliterated as **ö** or **ü** as appropriate. On the other hand, for the Phags-pa script inscription, the letter sequences **ėo** and **ėu** are transliterated as **ö** and **ü**, whereas the letters **O** and **U** are transliterated as **o** and **u**, indicating that the letters **O** and **U** in a "feminine" word are not considered to be glyph variants of the letters **OE** and **UE**.

It is also clear from this random sample that the two scripts have different rules for the writing of "feminine" words. Whereas, in the Mongolian script the "ui" and "oi" glyph forms of the letters **OE** and **UE** are always required in the first syllable, and never occur in later syllables, in the Phags-pa script the letter sequences **ėo** and **ėu** may occur in later syllables (e.g. the word **’ök‘ödeė** on line 9), or may be replaced by the letter **O** or **U** even in the first syllable, so that the "feminine" word **k‘uč‘un-dur** is written entirely with back vowels. Thus for Mongolian, the correct glyph form of the letters **OE** and **UE** can always be determined by the rendering engine from context, without any need for the user to apply one of the Mongolian Free Variation Selectors (an FVS would only be needed to select a particular contextual form of the letters **OE** and **UE** out of context). In contrast, the selection of Phags-pa **O**/**U** or **OE**/**UE** is not necessarily determinable from context, so that for words like **’ök‘ödeė** and **k‘uč‘un-dur** it would be necessary for the end-user to apply an appropriate variation selector to select the correct glyph form.

As Phags-pa scholars have the habit of transliterating the letters **O** and **U** as **o** and **u**, even in "feminine" words, it may be assumed that they would also tend to encode the letters **O** and **U** in words such as **ülu** and **k‘uč‘un-dur** with A854 and A855 rather than with A856 and A857 plus variation selector as suggested by N2745. Thus, the encoding model proposed in N2745 is again introducing the likelihood of different end-users encoding the same Phags-pa text using different character sequences. Especially as we are dealing with an historic script used to write stages of languages such as Mongolian and Chinese that were spoken some seven hundred years ago, and may thus be quite different from the modern languages in many respects, I do not believe that it is correct to ask end-users to make assumptions about the phonetic value of letters in a word. In the case of Phags-pa I believe that the safest encoding model, and the one most likely to be accepted by end-users, is one that allows the user to encode characters based on their graphic shape rather than their phonetic reconstruction.

2.5 Simple Glyph Variants for A852, A86B and A87A

As is the case with almost every known script in the world, some Phags-pa letters occur in variant glyph forms with no semantic or phonetic distinction between them. These variant glyph forms simply represent scribal preference. It is a fundamental principle of ISO/IEC 10646 not to encode simple glyph variants.

Nevertheless, the presentation forms listed in N2745-3 Section III include simple glyph variants for the proposed letters A852 [HPHAGS-PA LETTER EE], A86B [HPHAGS-PA LETTER HAN WA] and A87A [HPHAGS-PA LETTER HAN HALF YA]. These are referred to as "free variants" in the proposal.

The double-toothed form of A852  is proposed to be differentiated from the "nominal" single-toothed form  of the

letter by application of VS1 or VS3. The square form of A86B  is proposed to be differentiated from the "nominal" triangular form  of the letter by application of VS1. The rectangular form of A87A  is proposed to be differentiated from the "nominal" wedge-shaped form  of the letter by application of VS1. This situation is summarised in Table 2-5-1.

Table 2-5-1 : Simple Glyph Variants for A852 and A86B

Glyph ID	Glyph Description	Variant Selector	Simple Glyph Variant of
0003	hphal. ee second isolate form	VS1	A852 Nominal Form
0005	hphal. ee second initial form	VS1	Presentation Form 0004
0008	hphal. ee second medial form	VS1	Presentation Form 0007
000A	hphal. ee second final form	VS1	Presentation Form 0009
000C	hphal. ee second reversed final form	VS3	Presentation Form 000B
0034	hphal. han wa second isolate-final form	VS1	A86B Nominal Form
0036	hphal. han wa second initial-medial form	VS1	Presentation Form 0035
004F	hphal. han half ya second medial form	VS1	Presentation Form 004E
0050	hphal. han half ya second final form	VS1	A87A Nominal Form

There is no semantic or phonetic distinction between the use of the "nominal" glyph for A852,A86B or A87A and their corresponding variant glyph. The distinction between the "nominal" and variant glyph forms of these two letters is purely aesthetic, and the choice of which form to use should be a font consideration, not an encoding issue.

The encoding of variant glyph forms for these three particular letters appears to be arbitrary, given that several other Phags-pa letters also have distinctive glyph variants (see N2622 Table 2). Moreover, the letter **E** [A852] occurs in more than two distinct variant glyph forms (see N2719 pages 7-8, where various examples are provided, and it is noted that Professor Junast identifies three main glyph forms of the letter), and it is thus also strangely arbitrary to propose the encoding of only two of the glyph variants. Regardless of this, the fact remains that there is absolutely no reason to encode simple glyph variants of these or other letters, either as separately encoded characters or by means of variation selectors. To do so would turn ISO/IEC 10646 from being a character encoding standard into a glyph encoding standard.

2.6 HPHAGS-PA JOINER [A844]

N2745 proposes a "joiner" character corresponding to the Mongolian Nirugu [180A] to represent the short extender used to physically join adjacent characters in a syllable unit. The join between letters within a syllable unit normally occurs on the right-hand side, but may occur on the left-hand side after a reversed letter (**TTA**, **TTHA**, **DDA** and **NNA**). However, after the letter **O** the join occurs in the centre. Also, in Tibetan-style Phags-pa texts the joint may sometimes occur along the central axis after any letter. Examples of the right-hand and left-hand join between the letters **THA** and **-A** and between the letters **TTHA** and **-A** in the Juyong Guan inscriptions are shown below (the join is the short extender to the letter **THA**  or **TTHA** that joins it to the letter **-A** below) :

The Joiner character A844 is proposed to be used to represent the normal right-hand join, whereas Joiner+VS1 <A844, FE00> is proposed to be used to represent the less common left-hand join. Why is such a "joiner" character required ? According to N2745, "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" (N2745-1 page 6). I do not believe that any of these are valid reasons to encode this "joiner" as a distinct character.

A. Facilitate the connection of various letters

The implication of N2745 seems to be that the joiner character will need to be encoded between all letters that ligate together along the right-hand or left-hand axis within a syllable unit. Thus the two syllables shown above would need to be encoded as <A861, A844, A86E> and <A87C, A844, FE00, A86E>. This would take the burden of ligating adjacent letters away from the rendering system, and place it in the hands of the user, who would have to enter this extra character after every non-terminal letter within a syllable unit other that the letter **O**.

