

## ISO/IEC JTC 1/SC 2/WG 2 N3012

L2/05-346

**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 for guidelines and details before filling this form.)  
See <http://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html> for latest Form.  
See <http://www.dkuug.dk/JTC1/SC2/WG2/docs/principles.html> for latest Principles and Procedures document.  
See <http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html> for latest roadmaps.

(Form number: N2352-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09))

**A. Administrative****1. Title: Proposal to encode three archaic Tibetan characters**

2. Requester's name: Andrew C. West \_\_\_\_\_
3. Requester type: Expert contribution \_\_\_\_\_
4. Submission date: 24th October 2005
5. Requester's reference (if applicable): \_\_\_\_\_
6. (Choose one of the following:)
- This is a complete proposal: Complete proposal
- or, More information will be provided later: \_\_\_\_\_

**B. Technical - General**

1. (Choose one of the following:)
- a. This proposal is for a new script (set of characters): \_\_\_\_\_  
Proposed name of script: \_\_\_\_\_
- 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: 3 \_\_\_\_\_
3. Proposed category (see section II, Character Categories): A \_\_\_\_\_
4. Proposed Level of Implementation (1, 2 or 3)  
(see clause 14, ISO/IEC 10646-1: 2000): 2 \_\_\_\_\_
- Is a rationale provided for the choice? No \_\_\_\_\_
- If Yes, reference: \_\_\_\_\_
5. Is a repertoire including character names provided? Yes \_\_\_\_\_
- a. If YES, are the names in accordance with the  
'character naming guidelines in Annex L of ISO/IEC 10646-1: 2000? Yes \_\_\_\_\_
- b. Are the character shapes attached in a legible form suitable for review?  
Yes \_\_\_\_\_
6. Who will provide the appropriate computerized font (ordered preference:  
True Type, or PostScript format) for publishing the standard?  
A font can be provided if required, but it may be better to generate the \_\_\_  
a glyph for the character from the existing font used for the Tibetan \_\_\_\_\_  
block. Andrew West \_\_\_\_\_  
If available now, identify source(s) for the font (include address,  
e-mail, ftp-site, etc.) and indicate the tools used:  
\_\_\_\_\_
7. References:
- a. Are references (to other character sets, dictionaries, descriptive  
texts etc.) provided? No \_\_\_\_\_
- b. Are published examples of use (such as samples from newspapers,  
magazines, or other sources) of proposed characters attached? Yes \_\_\_\_\_
8. Special encoding issues:  
Does the proposal address other aspects of character data processing  
(if applicable) such as input, presentation, sorting, searching, indexing,  
transliteration etc. (if yes please enclose information)?  
No \_\_\_\_\_

**C. Technical - Justification**

1. Has this proposal for addition of character(s) been submitted before? No\_\_\_  
If YES explain \_\_\_\_\_
2. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)? Yes\_\_\_\_\_
  - If YES, with whom? Members of TIBEX mailing list (tibex@unicode.org)\_\_\_\_\_
  - If YES, available relevant documents: TIBEX mailing list archives\_\_\_\_\_
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? No\_\_\_\_\_
  - Reference: \_\_\_\_\_
4. The context of use for the proposed characters (type of use; common or rare) Common\_\_\_\_\_
  - Reference: \_\_\_\_\_
5. Are the proposed characters in current use by the user community? Yes\_\_\_\_\_
  - If YES, where? Reference: In reproductions of archaic Tibetan texts\_\_\_\_\_
6. After giving due considerations to the principles in *Principles and Procedures document* (a WG 2 standing document) must the proposed characters be entirely in the BMP? Yes\_\_\_\_\_
  - If YES, is a rationale provided? No\_\_\_\_\_
  - If YES, reference: \_\_\_\_\_
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)? Yes\_\_\_\_\_
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? No\_\_\_\_\_
  - If YES, is a rationale for its inclusion provided? \_\_\_\_\_
  - If YES, reference: \_\_\_\_\_
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? No\_\_\_
  - If YES, is a rationale for its inclusion provided? \_\_\_\_\_
  - If YES, reference: \_\_\_\_\_
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character? No\_\_\_\_\_
  - If YES, is a rationale for its inclusion provided? \_\_\_\_\_
  - If YES, reference: \_\_\_\_\_
11. Does the proposal include use of combining characters and/or use of composite sequences (see clauses 4.12 and 4.14 in ISO/IEC 10646-1: 2000)? No\_\_\_\_\_
  - If YES, is a rationale for such use provided? \_\_\_\_\_
  - If YES, reference: \_\_\_\_\_
  - Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? \_\_\_\_\_
  - If YES, reference: \_\_\_\_\_
12. Does the proposal contain characters with any special properties such as control function or similar semantics? No\_\_\_\_\_
  - If YES, describe in detail (include attachment if necessary) \_\_\_\_\_
13. Does the proposal contain any Ideographic compatibility character(s)? \_\_\_
  - If YES, is the equivalent corresponding unified ideographic character(s) identified? No\_\_\_\_\_
  - If YES, reference: \_\_\_\_\_

