Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Organisation Internationale de Normalisation
Международная организация по стандартизации

Doc Type:	**Working Group Document**
Title:	**Proposal to encode four Tibetan-Sanskrit letters used in Kalacakra texts**
Source:	**Andrew West and Christopher Fynn**
Status:	**Individual Contribution**
Action:	**For consideration by JTC1/SC2/WG2 and UTC**
Date:	**2009-01-24**

1. Introduction

This is a proposal to encode one Tibetan letter and three subjoined Tibetan letters which are used in Tibetan books on Sanskrit grammar, in ritual texts focused on the Kālacakra ("Wheel of Time") system, and in commentaries on the Kālacakra Tantra.

Proposed Code Point	Proposed Character Name	Glyph
0F8C	TIBETAN SIGN INVERTED MCHU CAN	ვ
0F8D	TIBETAN SUBJOINED SIGN LCE TSA CAN	◌
0F8E	TIBETAN SUBJOINED SIGN MCHU CAN	◌
0F8F	TIBETAN SUBJOINED SIGN INVERTED MCHU CAN	◌

Unicode Character Properties	Script	Line Break Property
0F8C;TIBETAN SIGN INVERTED MCHU CAN;Lo;0;L;;;;;N;;;;;	Tibetan	AL
0F8D;TIBETAN SUBJOINED SIGN LCE TSA CAN;Mn;0;NSM;;;;;N;;;;;	Tibetan	CM
0F8E;TIBETAN SUBJOINED SIGN MCHU CAN;Mn;0;NSM;;;;;N;;;;;	Tibetan	CM
0F8F;TIBETAN SUBJOINED SIGN INVERTED MCHU CAN;Mn;0;NSM;;;;;N;;;;;	Tibetan	CM

2. Inverted MCHU CAN

An inverted form of U+0F89 ꧀ TIBETAN SIGN MCHU CAN is occasionally found. In particular the ordinary and inverted forms of this sign are used contrastively in an important 18th century Chinese text, *Tóngwén Yùntǒng* 同文韻統, which describes the rules for transliterating Sanskrit and Tibetan into Mongolian and Manchu. A project to digitalise this text has identified the inverted MCHU CAN sign as a character that requires encoding.

Fig. 1 : *Tóngwén Yùntǒng* 同文韻統　page 135

Note that the inverted and ordinary forms of the Tibetan MCHU CAN sign corrrespond to U+1883 ᢃ MONGOLIAN LETTER ALI GALI UBADAMA and U+1884 ᢄ MONGOLIAN LETTER ALI GALI INVERTED UBADAMA respectively, although the names of these two characters are probably inadvertently swapped, as it is the glyph form of U+1883 that is the inverted ubadama. It is believed that these two Mongolian characters were encoded on the basis of the example shown in Fig.1.

3. Subjoined LCE TSA CAN and MCHU CAN

The characters U+0F88 ☲ TIBETAN SIGN LCE TSA CAN and U+0F89 ᠊ TIBETAN SIGN MCHU CAN are special signs used in Kālacakra ("Wheel of Time") texts. The former is used as a superfix to the letters KA ཀ and KHA ཁ, and the latter is used as a superfix to the letters PA པ and PHA ཕ. Although logically they are superfixes, and almost never occur independently, the Tibetan encoding model means that they are encoded as letters (General category = Lo), so that they can act as the base character in a stack, with one or more subjoined letters following.