The MONGOLIAN NIRUGU [180A] on which A844 is modelled is an artefact of mechanical typesetting, where it is necessary to typeset the physical join between Mongolian letters. However, as the ligature between Mongolian letters is achieved automatically with "smart font" technology such as OpenType, there is normally no need to include the nirugu character [180A] in the code sequence for Unicode Mongolian text. Likewise, for Phags-pa, it would be expected with modern font technology and sophisticated rendering systems that the user would not need to have to manually enter A844 after every non-terminal letter. Instead, it would be expected that the appropriate ligatures between adjacent Phags-pa letters would be achieved automatically by the rendering system. A844 should only be needed for "dumb" rendering systems which are unable to automatically ligate adjacent letters. However, not only would "smart" rendering system be required to deal with positional and contextual variants anyway, but **N2652R** explicitly states that characters should not be accepted for encoding merely to "overcome short-term deficiency of rendering technology". Furthermore, if A844 were to be accepted, users with "dumb" rendering systems might include A844 after almost every non-terminal letter, whereas users with "smart" rendering systems might never use A844 at all, with the result that the same Phags-pa texts would be encoded differently, which would greatly hinder basic data processing activities such as searching and collating.

For Unicode Mongolian the Nirugu [180A] is only strictly required when a long join between two letters is needed to separate the constituent elements of certain compound words. However, in Phags-pa texts the join between adjacent letters never needs to be artificially lengthened, so A844 would not be needed for this usage either.

B. Select the joined parts

This seems to imply that users might want to render the join between Phags-pa letters in isolation. I cannot imagine why anyone would want to do this, and I do not believe that a separate character should be introduced just in case anyone did want to do so.

It might also be pointed out that the fact that the so-called "joiner" character does not occur in isolation is a fairly good indicator that it is not actually a character, but merely a glyph element.

C. Decrease the number of variant presentation glyphs

The number of presentation glyphs is irrelevant, as that is a font issue not an encoding issue. Notwithstanding this claim, out of the ninety-one presentation glyphs given in N2745, no less than forty-eight of them are actually glyphs representing a given letter plus a joiner (see Table 2-9-1). Each of these presentation forms should, according to the N2745 encoding model, be represented as two characters <A8XX, A844>. The fact that these forty-eight presentation forms are included in N2745 indicates an ambiguity over the usage of A844. Is it required to be manually inserted into the code sequence ? Or should the rendering system select the appropriate "Letter+Joiner" presentation form in the initial or medial position ? If the former, why are the corresponding presentation forms included ? If the latter, what need is there to encode a separate "joiner" character ?

In conclusion, I do not believe that there is any need to encode a separate "joiner" character for Phags-pa. The Mongolian Nirugu [180A], which is the analogy for encoding a Phags-pa "joiner" character, is not required for achieving the actual join between Mongolian letters, and is only strictly required in situations that are not relevant to Phags-pa (to create an extra long join between letters in certain circumstances). Thus, by analogy with Mongolian, A844 is not actually required at all. Introducing a redundant "joiner" character would just cause confusion amongst end users about when, if and how it should be used.

2.7 Subjoined Letter RA and Superfixed Letter RA

The two characters PHAGS-PA SUBJOINED LETTER RA [A868] and PHAGS-PA SUPERFIXED LETTER RA [A869] proposed in N2622 are not included in the character repertoire proposed in N2745. Instead it is proposed that these forms of the letter RA be treated as positional or contextual variants. According to N2745 Superfixed Letter RA would be represented as RA+VS1 <A870, FE00>; whilst Subjoined Letter RA would be represented as an automatically selected positional form of the letter RA in medial position within a syllable (e.g. in the mantric syllable "bh-ruṁ"), and as RA+VS1 <A870, FE00> when carrying an inherent vowel [a] in the final position within a syllable (e.g. "ba dzra" = Sanskrit "vajra"). Whilst this is a workable solution, I believe it is not the best solution.

Consistency with Subjoined WA and YA

The Phags-pa subjoined letters WA  , YA  and RA  are all derived from the Tibetan script where these three letters occur in distinct subjoined forms after a consonant. Two of these subjoined forms, YA and WA, are used in writing Chinese, and were are included in the original set of forty-one Phags-pa letters (see Table 2-1-1 nos.40-41). They are consequently encoded for separate encoding in N2745 as HPHAGS-PA LETTER HAN HALF [A879] and HPHAGS-PA LETTER HAN HALF [187A]. The Phags-pa subjoined letter RA is used in Sanskrit texts, notably the Juyong Guan inscriptions, where it occurs with great frequency, together with subjoined WA and subjoined YA. These three special forms of the letters WA, YA and RA have exactly the same status in Tibetan, and are always grouped together in descriptions of the Tibetan script. To encode two of the three corresponding Phags-pa letters separately, but to treat one of them as a variant form of the ordinary letter is inconsistent.

Compatibility with Tibetan

In Tibetan subjoined forms of the letters WA [0FAD], YA [0FB1] and RA [0FB2] are encoded differently from the base forms of the letters WA [0F5D], YA [0F61] and RA [0F62]. Encoding subjoined RA as a distinct character would simplify transcoding between Tibetan and Phags-pa scripts.

Simplicity

The superfixed form of the letter RA is used exclusively in writing Tibetan words, and occurs rarely in 13th and 14th century Phags-pa texts (but see N2622 Example 6 for one example). In the Tibetan encoding model the superfixed "head form" of the letter RA is not encoded separately, but the rendering system automatically selects the ordinary form or head form of the letter contextually. In Phags-pa it is not possible to contextually determine whether an initial letter RA is a head form or ordinary form of the letter when it carries an inherent [a] vowel (see N2719 Point 2 for details), and so I proposed that the superfixed head form of the letter be encoded separately. I agree that using variation selectors would be an acceptable alternative, but it would be simpler for end users if the letter were encoded separately.

2.8 HPHAGS-PA LETTERS HAN YA/SHA/HA/FHA [A87F..A882]

The four proposed characters HPHAGS-PA LETTER HAN YA [A87F], HPHAGS-PA LETTER HAN SHA [A880], HPHAGS-PA LETTER HAN HA [A881] and HPHAGS-PA LETTER HAN FHA [A882] correspond to the four Phags-pa standardized variants proposed in N2622 (see section 9 of that document for details). Note that the "FHA" in the proposed name "HPHAGS-PA LETTER HAN FHA" seems to have been chosen in order to differentiate it from "HPHAGS-PA LETTER HAN FA" [A877], and it is not normally transliterated as "fha".