**ADDITIONAL INFORMATION****1. PROPOSED CHARACTERS**

This proposal covers three punctuation marks that are required for encoding in order to represent early Tibetan manuscripts and monumental inscriptions.

Code Point	Representative Glyph	Proposed Character Name
0FD3	ུ	TIBETAN MARK NYIS TSHEG
0FD4	འ	TIBETAN MARK INITIAL BRDA RNYING YIG MGO MDUN MA
0FD5	ཡ	TIBETAN MARK CLOSING BRDA RNYING YIG MGO SGAB MA

**Unicode Properties:**

0FD3;TIBETAN MARK NYIS TSHEG;Po;0;L;;;;;N;nyi tsek;;;  
 0FD4;TIBETAN MARK INITIAL BRDA RNYING YIG MGO MDUN MA;Po;0;L;;;;;N;da nying yik go dun ma;;;  
 0FD5;TIBETAN MARK CLOSING BRDA RNYING YIG MGO SGAB MA;Po;0;L;;;;;N;da nying yik go kab ma;;;

**Linebreak Properties:**

0FD3;BA # TIBETAN MARK NYIS TSHEG  
 0FD4;BB # TIBETAN MARK INITIAL BRDA RNYING YIG MGO MDUN MA  
 0FD5;AL # TIBETAN MARK CLOSING BRDA RNYING YIG MGO SGAB MA

**Notes on Proposed Names:**

TIBETAN MARK NYIS TSHEG : NYIS TSHEG ༴ུ་ཚེག་། means "double tsheg" (cf. 0F0E TIBETAN MARK NYIS SHAD "double shad").

TIBETAN MARK INITIAL BRDA RNYING YIG MGO MDUN MA and TIBETAN MARK CLOSING BRDA RNYING YIG MGO SGAB MA : BRDA RNYING བར་རྣམ་འགྲུབ་། means "old orthography", and refers to Tibetan letters and marks used in archaic texts before the orthographic reforms of the Tibetan script.

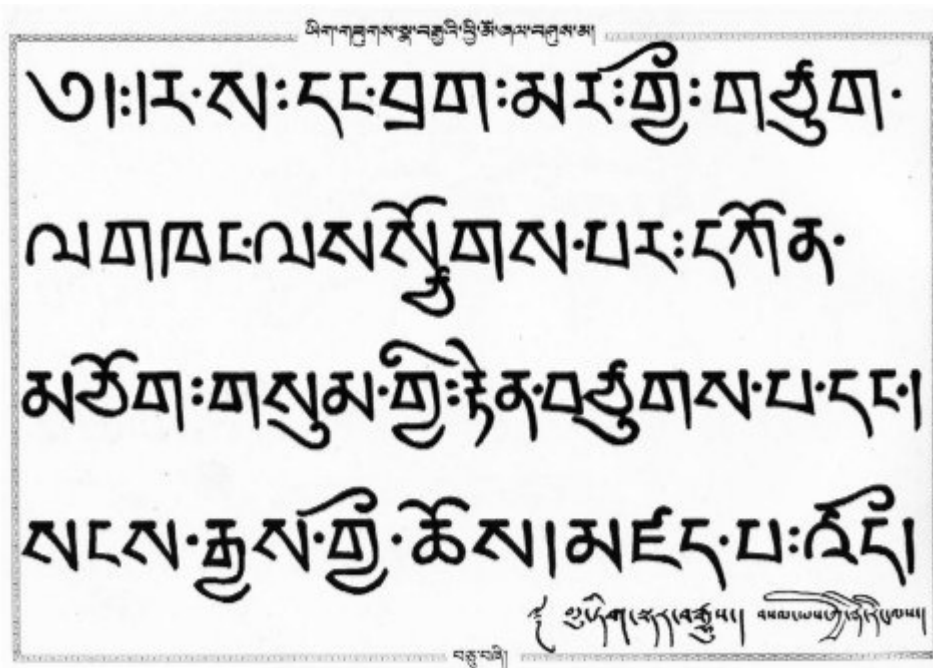
## 2. Double Tsheg

The double tsheg mark (Tibetan NYIS TSHEG) is a form of the intersyllabic tsheg mark that comprises two dots in vertical juxtaposition, in contrast with the normal tsheg mark [0F0B] that comprises a single dot.