However, a Kangyur input project has identified a number of instances where these two signs occur superfixed to another letter *within a stack*, and at present it is impossible to represent such stacks because U+0F88 and U+0F89 can only occur at the head of a stack. Figs. 2 and 3 show examples of the MCHU CAN sign and the LCE TSA CAN sign occuring both at the head and in the middle of a stack.
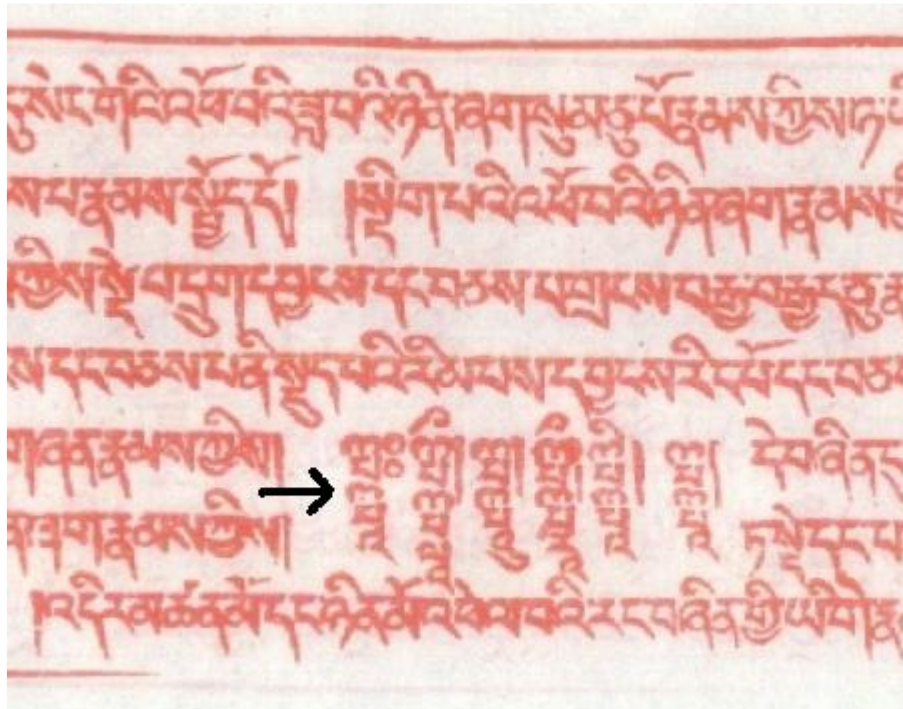
Fig. 2 : *bka' 'gyur (sde dge par phud)* བཀའ་འགྱུར་(སྡེ་དགེ་པར་ཕུད་) (Delhi, 1976-1979) page 112
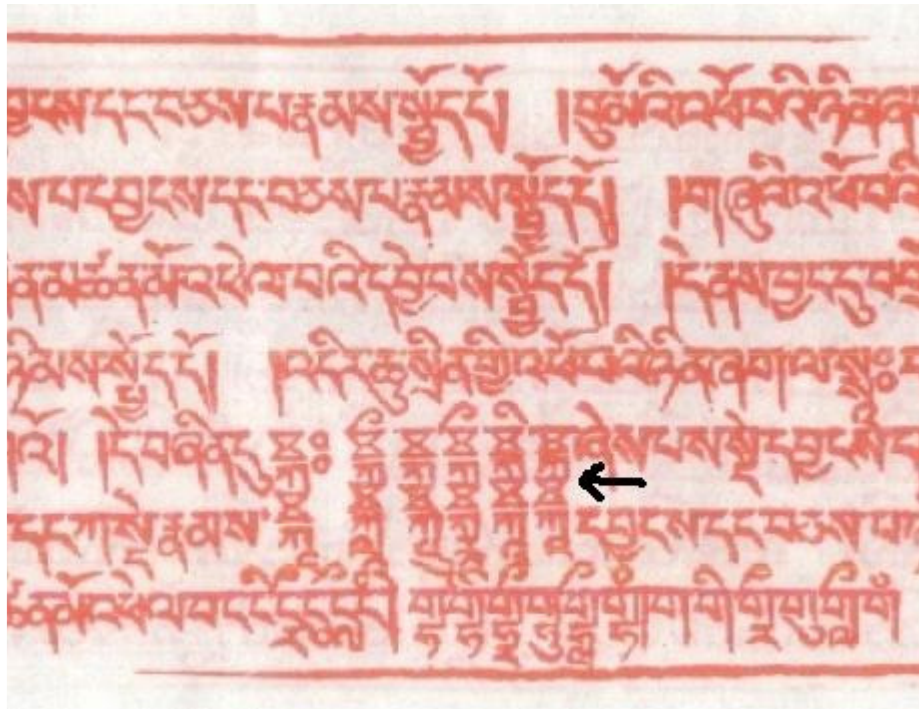
Fig. 3 : *bka' 'gyur (sde dge par phud)* བཀའ་འགྱུར་(སྡེ་དགེ་པར་ཕུད་) (Delhi, 1976-1979) page 112

Although occurences such as these are very rare, there is a pressing need for the Kangyur input project to be able to represent such stacks in plain text. We therefore propose that subjoined versions of these two signs are encoded to meet this need. Although no examples of the inverted MCHU CAN sign occuring within a stack have yet been found, there is no reason why, in principle, the inverted form of the MCHU CAN sign should not also be used in the same way that the ordinary MCHU CAN sign is, *within a stack*, and it therefore seems prudent to also encode a subjoined inverted MCHU CAN sign, so that all theoretical possibilities are catered for by the Tibetan encoding model.