These graphically distinct forms of the letters only occur contrastively in a single text, Menggu Ziyun 蒙古字韻 (a rhyming dictionary of Chinese with Phags-pa spellings, revised and edited by Zhu Zongwen 朱宗文 in 1308), where an artificial graphic distinction is used to indicate historical phonetic differences between Chinese words which were pronounced the same in 14th-century Northern Chinese (i.e. Old Mandarin), and which were written the same in other Phags-pa texts and inscriptions. Thus, for example, the two Chinese ideographs 施 **shi** and 時 **shi** are both spelled as **shi** in the Phags-pa script,

but in Menggu Ziyun the initial letter **SHA** in each word is written differently in order to distinguish the fact that historically the two words had different initial consonants (the former [ɕ] and the latter [ʑ]) This distinction was important for the author of Menggu Ziyun as rhyming dictionaries of Chinese traditionally ordered ideographs within rhyme categories according to the "36 initials" of Chinese spoken during the Tang and Song dynasties, and as Menggu Ziyun continued this tradition it needed to indicate the relationship between Yuan dynasty Phags-pa spelling and the earlier "36 initials".

Whilst I am pleased that N2745 recognises the need to distinguish these variant forms from the normal forms of the letters **YA**, **SHA**, **HA** and **FA**, I believe that treating them as standardized variants rather than encoding them as distinct characters is a preferable solution, for the following reasons :

1. The variant forms are not distinct characters

The primary reason for not encoding the letters YA, SHA, HA and FA as distinct characters is that they are not distinct characters in their own right, but are only variant forms of the same character that are used to indicate semantic distinctions in one particular text only.

2. Standardized Variants avoid accidental misuse

The only occasions that these four variant forms of the letters YA, SHA, HA and FA should be used is in quoting from entries in Menggu Ziyun, and they should never be used in encoding any other Phags-pa texts, even if a particular text does use a form of one of these letters that is closer to the variant glyph form than the nominal glyph form (this is because, in texts other than Menggu Ziyun, such glyph usage is purely accidental, and carries no semantic significance whatsoever). By providing two sets of characters for the letters YA, SHA, HA and FA, each graphically very similar one to another, the end user may well be confused as to which particular character to use to encode ordinary Phags-pa texts. Especially as the names of the standard letters YA, SHA, HA and FA [A86F, A872, A874 and A877] do not contain the word "HAN", whereas the names of the variant forms of these letters [A87F..A882] do all contain the word "HAN", many users might mistakenly believe that A86F, A872, A874 and A877 should be used for encoding YA, SHA, HA and FA Mongolian Phags-pa texts, whilst A87F..A882 should be used for encoding YA, SHA, HA and FA in Chinese Phags-pa texts. This is categorically not the case.

By treating the variant letters as standardized variants, ordinary users will not accidentally misuse the variant forms of the letters YA, SHA, HA and FA for texts other than Menggu Ziyun. On the other hand, if the variant forms of the letters are encoded separately, as proposed in N2745, there is a high degree of probability that some end users will mistakenly use the variant forms of these letters for encoding texts other than Menggu Ziyun, which would cause considerable problems for searching, collation, etc.

3. Variation Selector is easy to strip out or ignore

For many purposes the variant forms of the letters YA, SHA, HA and FA should be treated as if they were the standard forms of the letters YA, SHA, HA and FA. For example, searching for Phags-pa **shi** should return both **shi** with the standard form of the letter **SHA** and **shi** with the variant form of the letter **SHA**.By encoding the variant forms of these letters as standardized variants it is possible for processes to treat the standard and variant forms of the letters identically by simply ignoring the Variation Selector [FE00].

Another good example of where it would be advantageous to encode these variant letters as standardized variants would be in a list of Chinese words in Phags-pa spelling that have been derived from Menggu Ziyun. In order to convert the peculiar spellings of Chinese words in Menggu Ziyun (reflecting historic phonetic differences by means of variant letters) into a form that reflected actual Phags-pa spellings used in monumental inscriptions all that would be necessary would be to strip out all occurrences of the variation selector FE00.

Although I would rather see these four variant letters encoded as standardized variants, if they were to be encoded as separate characters in there own right, I would strongly suggest that their names be changed to something like "PHAGS-PA LETTER MENGGU ZIYUN YA/SHA/HA/FA", and that a note indicating their limited intended scope of use be added to the code charts.

2.9 Presentation Forms

N2745-3 Section III lists ninety-one presentation forms, as listed in Table 2-9-1.

Table 2-9-1 : Proposed Presentation Forms

Presentation Glyph	Description	Category	Comment
0000	hpha . left joiner 1	C	After A87B..A87E
0001	hphal. a initial form	A	A850 plus A844
0002	hphal. e medial form	A	A851 plus A844
0003	hphal. ee second isolate form	D	Glyph variant of A852 (nominal form)
0004	hphal. ee first initial form	B	A852 initial form
0005	hphal. ee second initial form	D	Glyph variant of 0004
0006	hphal. ee third initial form	E	A850 plus A852
0007	hphal. ee first medial form	B	A852 medial form
0008	hphal. ee second medial form	D	Glyph variant of 0007
0009	hphal. ee first final form	B	A852 final form
000A	hphal. ee second final form	D	Glyph variant of 0009
000B	hphal. ee first reversed final form	C	After A87B..A87E
000C	hphal. ee second reversed final form	D	Glyph variant of 000B
000D	hphal. i initial form	B	A853 initial form
000E	hphal. i medial form	B	A853 medial form
000F	hphal. i final form	B	A853 final form
0010	hphal. i reversed final form	C	After A87B..A87E
0011	hphal. o first initial form	B	A854 initial form
0012	hphal. o second initial form	E	A850 plus A854
0013	hphal. o medial form	B	A854 medial form
0014	hphal. o final form	B	A854 final form
0015	hphal. u first initial form	B	A855 initial form
0016	hphal. u second initial form	E	A850 plus A855
0017	hphal. u third initial form	A	A855 plus A844
0018	hphal. u medial form	B	A855 medial form
0019	hphal. u final form	B	A855 final form
001A	hphal. u reversed final form	C	After A87B..A87E
001B	hphal. oe initial form	E	A850 plus A856 (A851, A854)
001C	hphal. oe first medial form	B	A856 medial form
001D	hphal. oe first final form	B	A856 final form
001E	hphal. ue initial form	E	A850 plus A857 (A851, A855)
001F	hphal. ue first medial form	B	A857 medial form
0020	hphal. ue first final form	B	A857 final form
0021	hphal. ka initial-medial form	A	A858 plus A844
0022	hphal. kha initial-medial form	A	A859 plus A844
0023	hphal. ga initial-medial form	A	A85A plus A844
0024	hphal. nga initial-medial form	A	A85B plus A844
0025	hphal. ca initial-medial form	A	A85C plus A844
0026	hphal. cha initial-medial form	A	A85D plus A844