The double tsheg mark is common in early Tibetan monumental inscriptions and manuscript texts, as exemplified by Illustrations 1-3: Illustration 1 shows a modern calligraphic copy of the first part of the inscription on a famous 9th-century stone pillar at Samye; Illustration 2 shows a paper fragment from Dunhuang; and Illustration 3 shows a woodslip from the fort at Mazartagh at the south of the Taklamakan Desert (datable to the late 8th or early 9th century).

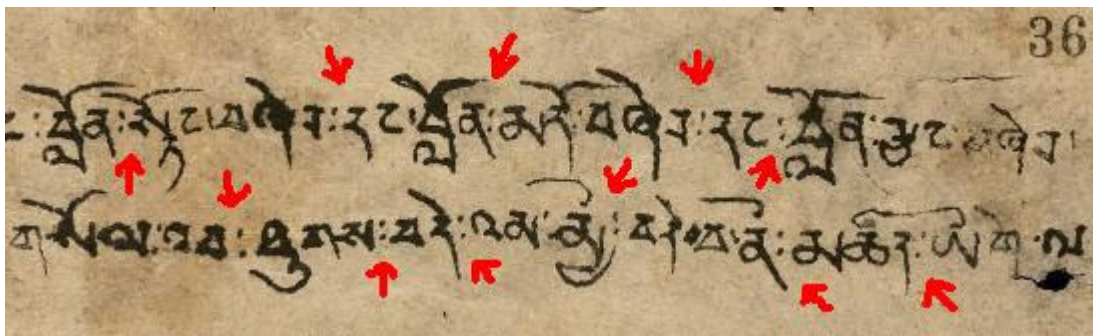
Note how both single tsheg and double tsheg marks occur together within the same text in Illustration 1.

### Illustration 1 : Inscription on the Stone Pillar at Samye

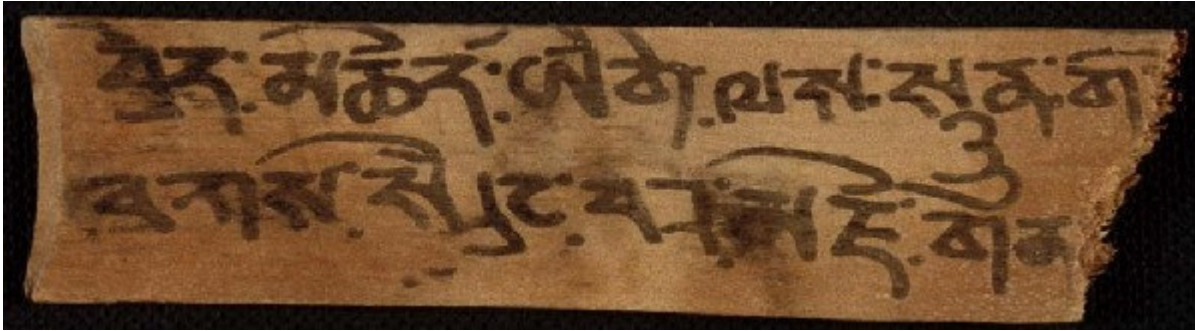


**Source :** ཡིག་གཟུགས་སྐབ་བྱུང་གི་ཕྱི་མོ་ཞལ་བཤུས་མ། (《藏文书体百种》) [One Hundred Examples of Tibetan Script Styles] (Beijing : Nationalities Publishing House, 1999) page 14.

