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 two Tibetan annotation marks**
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 two Tibetan annotation marks. These marks consist of a series of tsheg-like dots that are used in traditional Tibetan books to link the text being annotated to its annotation (which is normally written in a smaller size than the main text). They are found in many books, and when used they are often used extensively, as shown in Fig. 1.

Proposed Code Point	Proposed Character Name	Glyph
0FD9	TIBETAN MARK LEADING MCHAN RTAGS	⸰⸰⸰
0FDA	TIBETAN MARK TRAILING MCHAN RTAGS	⸰⸰⸰

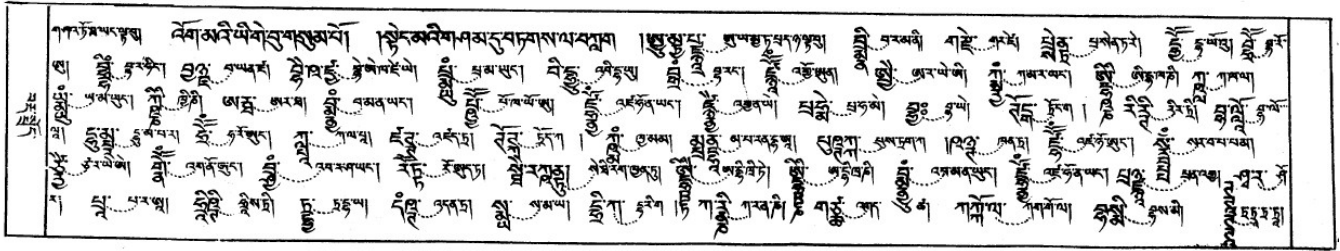| Unicode Character Properties	Script	Line Break Property
`0FD9;TIBETAN MARK LEADING MCHAN RTAGS;Po;0;L;;;;;N;;;;;`	Tibetan	GL
`0FDA;TIBETAN MARK TRAILING MCHAN RTAGS;Po;0;L;;;;;N;;;;;`	Tibetan	GL

Fig. 1 : *sngags kyi klog thabs shes rab mig 'byed* སྔགས་ཀྱི་ཀློག་ཐབས་ཤེས་རབ་མིག་འབྱེད། (New Delhi, 1999) page 498

In most cases the annotation is to the right of the text being annotated (as in Fig. 1), but sometimes the annotation is to the left of the text being annotated. The actual realisation of annotation marks in traditional books varies from case to case, depending on the relative position of the annotation text to the annotated text and the shapes of the terminal stacks it joins together, and so there may not always be an obvious difference in shape between an annotation mark to the left and an annotation mark to the right. However, in order to distinguish in plain text whether the annotation mark links to or from an annotation, two annotation marks with nominal glyph shapes are proposed for encoding: a leading annotation mark representing a link from the annotation on the left to its target text on the right, and a trailing annotation mark representing a link from the target text on the left to its annotation on the right.

Fig. 2 : Examples of Leading (left) and Trailing (right) Annotation Marks in Computer Typeset Tibetan

Although it is not possible to mimic the potentially meandering course of an annotation mark across the page in a plain text encoding, it is still essential to be able to be able to indicate in plain text that there is an annotation relationship between two sections of adjacent text using the proposed characters. Furthermore, some non-Unicode Tibetan text processing systems, which were in widespread use, include these two characters (see Fig. 3), and so they are also required to be encoded for compatibility with such systems. In particular there is a need to convert many texts which were input using these legacy systems to Unicode.

c) Annotation Markers (mchan.rtags).

One type of Tibetan commentary uses annotations in the body of the text itself as a way of commenting on the body of the text, just as we use footnotes or parenthetical notes are used in English. The annotations (called mchan) usually follow the text that they are commenting upon and amplify or clarify it. The annotations are usually preceded by a series of tshegs and occasionally succeeded by them. The mark made by the group of tshegs joins the annotation to the text that it is commenting on.

mchan rtags trailing	◌	[99]	Skt4

Character List for Windows Encoding Tibetan Support Fonts: Font Skt4: page 90

Description	Tibetan	Dec.	Font	Gi.gu	Zhabs.kyu	'Greng.bu	Na.ro	A.chung	A.chung+ Zhabs.kyu	Unicode 3	Notes
mchan rtags leading	◌	[100]	Skt4								

Fig. 3 : Word Tibetan! 5.1 pages 90-91

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 two Tibetan annotation marks*
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: *2*
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: *TIBEX mailing list*
3. Information on the user community for the proposed characters (for example:
     size, demographics, information technology use, or publishing use) is included?   *Tibetans and Tibetologists.*
       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 traditional Tibetan 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*