0027	hphal. ja initial-medial form	A	A85E plus A844
0028	hphal. nya initial-medial form	A	A85F plus A844
0029	hphal. ta first initial-medial form	A	A860 plus A844
002A	hphal. tha initial-medial form	A	A861 plus A844
002B	hphal. da initial-medial form	A	A862 plus A844
002C	hphal. na initial-medial form	A	A863 plus A844
002D	hphal. pa initial-medial form	A	A864 plus A844
002E	hphal. pha initial-medial form	A	A865 plus A844
002F	hphal. ba initial-medial form	A	A866 plus A844
0030	hphal. ma initial-media form	A	A867 plus A844
0031	hphal. han tsa initial-media form	A	A868 plus A844
0032	hphal. han tsha initial-media form	A	A869 plus A844
0033	hphal. han dza first initial-medial form	A	A86A plus A844
0034	hphal. han wa second isolate-final form	D	Glyph variant of A86B (nominal form)
0035	hphal. han wa first initial-medial form	A	A86B plus A844
0036	hphal. han wa second initial-medial form	D	Glyph variant of 0035
0037	hphal. han nya second initial-medial form	A	A86C plus A844
0038	hphal. han za second initial-medial form	A	A86D plus A844
0039	hphal. minuscule a initial-medial form	A	A86E plus A844
003A	hphal. minuscule a reversed initial-medial form	C	After A87B..A87E
003B	hphal. ya initial-medial form	A	A86F plus A844
003C	hphal. ra first initial-medial form	A	A870 plus A844
003D	hphal. ra second initial form	F	RA superfixed form
003E	hphal. ra second medial form	F	RA subjoined form
003F	hphal. ra second final form	F	RA subjoined form
0040	hphal. la initial-medial form	A	A871 plus A844
0041	hphal. sha initial-medial form	A	A872 plus A844
0042	hphal. sa initial-medial form	A	A873 plus A844
0043	hphal. ha initial-medial form	A	A874 plus A844
0044	hphal. ha reversed medial form	C	After A87B..A87E
0045	hphal. ha reversed final form	C	After A87B..A87E
0046	hphal. qha initial-medial form	A	A875 plus A844
0047	hphal. han qwa initial-medial form	A	A876 plus A844
0048	hphal. han fa initial-medial form	A	A877 plus A844
0049	hphal. qa initial-medial form	A	A878 plus A844
004A	hphal. han half u initial form	E	A850 plus A879
004B	hphal. han half u medial form	A	A879 plus A844
004C	hphal. han half u final form	A	A879 plus A844
004D	hphal. han half u reversed final form	C	After A87B..A87E
004E	hphal. han half ya first medial form	A	A87A plus A844
004F	hphal. han half ya second medial form	D	Glyph variant of 004E
0050	hphal. han half ya second final form	D	Glyph variant of A87A (nominal form)
0051	hphal. han half ya reversed final form	C	After A87B..A87E
0052	hphal. sanskrit tta initial-medial form	A	A87B plus A844
0053	hphal. sanakrit ttha first initial-medial form	A	A87C plus A844
0054	hphal. sanakrit ttha second initial-medial form	A	A87C plus A844

0055	hphal. sanakrit dda initial-medial form	A	A87D plus A844
0056	hphal. sanakrit nna initial-medial form	A	A87E plus A844
0057	hphal. han ya initial-medial form	A	A87F plus A844
0058	hphal. han sha initial-medial form	A	A880 plus A844
0059	hphal. han ha initial-medial form	A	A881 plus A844
005A	hphal. han fha initial-medial form	A	A882 plus A844

A. Initial/Medial Forms with Joiner [48 presentation forms]

These presentation forms are the same as the "nominal" form of the corresponding letter, but with a slight downward extension on the left or right side of the letter corresponding to the "joiner" character A844. These represent the initial or medial forms of letters which ligate to the following letter by means of a short vertical extension. These presentation forms are all selected by application of ZWJ [200D] after a Phags-pa letter.

These presentation forms are not explicitly listed in N2622 as it is difficult to imagine circumstances when anybody would want to display an initial form with an extender in isolation, although, as stated in the last paragraph of Section 5 of N2622, I recognise that Phags-pa fonts may need to render initial and medial forms of some letters with a short extender in order to ligate to the following letter. That ZWJ should produce such presentation forms is implicit in N2622.

It may be noted that one of the reasons given in N2745 for the encoding of a separate "joiner" character A844 is to "decrease the number of variant presentation glyphs", and, as discussed in Section 2-6 above, N2745 seems to imply that separate initial and medial presentation forms such as these would not be required as the extender used to ligate initial/medial forms of letters to the following letter would need to be hard-coded with A844. Whilst I strongly disagree with the need to hard-code a "joiner" character after every initial or medial occurrence of almost all Phags-pa letters, and agree that presentation forms such as these are implicitly required, the conflict between the encoding of a "joiner" character A844 on the one hand and the definition of presentation forms representing letters plus joiner on the other hand does seem to be indicative of the instability of the encoding model proposed in N2745.

B. Positional Vowel Forms [16 presentation forms]

These presentation forms represent the initial, medial and final forms of the vowel letters **E**, **I**, **O** and**U**, as well as the medial and final forms of the precomposed vowel letters **OE** and **UE**. N2745 refers to these as "positional variants". The "nominal" forms of these letters represent the isolate form. The positional forms of the letters **E**, **I**, **O** and **U** correspond to the positional forms for vowels listed in N2622 Section 7 Table 4. The correct positional vowel form would be selected contextually by the rendering engine, although ZWJ and ZWNJ could be used to override the contextual form. This is in accord with the model proposed in N2622.

C. Reversed Variants [8 presentation forms]

These presentation forms represent the reversed forms of letters and the proposed "joiner" character that occur after the Sanskrit letters **TTA**, **TTHA**, **DDA** and **NNA** [A87B..A87E]. These are contextual variants that should be achieved automatically after an initial letter **TTA**, **TTHA**, **DDA** or **NNA**. N2745 refers to these as "postpositive variants". These presentation forms correspond to the contextual variants given in N2622 Section 8 (see especially Table 5), with the following exceptions :

- I do not believe there is any reason to encode a distinct "joiner" character (see Section 2-6 above), and so N2622 does not recognise the need for a reversed joiner presentation form.
- Presentation Glyph 003A ("hphal. minuscule a reversed initial-medial form") is missing from N2622 Table 5. This is a mistake by me. Although the majority of instances of the letter **-A** after **TTA** or **TTHA** in the Juyong Guan inscriptions are not reversed (seven out of eight cases), a reversed letter **-A** does occur once, but this single occurrence was accidentally overlooked by myself. I agree that there should be a method of selecting the reversed form of the letter **-A**, and I think this example gives added weight to the need to encode a PHAGS-PA FREE VARIATION SELECTOR, as mooted on page 12 of N2719.
- Presentation Glyph 004D ("hphal. han half u reversed final form") is missing from N2622 Table 5. The Juyong Guan inscriptions do not appear to use a reversed **Subscript WA**, and I have not seen reversed forms of letters used in any other Phags-pa text. Clarification on the occurrence of this presentation form would be welcome.

