



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 add four Tibetan characters for Balti to the BMP of the UCS****Source: Michael Everson****Status: Individual Contribution****Action: For consideration by JTC1/SC2/WG2 and UTC****Date: 2005-09-04**

This document requests four additional Tibetan characters to be added to the UCS and contains the proposal summary form. These letters are used to represent sounds of the Balti language when written in Tibetan script. The Balti language belongs to the archaic western branch of the Tibetan language family, most closely related to Purik and Ladakhi in the western Himalayas, and some dialects in Amdo (eastern Tibet), especially around the Blue Lake.

The letters are reversed KA and reversed RA, used for [q] and [ɽ] respectively, and KHA and GA modified with a TSA -PHUR ear, for [x] and [ɣ] respectively. These could in principle be composed with TSA -PHRU, but there is a serious problem with the canonical combining class of TSA -PHRU which makes it a problem to use. TSA -PHRU is added to a base letter when forming contractions and usually indicates TSA, TSHA, DZA, or ZA in the contraction or abbreviation. Here, we are talking about separate (new) letters for transcribing Urdu or Arabic words. Accordingly, the KHHA and GHHA should be considered atomic, as Tibetan TSA, TSHA, and DZA are.

0F6B		TIBETAN LETTER KKA
0F6C		TIBETAN LETTER KHHA
0F6D		TIBETAN LETTER GHHA
0F6E		TIBETAN LETTER RRA

Unicode properties

```
0F6B;TIBETAN LETTER KKA;Lo;0;L;;;;N;;;;;
0F6C;TIBETAN LETTER KHHA;Lo;0;L;;;;N;;;;;
0F6D;TIBETAN LETTER GHHA;Lo;0;L;;;;N;;;;;
0F6E;TIBETAN LETTER RRA;Lo;0;L;;;;N;;;;;
```

Reference

Hussainabadi, Mohamad Yusuf. 1990. *Balti Zaban*. [s.l.]: [s.n.].

Figures



Figure 1. Cover of the book *Balti Zaban*.



Figure 2. Sample from the book *Balti Zaban*, showing the regular set of Tibetan letters.

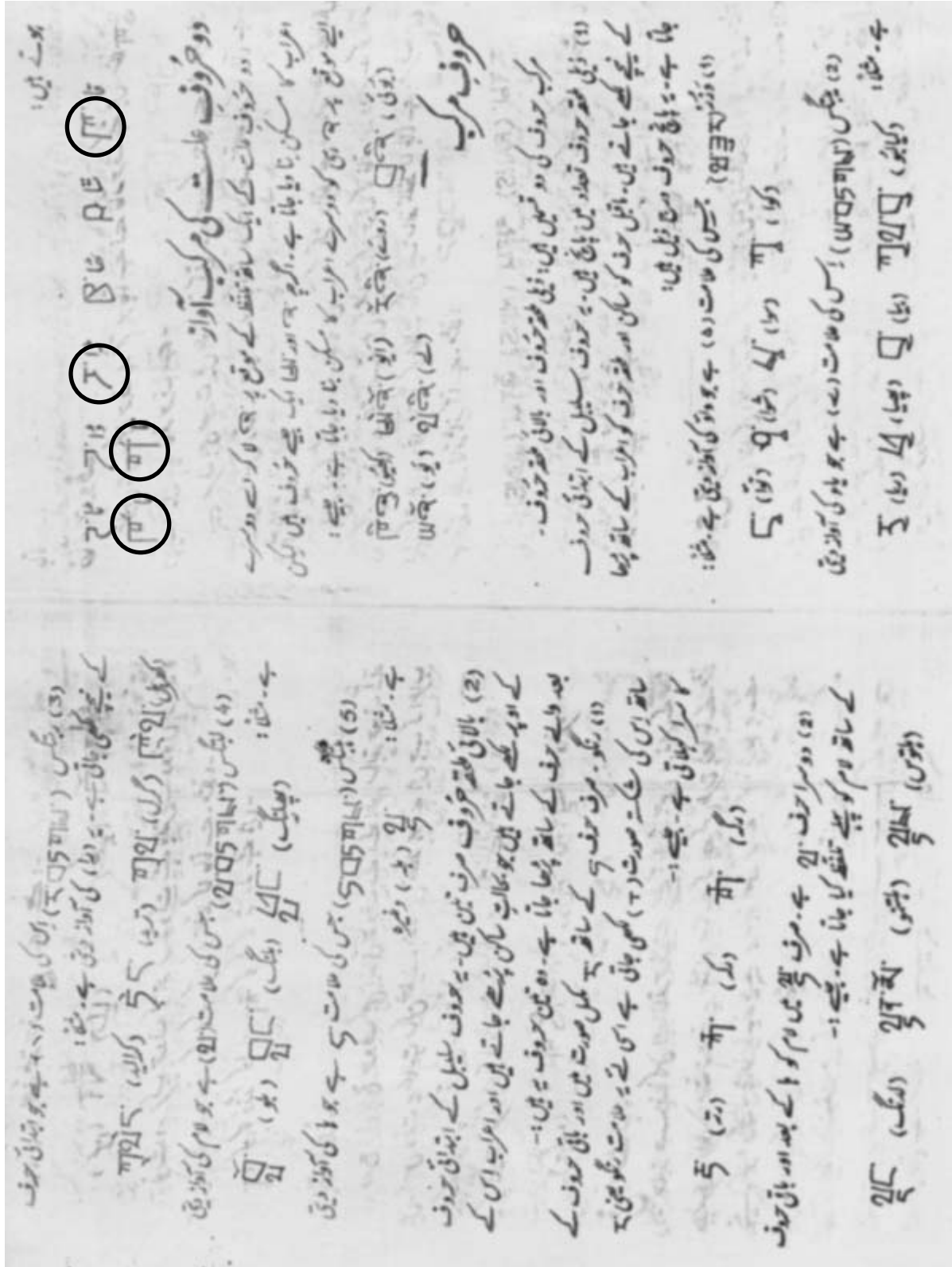


Figure 3. Sample from the book *Balti Zaban*, showing extensions for Balti. Four of those characters are already encoded in the Tibetan block (for Sanskrit extensions); the four here are shown with their Arabic transliterations, and I give the IPA here as well: 𑄗 [a], 𑄘 [qa], 𑄙 [xa], and 𑄚 [ya].