### Illustration 2 : Paper Fragment from Dunhuang

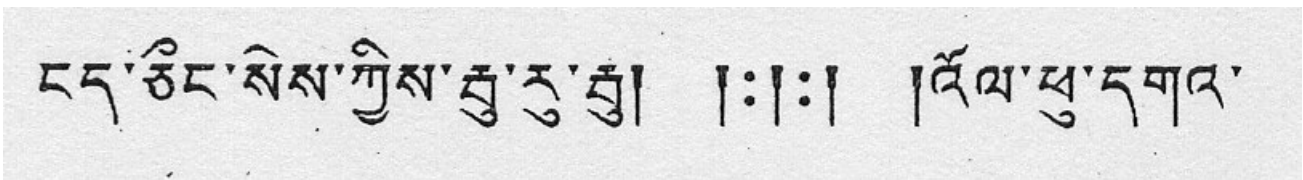


**Source :** International Dunhuang Project – <http://idp.bl.uk/GetObjectOverview/30102>.

**Illustration 3 : Woodslip from Mazartagh**

**Source :** International Dunhuang Project – <http://idp.bl.uk/GetObjectOverview/7056>.

In addition to its normal syllable demarcation function, the double tsheg is often placed between shad marks to mark off the end of a section, as shown in Illustration 4, which is from a typeset edition of the early Tibetan manuscripts held at the Bibliothèque Nationale in Paris.

**Illustration 4 : Typeset copy of an early Tibetan Manuscript**

**Source :** Choix de Documents Tibétains conservés à la Bibliothèque Nationale Vol.IV (ICCAA, 2001) p.387.

The double tsheg mark is also used in some modern editions of texts to indicate the point in the text from which a textual note refers. See Illustration 5 for an example of this usage taken from a modern edition of the "Comparative Tibetan Canon" བཀའ་བསྟན་འཕེལ་བཞུགས་ published in China. It is proposed that this less common modern usage of the double tsheg mark be unified with the archaic double tsheg mark.












Illustration 5 : Annotated edition of the Comparative Tibetan Canon

རང་བཞིན་ཡོད་མིན་ན། །གལ་ཏེ་འགོག་སྲ་འདགག་འབྱི་ཞེས། །ཟེ་  
 འཐད་པ་མ་ཡིན་ལ། །འབྱི་དང་ཅིག་ཅར་མི་འཐད་དོ། །ཞེས་སྐུས་པ་  
 ཡིན་པ་འདི་ལ་བཤད་པར་བྱ་སྟེ། རུས་གསུམ་གཏན་ཚིགས་གང་ཡོད་སྲ།  
 ལན་བཏབ་དེ་དང་མཚུངས་ཕྱིར་རོ། །རུས་གསུམ་ཕྱིད་ཀྱི་གཏན་ཚིགས་ནི  
 ལྷོང་ཉིད་སྐྱབ་ནམས་ལ་རུང་། །གཏན་ཚིགས་རུས་གསུམ་འགོག་པར་བྱེད  
 ཞེས་བྱ་བ་གང་ཡིན་པ་དེ་ནི་ལན་བཏབ་བྱིན་ཏེ། གང་གི་ཕྱིར་ཞེ་ན་བསྐྱ

- ① «ལེ» «སྐར» རྫོག
- ② «ལེ» «སྐར» མཁའ།
- ③ «ལེ» «སྐར» བཟུང།
- ④ «ལེ» «སྐར» རྫོག
- ⑤ «ལེ» «སྐར» རྩུག།
- ⑥ «ལེ» «སྐར» གལ་འདོད་དེ།
- ⑦ «ལེ» «སྐར» ལྟ།
- ⑧ «ལེ» «སྐར» ལྟ།
- ⑨ «ལེ» «སྐར» ཕྱིད་ཀྱིས།
- ⑩ «ལེ» «སྐར» ཞེ་ན།