D. Simple Glyph Variants [9 presentation forms]

These presentation forms represent simple glyph variants of the letters **E**, **WA** and **Half YA** (Subjoined YA) that N2745 proposes to encode by means of variation selectors. N2745 refers to these as "free variants". As discussed in Section 2-5 above, there is absolutely no justification for encoding simple glyph variants of these letters.

E. Null Consonant Plus Vowel or Semi-Vowel [6 presentation forms]

These presentation forms correspond to the null consonant A850 followed by a vowel. As discussed in Section 2-2 above, the null consonant preceding a vowel letter should not be considered as part of the initial form of a vowel. These presentation forms are in reality precomposed sequences of null consonant plus vowel that are mapped to the initial form of the vowel.

F. Subjoined and Superfixed Letter RA [3 presentation forms]

These presentation forms correspond to the separately encoded letters **Superfixed RA** and **Subjoined RA** that are proposed in N2622. As discussed in Section 2-7 above, I believe that encoding these letters as distinct characters rather than as presentation forms selectable by means of variation selectors is a preferable encoding solution.

2.10 Variation Selectors

N2745 proposes to use three variation selectors, VS-1..VS-2 [FE00.FE02], to select variant forms of characters (see N2745-3 Section IV "Reference Table for HPhags-pa Script" for details). There are certain issues with this, as discussed below.

A. Variation Selector or Free Variation Selector ?

N2745 proposes to use standard variation selectors for Phags-pa in situations that are analogous to the Mongolian use of Free Variation Selectors [180B..180D]. Although the difference between standard variation selectors and Free Variation Selectors is not clearly defined, it is my feeling that there is a real difference in intended usage.

Variation Selectors would seem to me to be intended to select a single fixed glyph form of a character that is semantically different from the standard glyph form of that character only in a particular context. It is this concept of what a standardized variant should be that informed my suggestion to encode the variant letters SHA, YA, HA and Fa used in Menggu Ziyun as standardized variants.

On the other hand, it seems to me that FVSs have a less rigid function, and are not intended to be used to select a particular glyph variant in isolation, but rather are intended to override the default selection of a contextual glyph for a particular character by the rendering system. Thus in Mongolian FVSs are used to force the selection of an unexpected positional form of a letter, or to select a particular form of letter that is used for writing foreign words. In this way, a particular FVS may select a number of different glyph forms depending upon context. This is in contrast to standard variation selectors which should only select one particular glyph for any base character.

The usage of variation selectors in N2745 is clearly akin to the usage of Mongolian FVSs rather than ordinary variation selectors, with individual variation selectors used to select up to four different glyph forms for the same base character depending on context. For example, **EE plus VS-1** <A852, FE00< selects presentation glyphs 0003, 0005, 0008 or 000A depending on positional context. As a further example, **RA plus VS-1** <A870, FE00< selects the superfixed "head form" of the letter RA in a syllable-initial position, but selects the subjoined form of the letter RA in a syllable-final position. This multiplicity of glyph selection for a single variation selector seems to me to be at variance with the expected behaviour of standard variation selectors, and I therefore think that such usage would be better assigned to Phags-pa specific Free Variation Selector characters.

B. Number of Variation Selectors

N2745 proposes to use up to three variations selectors for a single character (e.g. A852). These variation selectors are used for the following purposes :

1. To select simple glyph variants (presentation glyphs 0003, 0005, 0008, 000A, 000C, 0034, 0036, 004F and 0050).
2. To select precomposed sequences of null consonant followed by a vowel or semi-vowel (presentation glyphs 0006, 0012, 0016, 001B, 001E and 004A).
3. To select reversed glyph forms (presentation glyphs 0000, 000B, 0010, 001A, 003A, 0044, 0045, 004D and 0051)
4. To select the superfixed and subjoined forms of the letter RA (presentation glyphs 003D and 003F)
5. To select the initial form of the letter **U** with joiner at the side rather than in the middle (presentation glyph 0017)
6. To select the initial form of the letter **TTHA** with joiner at the right side rather than on the left side (presentation glyph 0054)

I have suggested in N2622 that how the ligature between adjacent Phags-pa letters is made is matter for the font designer, and so I do not subscribe to the belief that variation selectors should be used to select variant ways of ligating adjacent Phags-pa letters, as is the case for presentation forms 0017 and 0054. In Tibetan Phags-pa texts the "joiner" between letters is frequently along a central axis rather than down the right-hand side, which according to the N2745 model would require many more variation sequences than are already defined in the document. I maintain that although the join should normally be made along the right, or along the left for reversed letters, a font designer should be free to make the join as he feels fit. In the Juyong Guan inscriptions the letter **-A** is ligated to a preceding letter **TTHA** on the left in four cases and on the right in two cases. According to N2745 the cases where the join is on the right should be selected by means of a variation selector (either **TTHA plus VS-1** or **TTHA plus Joiner plus VS-1** depending upon your interpretation of N2745). I think that this situation is analogous to the different ways that Latin letters can be connected to each other in cursive fonts, and is not a matter for encoding, but for should be a matter of choice for the individual font.

If we discount the uses of variation selectors to select simple glyph variants, precomposed sequences of null consonant followed by vowel or semi-vowel, and superfixed/subjoined forms of the letter RA, as I have already argued that these particular aspects of N2745 should not be accepted, then all we are left with are the use of variation selectors to select reversed forms of letters. I would suggest that this is the only valid use of variation selectors in N2745. As any Phags-pa letter has at most one reversed form (most have none), then only one variation selector per character would ever be required, and that would always be VS-1. Thus, I believe, only one variation selector is strictly needed for Phags-pa purposes, not three as proposed in N2745.

The usage of a variation selector (VS-1) to select reversed forms of letters corresponds directly the suggested use of a "Contextual Variant Override" character in N2622 (see Section 8 of that document) and the suggested use of a Phags-pa Free Variation Selector character in N2719. My original proposal did not actually propose the encoding of a specific control character for selecting reversed forms of letters, but merely raised the point that such a character was required, as I thought that the UTC and WG2 were more qualified than myself to decide exactly what character would be most appropriate for this purpose. However, since this issue does not appear to have been addressed, I would now suggest that a Phags-pa Free Variation Selector be added to the Phags-pa block, as per N2719 (see page 12 of that document). This Phags-pa FVS would correspond to the use of VS-1 in N2745 to select presentation glyphs 000B, 0010, 001A, 003A, 0044, 0045, 004D and 0051.