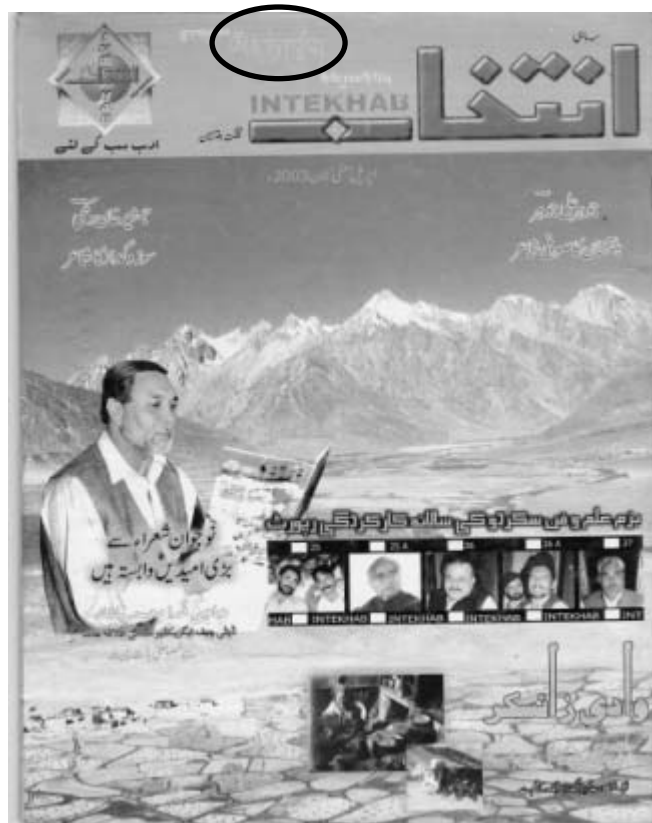


Figure 2. Sample from the magazine *Intekhab*, showing the title with KHHA: 𑂔𑂗𑂢𑂰𑂱𑂲𑂳𑂴𑂵.

A. Administrative

1. Title

Proposal to add four Tibetan characters for Balti to the BMP of the UCS.

2. Requester's name

Michael Everson

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

Individual contribution.

4. Submission date

2005-09-05

5. Requester's reference (if applicable)

6. Choose one of the following:

6a. This is a complete proposal

Yes.

6b. More information will be provided later

No.

B. Technical – General

1. Choose one of the following:

1a. This proposal is for a new script (set of characters)

No.

Proposed name of script

1b. The proposal is for addition of character(s) to an existing block

Yes.

1b. Name of the existing block

Tibetan

2. Number of characters in proposal

4

3. Proposed category (see section II, Character Categories)

Category A

4a. Proposed Level of Implementation (1, 2 or 3) (see clause 14, ISO/IEC 10646-1: 2000)

Level 1.

4b. Is a rationale provided for the choice?

Yes.

4c. If YES, reference

Spacing letter characters.

5a. Is a repertoire including character names provided?

Yes.

5b. If YES, are the names in accordance with the character naming guidelines in Annex L of ISO/IEC 10646-1: 2000?

Yes.

5c. Are the character shapes attached in a legible form suitable for review?

Yes.

6a. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?

Michael Everson. TrueType.

6b. If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:

Michael Everson. Fontographer.

7a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?

Yes, see above.

7b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?

Yes.

8. Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)?

No.

9. 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.

Yes, see Unicode properties above.

C. Technical – Justification

1. Has this proposal for addition of character(s) been submitted before? If YES, explain.

No.

2a. 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.

2b. If YES, with whom?

Julian Green, who has worked with the foremost scholars in Baltistan, including Mohammad Yusuf Hussainabadi, Raza Tasawor, Ghulam Hassan Lobsang, and Syed Abbas Kazmi. The proposal has been seen by Chris Fynn, Tibetan script expert.

2c. If YES, available relevant documents

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

Yes. There are some 300,000 speakers of Balti.

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

Common when Balti is written in Tibetan script.

4b. Reference

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

Yes.

5b. If YES, where?

In Pakistan.

6a. 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.

6b. If YES, is a rationale provided?

Yes.

6c. If YES, reference

All Tibetan points are in the BMP.

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

No.

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

No.

8b. If YES, is a rationale for its inclusion provided?

8c. If YES, reference

9a. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?

Two of them could in principle be composed with TSA -PHRU, but this would cause processing difficulty when interacting with vowel matras. Accordingly, the KHHA and GHHA should be considered atomic, as Tibetan TSA, TSHA, and DZA are.

9b. If YES, is a rationale for its inclusion provided?

9c. If YES, reference

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

No.

10b. If YES, is a rationale for its inclusion provided?

10c. If YES, reference

11a. 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)?

Yes.

11b. If YES, is a rationale for such use provided?

No.

11c. If YES, reference

12a. Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?

No.

12b. If YES, reference

13a. Does the proposal contain characters with any special properties such as control function or similar semantics?

No.

13b. If YES, describe in detail (include attachment if necessary)

14a. Does the proposal contain any Ideographic compatibility character(s)?

No.

14b. If YES, is the equivalent corresponding unified ideographic character(s) identified?

14c. If YES, reference