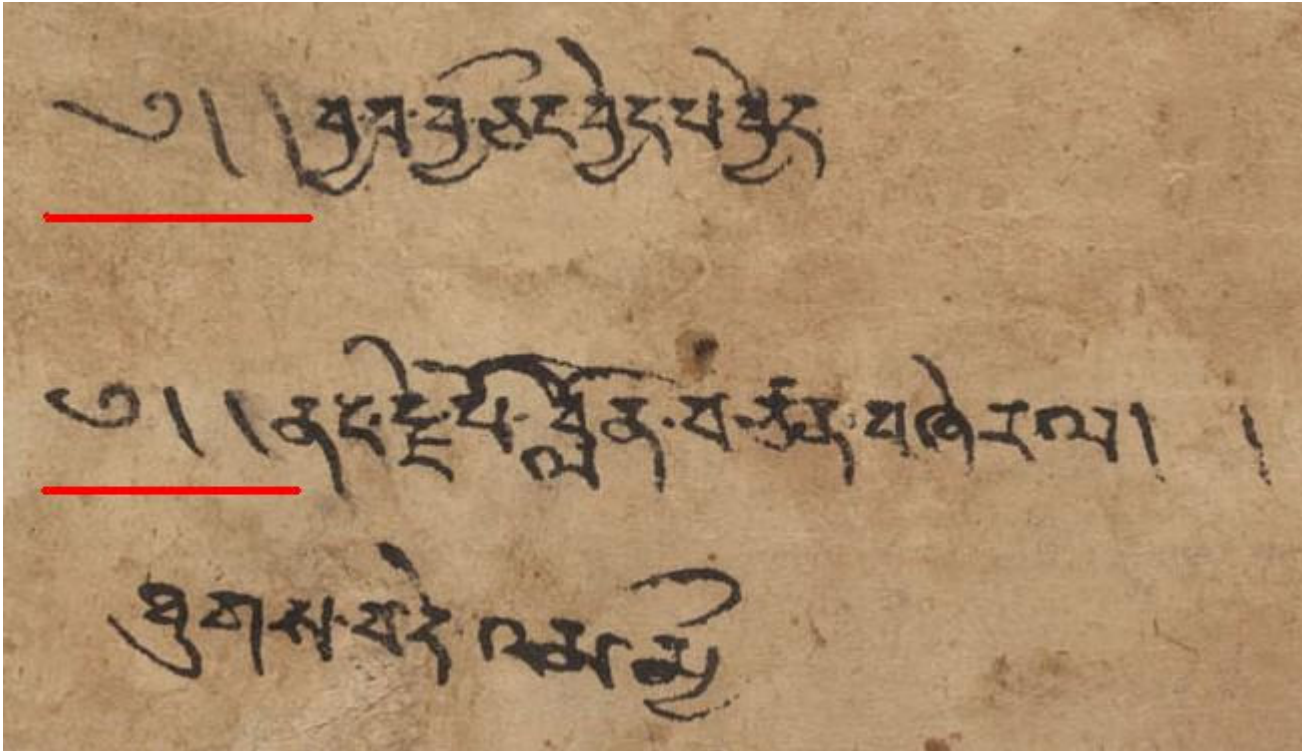
### 3. Archaic Head Marks

In traditional Tibetan texts a single-looped , double-looped  or even triple-looped  mark is used to mark the start of text or the front folio of a page. As noted in the Unicode Standard section 9.11 : "The head mark can and does vary from text to text; there are many different ways to write it. The common type of head mark has been provided for with U+0F04 TIBETAN MARK INITIAL YIG MGO MDUN MA and its extension U+0F05 TIBETAN MARK CLOSING YIG MGO SGAB MA." In early Tibetan texts the head mark is normally written with a simple single line  rather than the ornate double line that is seen in the modern glyph form. Whilst these archaic style head marks may be considered to be glyph variants of 0F04/0F05, and are indeed sometimes seen in more recent texts (although rarely in traditional printed books using the standard book-style of Tibetan script), it would be of great convenience to scholars working with the many thousands of early Tibetan manuscripts that are held in collections in China, Japan, Russia, Germany, France and Britain to be able to uniquely represent the archaic single-line form of the head mark.

One precomposed sequence of archaic form single-looped head mark followed by single shad, double tsheg and single shad is already encoded as 0F07 |:| [TIBETAN MARK YIG MGO TSHEG SHAD MA]. Whilst this allows some initial sequences in early manuscripts and inscriptions to be represented uniquely, it means that other equally common initial sequences such as head mark followed by shad  |, head mark followed by shad, single tsheg and shad  |' |, or head mark followed by double shad  ||, cannot be represented at present. If the characters proposed here are accepted for encoding, then I would recommend that the existing character 0F07 [TIBETAN MARK YIG MGO TSHEG SHAD MA] be deprecated in favour of the decomposed sequence <0FD4 0F0D 0FD3 0F0D>  | : |.