2.11 Punctuation Marks [A840..A843]

The great majority of Phags-pa texts dating from the 13th and 14th centuries do not use punctuation marks of any kind. When punctuation marks are used, they are borrowed from Chinese or Mongolian. Thus the Juyong Guan inscriptions use the Mongolian punctuation marks ᠂ [1802 : MONGOLIAN COMMA], ᠃ [1803 : MONGOLIAN FULL STOP] and ᠅ [1805 : MONGOLIAN FOUR DOTS] (see N2622 Example 3); whereas the fragments of the printed edition of the Mongolian translation of the Subhāṣitaratnanidhi use a small circle 。 [3002 : IDEOGRAPHIC FULL STOP] as a punctuation mark.

N2745 proposes the encoding of separate characters to represent the usage of these borrowed Chinese and Mongolian punctuation marks in Phags-pa texts [A840..A843]. I do not believe there is any need to duplicate the encoding of these already encoded punctuation marks with Phags-pa clones, especially as the use of such punctuation marks in Phags-pa texts is the exception rather than the rule.

3002 [IDEOGRAPHIC FULL STOP] is already classified as "Common" in "Scripts.txt", and would thus be appropriate to use for Phags-pa texts (it is, for example, part of the official punctuation set designated for use with the Yi script). Although the

Mongolian punctuation marks are currently classified as "Mongolian" only in "Scripts.txt", as Phags-pa shares the same directionality as Mongolian, and as there is no difference in glyph shape between the marks used in Mongolian and Phags-pa texts, I think that it would be appropriate to simply use 1802, 1803 and 1805 to represent the occasional usage of these marks in Phags-pa texts, and change their designation from "Mongolian" to "Common".

2.12 Tibetan Phags-pa Characters

N2745 does not include any of the special Tibetan-usage Phags-pa punctuation marks proposed in N2622. I have examined quite a few examples of modern Tibetan Phags-pa texts, both traditional woodblock prints and examples in books on Tibetan calligraphy published in the People's Republic of China, and Tibetan usage of the Phags-pa script does not appear to be problematic in any way. Whilst the Tibetan style of Phags-pa letters are distinct from the style of Phags-pa letters used during 13th and 14th centuries, this difference can be dealt with easily by means of dedicated Tibetan-style Phags-pa fonts, and in no way affects the encoding of the script.

Tibetan Phags-pa texts do not appear to use any additional letters, but they do use a number of punctuation marks that are based on Tibetan punctuation marks, that is two types of "head marks" which are used to mark the start of a Phags-pa passage, and "shad" and "double shad" marks that are used to mark the end of sections of text. Whilst these additional punctuation marks are mostly found in 19th and 20th century Tibetan Phags-pa texts, there is at least one example of a 14th-century usage of the Phags-pa "shad" mark (see N2622 Example 6).

It should be noted that although these marks correspond to already-encoded Tibetan punctuation marks, it would not be possible to use the corresponding Tibetan marks in Phags-pa texts as the Phags-pa glyphs for these punctuation marks are distinct from the corresponding Tibetan glyphs. Furthermore, as Tibetan is written horizontally, if a Tibetan head mark were to be used in vertical Phags-pa text the glyph would be rotated incorrectly. For these reasons, it is necessary to encode separate Phags-pa versions of these Tibetan punctuation marks. This is in marked contrast to the case of Mongolian and Chinese punctuation marks that are occasionally used in Phags-pa texts. In these cases there is no difference in glyph shape between the punctuation marks used in Chinese or Mongolian texts and the same punctuation marks used in Phags-pa texts; and as Mongolian and Chinese are both written vertically, the problem of glyph rotation does not occur.

As the "head mark" and "shad" punctuation marks are used with great frequency in Tibetan Phags-pa texts, if these marks were not to be encoded at this time it would effectively mean that Tibetan Phags-pa texts could not be represented in Unicode. On the other hand, if the Tibetan Phags-pa punctuation marks proposed in N2622 were to be accepted for encoding now, then there would be no obstacle to encoding Tibetan Phags-pa texts in Unicode. As Tibetan usage of the Phags-pa script is the only living usage of the script, it would seem ridiculous not to facilitate Tibetan usage by encoding these extra characters at the earliest opportunity.

2.13 Character Names

Some of the proposed character names in N2745 include a linguistic qualifier :

- No linguistic qualifier for letters that are used in writing Mongolian (e.g. HPHAGS-PA LETTER KHA)
- The qualifier "HAN" (i.e. Chinese) is used for letters that are used for writing Chinese (e.g. HPHAGS-PA LETTER HAN NYA)
- The qualifier "SANSKRIT" is used for letters that are used for writing Sanskrit (e.g. HPHAGS-PA LETTER SANSKRIT TTA)
- The qualifier "TIBETAN" is used for letters that are used for writing Tibetan (e.g. HPHAGS-PA TIBETAN ANUSVARA)

Linguistic qualifiers such as these are expressly noted as being inappropriate in Rule 9 of the "Character-naming guidelines" (**N2652R** Annex L) :

> In principle when a character of a given script is used in more than one language, no language name is specified. Exceptions are tolerated where an ambiguity would otherwise result.

I would make the following points against the inclusion of any linguistic qualifiers in the names of any Phags-pa characters :

1. No ambiguity would result from the omission of the linguistic qualifiers of any of the letters, except for A87F..A881 [HPHAGS-PA LETTER HAN YA/SHA/HA]. For these three letters the qualifier "HAN" is inappropriate anyway, as the letters A86F [HPHAGS-PA LETTER YA], A872 [HPHAGS-PA LETTER SHA] and A874 [HPHAGS-PA LETTER HA] are the normal letters for writing Chinese, and A87F..A881 are only used in a single text as specialized variants of A86F, A872 and A874.
2. The HAN and SANSKRIT qualifiers are probably modelled after the "TODO", "MANCHU" and "SIBE" qualifiers used in naming letters in the Mongolian block. This analogy is false, as Mongolian, Todo, Manchu and Sibe are four distinct scripts belonging to the Mongolian script family, which as they share many letters have been unified in a single "Mongolian" block. Thus for Mongolian, the "TODO", "MANCHU" and "SIBE" qualifiers represent script usage not linguistic usage (even though Manchu and Sibe are also languages). On the other hand, Phags-pa is a single script used for writing Mongolian, Chinese and other languages. There is no separate "Mongolian Phags-pa" script or "Chinese Phags-pa" script; just Phags-pa.
3. The absence of the qualifier "HAN" is misleading, as all except for three of the letters without the HAN qualifier are also used for writing Chinese, and many of the letters are also used for writing languages such as Uighur and Tibetan.
4. The presence of the qualifier "HAN" is misleading, as many of these letters are also used for writing other languages, such as Sanskrit and Tibetan (indeed, there is not a single "HAN" letter that is used exclusively for writing Chinese).
5. The qualifier "TIBETAN" for A845 [HPHAGS-PA TIBETAN ANUSVARA] is incorrect, as Tibetan Phags-pa texts actually use the Phags-pa letter **MA** instead of a special Phags-pa anusvara letter. In fact the only extensive texts that the Phags-pa "anusvara" is used in are the Sanskrit texts at Juyong Guan.
6. Assignment of linguistic usage to a particular letter may be problematic. For example, although the letters A869 [HPHAGS-PA LETTER HAN TSHA] and A86B [HPHAGS-PA LETTER HAN WA] do not occur in native Mongolian words, they do occur in Mongolian loan words from Chinese, Sanskrit and other languages (e.g. the letter **TSHA** or **c‘**) is used in writing the common Mongolian word **šac‘in** "religion").