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646
Please fill all the sections A, B and C below.
Please read Principles and Procedures Document (P & P) from
http://www.dkuug.dk/JTC1/SC2/WG2/docs/principles.html for guidelines and details before filling this form.
Please ensure you are using the latest Form from http://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html.
See also http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html for latest *Roadmaps*.
Form number: N3452-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05,
2001-09, 2003-11, 2005-01, 2005-09, 2005-10, 2007-03, 2008-05)

A. Administrative

1. **Title:** *Proposal to encode four Tibetan-Sanskrit letters used in Kalacakra texts*

2. Requester's name: *Andrew West and Christopher Fynn*

3. Requester type (Member body/Liaison/Individual contribution): *Individual contribution*

4. Submission date: *2009-01-24*

5. Requester's reference (if applicable): *N/A*

6. Choose one of the following:

   This is a complete proposal: *YES*

   (or) More information will be provided later: *NO*

B. Technical - General

1. Choose one of the following:

   a. This proposal is for a new script (set of characters): *NO*

   Proposed name of script: *N/A*

   b. The proposal is for addition of character(s) to an existing block: *YES*

   Name of the existing block: *Tibetan*

2. Number of characters in proposal: *4*

3. Proposed category (select one from below - see section 2.2 of P&P document):

   A-Contemporary _____ B.1-Specialized (small collection) _____ *X* B.2-Specialized (large collection) _____

   C-Major extinct _____ D-Attested extinct _____ E-Minor extinct _____

   F-Archaic Hieroglyphic or Ideographic _____ G-Obscure or questionable usage symbols _____

4. Is a repertoire including character names provided? *YES*

   a. If YES, are the names in accordance with the "character naming guidelines" *YES*

   b. Are the character shapes attached in a legible form suitable for review? *YES*

5. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for

   publishing the standard? *Christopher Fynn*

   If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools

   used:

6. References:

   a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided? *YES*

   b. Are published examples of use (such as samples from newspapers, magazines, or other sources)

   of proposed characters attached? *YES*

7. Special encoding issue

   Does the proposal address other aspects of character data processing (if applicable) such as input,

   presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)? *N/A*

8. Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at http://www.unicode.org for such information on other scripts. Also see http://www.unicode.org/Public/UNIDATA/UCD.html and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before?   *NO*
   If YES explain   *N/A*
2. Has contact been made to members of the user community (for example: National Body,
   user groups of the script or characters, other experts, etc.)?   *YES*
   If YES, available relevant documents: *N/A*
3. Information on the user community for the proposed characters (for example:
   size, demographics, information technology use, or publishing use) is included?   *Scholarly use.*
   Reference: *N/A*
4. The context of use for the proposed characters type of use; common or rare)   *Rare*
   Reference: *N/A*
5. Are the proposed characters in current use by the user community?   *YES*
   If YES, where? Reference: *In Kālacakra texts.*
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely
   in the BMP?   *YES*
   If YES, is a rationale provided?   *NO*
   If Yes, reference: *N/A*
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?   *N/A*
8. Can any of the proposed characters be considered a presentation form of an existing
   character or character sequence?   *NO*
   If YES, is a rationale for its inclusion provided?   *N/A*
   If Yes, reference: *N/A*
9. Can any of the proposed characters be encoded using a composed character sequence of either
   existing characters or other proposed characters?   *NO*
   If YES, is a rationale for its inclusion provided?   *N/A*
   If Yes, reference: *N/A*
10. Can any of the proposed character(s) be considered to be similar (in appearance or function)
    to an existing character?   *YES*
    If YES, is a rationale for its inclusion provided?   *YES*
    If Yes, reference: *See Proposal Summary below*
11. Does the proposal include use of combining characters and/or use of composite sequences?   *NO*
    If YES, is a rationale for such use provided?   *N/A*
    If Yes, reference: *N/A*
    Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?   *N/A*
    If Yes, reference: *N/A*
12. Does the proposal contain characters with any special properties such as
    control function or similar semantics?   *NO*
    If YES, describe in detail (include attachment if necessary)   *N/A*

13. Does the proposal contain any Ideographic compatibility character(s)?   *NO*
    If YES, is the equivalent corresponding unified ideographic character(s) identified?   *N/A*
    If Yes, reference: *N/A*