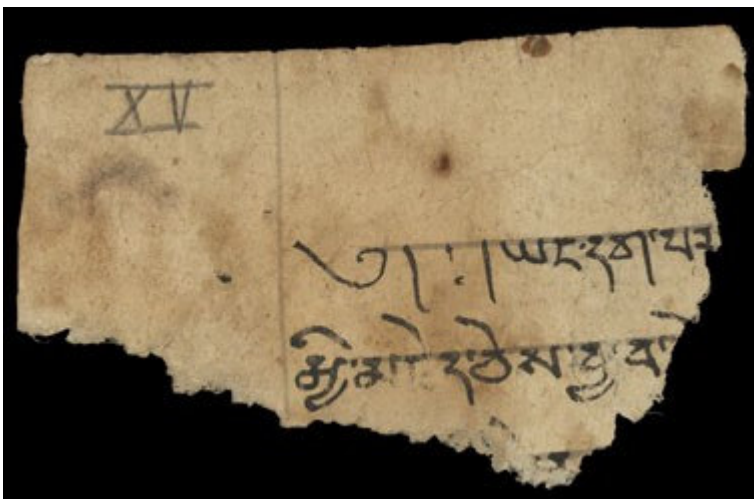
Illustrations 6-8 show examples of this head mark in early manuscripts: Illustration 6 shows the head mark followed by a double shad; Illustration 7 shows the head mark followed by a shad, a double tsheg and a shad (cf. Illustration 1); and Illustration 8 shows the head mark followed by a shad, a single tsheg and a shad.

**Illustration 6 : Paper Fragment from Dunhuang**



**Source :** International Dunhuang Project – <http://idp.bl.uk/GetObjectOverview/30142>.

**Illustration 7 : Paper Fragment from Dunhuang**



**Source :** International Dunhuang Project – <http://idp.bl.uk/GetObjectOverview/4679>.

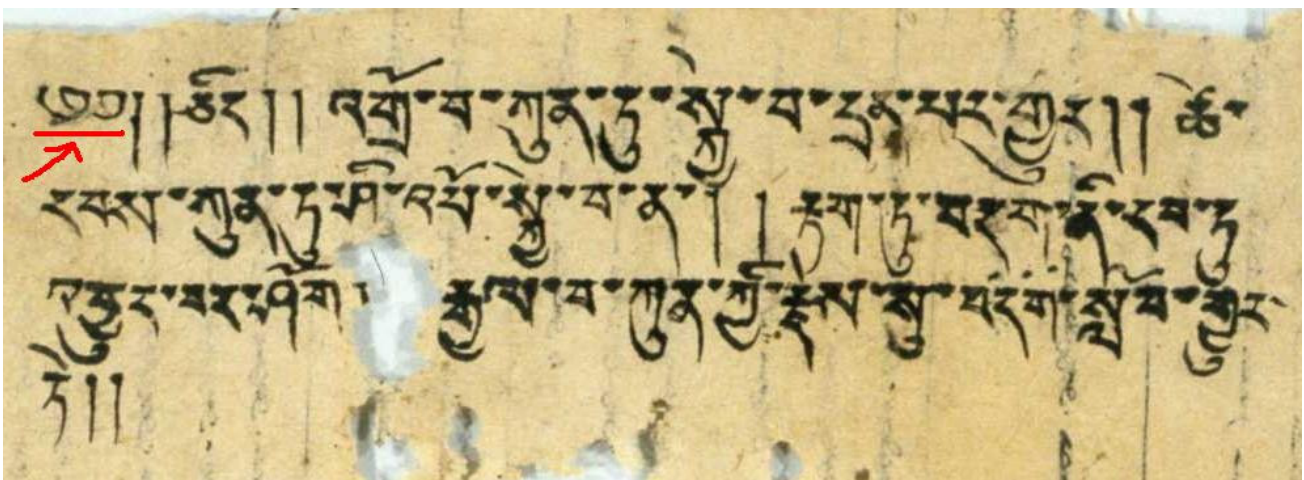
Illustration 8 : Paper Fragment from Domoko



Source : International Dunhuang Project – <http://idp.bl.uk/GetObjectOverview/30120>.

Although single-looped head marks are the norm in early manuscripts, double-looped single-line head marks do occur in some later manuscripts, as shown in Illustration 9. In the same way that ordinary double-looped head marks are represented by the sequence <0F04 0F05> ཨ, archaic form double-looped head marks should be encoded by the sequence <0FD4 0FD5> ཨ.

Illustration 9 : Paper Fragment from Turfan



Source : Digitales Turfan-Archiv (Berlin Turfan-Collection) – [http://www.bbaw.de/forschung/turfanforschung/dta/monght/images/monght057\\_verso.jpg](http://www.bbaw.de/forschung/turfanforschung/dta/monght/images/monght057_verso.jpg).