I would suggest that linguistic usage is better relegated to the code chart notes, as was proposed in N2719, where more complete linguistic usage for each letter is enumerated.

2.14 Reference Glyphs

As is noted in N2745-1 Section II.9 "Style of Script" (pp.7-8), there are various styles of Phags-pa script. In particular, monumental inscriptions tend to use broader strokes, whilst printed and manuscript texts end to use thinner strokes and cleaner glyph outlines. The choice of font in N2745 is explained thus :

> 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).

Whilst the style of Phags-pa letters used in the monumental inscriptions of Khubilai Khan’s edicts are undoubtedly very beautiful, that does not necessarily make them the most suitable style of letters to use in the code charts of an encoding standard.

The font style created specifically for use in N2622 is modelled after the style of Phags-pa letters used in 14th-century printed texts such as Baijiaxing 百家姓 [The Hundred Chinese Family Names], and emphasises clarity and simplicity of stroke line. This font style may not be as elegant as the style used in N2745, but it is a recognised Phags-pa font style — indeed it may be noted that the "fine style" Phags-pa glyphs listed on page 7 of N2745-1 bear an uncanny resemblance to the glyphs used in N2622. Moreover, the font face used in N2622 is typical of the style of Phags-pa lettering used in modern academic books and articles.

It should be remembered that the purpose of the code charts is to facilitate recognition of a given character, and I believe that the font face used in N2622 achieves this aim extremely well :

> Each character in these code charts is shown with a representative glyph. A representative glyph is not a prescriptive form of the character, but one that enables recognition of the intended character to a knowledgeable user and facilitates lookup of the character in the code charts. In many cases, there are more or less well-established alternative glyphic representations for the same character.
>
> The Unicode Standard, Version 4.0 Section 16.1

3. Bibliography

3.1 Phags-pa Script

Chen Liming 陈历明. "Guangdong Chao'an chutu Yuandai tongyin" 广东潮安出土元代铜印. In Wen Wu 文物 1984.8 : 39.

Cheng Tsai-fa 鄭再發 [Zheng Zaifa]. Menggu Ziyun gen gen Basibazi Youguan de Yunshu 蒙古字韻跟跟八思巴字有關的韻書 [Meng-ku-tzŭ-yün and Some Other Rime Books Connected with *h*Phags-pa Script]. Taipei : Guoli Taiwan Daxue 國立臺灣大學, 1965.

Dao Bu 道布 and Junast 照那斯图. "Henan Dengfeng Shaolinsi chutu de huihushi mengguwen he basibazi shengzhi bei kaoshi" 河南登封少林寺出土的回鹘式蒙古文和八思巴字圣旨碑考释. In Minzu Yuwen 民族语文 1993.5, 1993.6 and 1994.1.

Dong Xuezeng 董学增. "Jilin Jiaohe faxian Yuan basibawen tongyin" 吉林蛟河发现元八思巴文铜印. In Wen Wu 文物 1982.3 : 36.

Hua Yuan 花原. "Jinnian chutu de bufen Jin-Yuan guanyin jianjie" 近年出土的部分金元官印简介. In Wen Wu 文物 1993.12 : 87-88.

Ji Fu 忌浮. "Menggu Ziyun danzi jiaokan buzheng" 《蒙古字韵》单字校勘补正. In Minzu Yuwen 民族语文 1994.2 : 71-75.

Junast 照那斯图. "Basibazi Baijiaxing jiaokan" 八思巴字百家姓校勘. In Minzu Yuwen Lunji 民族语文论集. Beijing : Zhongguo Shehui Kexue Chubanshe 中国社会科学出版社, 1981.

Junast 照那斯图. "Guanyu Xuanzhongsi basisbazi mengguyu shengzhi beike" 关于玄中寺八思巴字蒙古语圣旨碑刻. In Minzu Yuwen 民族语文 1986.6 : 43-46.

Junast 照那斯图 and Yang Naisi 杨耐思. Menggu Ziyun Jiaoben 蒙古字韵校本 [A critical edition of the Phags-pa Chinese Rhyming Dictionary]. Beijing : Minzu Chubanshe 民族出版社, 1987.

Junast 照那斯图. "Basibawen yuanyin zimu zixing wenti shang de liangzhong tixi" 八思巴文元音字母字形问题上的两种体系 [Two series of Phags-pa vowel letter shapes]. In Minzu Yuwen 民族语文 1987.4 : 1-6.

Junast 照那斯图. "Youguan Basiba zimu ė de jige wenti" 有关八思巴字母 ė 的几个问题 [About certain issues relating to the Phags-pa letter ė]. In Minzu Yuwen 民族语文 1988.1 : 1-8.

Junast 照那斯图. "Basibazi zhong de ling shengmu fuhao" 八思巴字中的零声母符号 [The Phags-pa null consonant sign]. In Minzu Yuwen 民族语文 1989.2 : 29-36.

Junast 照那斯图. "Yuandai zhibi basibazi guanyin wenzi kao" 元代纸币八思巴字官印文字考. In Minzu Yuwen 民族语文 1995.6 : 43-48.

Junast 照那斯图 and Hu Haifan 胡海帆. "Linxian Baoyansi liangdao basibazi mengguyu shengzhi" 林县宝严寺两道八思巴字蒙古语圣旨. In Minzu Yuwen 民族语文 1996.3 : 48-54.

Junast 照那斯图. "Dui Jilin Chutu Gudai Guanyin liangfang basibazi guanyin shiwen de dingzheng" 对《吉林出土古代官印》两方八思巴字官印释文的订正. In Wen Wu 文物 1996.3 : 94-95.

Junast 照那斯图. "Basibazi mengguyu longnian shengzhi" 八思巴字蒙古语龙年圣旨. In Minzu Yuwen 民族语文 1996.4 : 45-49.

Junast 照那斯图. "Meng-Yuan shiqi gongting wenshu de yinzhang wenzi" 蒙元时期宫廷文书的印章文字. In Minzu Yuwen 民族语文 1997.3 : 43-50.

Junast 照那斯图. "Shanxi chutu lidai xiyin xuanbian he xubian liangshu zhong ruogan basibazi yinzhang wenzi de dingzheng he bushi" 陕西出土历代玺印选编和续编两书中若干八思巴字印章文字的订正和补释. In Wen Wu 文物 1997.3 : 67-72.

Junast 照那斯图. "Taiwan Gugong cang jifang Yuan basibazi yinwenzi yishi" 台湾故宫藏几方元八思巴字印文字译释. In Minzu Yuwen 民族语文 1997.5 : 41-43.

Junast 照那斯图, Dao Bu 道布 and Liu Zhaohe 刘兆鹤. "Ananda Qinwang basibazi mengguyu manian lingzhi" 阿难答秦王八思巴字蒙古语马年令旨. In Minzu Yuwen 民族语文 1998.3 : 13-17.

Junast 照那斯图. "Huihuwen zimu de Basibazi zhuyin" 回鹘文字母的八思巴字注音 [The Phags-pa phonetic glosses of Uighur letters]. In Minzu Yuwen 民族语文 2000.4 : 50-51.

Ligeti, Louis. "Le Po Kia Sing en Écriture 'Phags-Pa". In Acta Orientalia Vol. VI (1956): 1-52.

Luo Changpei 羅常培 and Cai Meibiao 蔡美彪. Basibazi yu Yuandai Hanyu [Ziliao Huibian] 八思巴字與元代漢語［資料彙編］. Beijing : Kexue Chubanshe 科學出版社, 1959.

Lü Chenglong 吕成龙. "Guanyu Basibazi kuan qinghua ciqi niandai zhi wojian" 关于八思巴字款青花瓷器年代之我见. In Wen Wu 文物 2001.8 : 77-83.

Murata Jiro 村田治郎 (ed.). Chü-Yung-Kuan 居庸關 : The Buddhist Arch of the Fourteenth Century A.D. at the Pass of the Great Wall Northwest of Peking. Kyoto : Kyoto University Faculty of Engineering 京都大學工學部, 1957.

Poppe, Nicholas. The Mongolian Monuments in ḤP'AGS-PA Script. Translated and edited by John R. Krueger. Wiesbaden : Otto Harrassowitz, 1957.

Pozdneyev, Alexei Matveevich (1851-1920). Лекціи по исторіи монгольской литературы, читанныя въ 1896/1897 акад. году [Lectures on the History of Mongolian Literature given during the Academic Year 1896-1897]. St. Petersburg, 1897.

Sheng Ximing 盛熙明. Fashu Kao 法書攷 (1334). In Lianting Shier Zhong 楝亭十二種. Shanghai : Shanghai Gushu Liutongchu 上海古書流通處, 1921.

Sun Jiatan 孙家潭. "Yifang Yuan basibazi taoyin kaoshi" 一方元八思巴字套印考释. In Minzu Yuwen 民族语文 2000.2 : 69-70.

Tao Zongyi 陶宗儀. Shushi Huiyao 書史會要 (1376). Shanghai : Shanghai Shudian 上海書店, 1984.

West, Andrew C. "Overview of the 'Phags script" (comprising a brief history of its rise and fall, and a summary of the extant corpus of texts and inscriptions written in the 'Phags-pa script). http://uk.geocities.com/BabelStone1357/hPhags-pa/Overview.html

West, Andrew C. "Description of the 'Phags-pa script" (including an explanation of the usage of its letters in Mongolian and Chinese). http://uk.geocities.com/BabelStone1357/hPhags-pa/Description.html

West, Andrew C. "The Forty-One Letters of the 'Phags script" (an analysis of the earliest Chinese descriptions of the original forty-one letters of the 'Phags-pa script). http://uk.geocities.com/BabelStone1357/hPhags-pa/41Letters.html

West, Andrew C. "The Uighur language in the 'Phags script" (an analysis of the description of the Old Uighur script given by Tao Zongyi). http://uk.geocities.com/BabelStone1357/hPhags-pa/Uighur.html

West, Andrew C. "The Hundred Chinese Surnames in the 'Phags script" (an annotated edition of the 'Phags-pa version of the "Hundred Chinese Surnames"). http://uk.geocities.com/BabelStone1357/hPhags-pa/Baijiaxing.html

West, Andrew C. "Menggu Ziyun" (an analysis of the 14th-century rhyming dictionary of Chinese with 'Phags-pa spellings). http://uk.geocities.com/BabelStone1357/hPhags-pa/MengguZiyun.html

Yang Huancheng 杨焕成 and Deng Hongli 邓宏理. "Henan Xinyang chutu Yuandai basiba zhuanwen tongyin" 河南沁阳出土元代八思巴篆文铜印. In Wen Wu 文物 1981.6 : 25.

Yao Lei 姚垒 and Song Renfu 宋仁甫. "Henan Yexian chutu Yuandai 'Tangxian kaolao yibing wanhufu zhi yin'" 河南叶县出土元代“唐县栲栳义兵万户府之印”. In Wen Wu 文物 1983.6 : 95.

Yao Shuomin 姚朔民. "Zaishi Heibei Pingshan suochu basibawen guanyin" 再释河北平山所出八思巴文官印. In Wen Wu 文物 1990.10 : 93 & 78.

Ye Qifeng 叶其峰. "Gugong cang Yuan basibazi yin jixiangguan wenti" 故宫藏元八思巴字印及相关问题. In Wen Wu 文物 1987.10 : 82-85.

Zheng Yuanwen 郑远文. "Guizhousheng Qianxixian faxian yimei Yuandai basibawen guanyin" 贵州省黔西县发现一枚元代八思巴文官印. In Wen Wu 文物 2001.7 : 86.

Zhou Jiuyi 周九宜. "Hunan Ningyuan chutu yimei tongzhi basibawen 'Ningyuan wu' guanya" 湖南宁远出土一枚铜质八思巴文“宁远务”官押. In Wen Wu 文物 1993.6 : 88.

3.2 WG2 Documents

N2622 "Proposal to encode the Phags-pa Script". Andrew C. West. 2003-09-18.

N2652R "Principles and Procedures for Allocation of New Characters and Scripts and handling of Defect Reports on Character Names". Ad hoc group on Principles and Procedures. 2003-11-24.

N2706 "Summary of Voting on ISO/IEC JTC 1/SC 2 N 3696 : Project Subdivision Proposal for ISO/IEC 10646: 2003/Amendment 1". 2004-02-03.

N2719 "Response to Comments on Phags-pa Proposal in N2706". Andrew C. West. 2004-03-10.

N2745 "HPhags-pa script encoding". China National Information Technology Standardization Technical Committee and Mongolian Agency for Standardization and Metrology. 2004-04-02.

N2745-1 "Proposal to Encode the HPhags-pa Script".

N2745-2 "Summary of the 1st Joint Meeting on the HPhags-Pa Script Encoding".

N2745-3 "Summary of the User’s Agreement Related to HPhags-pa Script".