Proposal to encode additional head marks for Zanabazar Square

Anshuman Pandey
Department of Linguistics
University of California, Berkeley
Berkeley, California, U.S.A.
anshuman.pandey@berkeley.edu

December 30, 2015

1   Introduction

This is a proposal to encode the following two characters in the Zanabazar Square block, which was approved for inclusion in Unicode (see L2/14-044 and the updated proposal L2/15-337):

Glyph	Code	Character name
凹	11A45	ZANABAZAR SQUARE INITIAL DOUBLE-LINED HEAD MARK
卪	11A46	ZANABAZAR SQUARE CLOSING DOUBLE-LINED HEAD MARK

2   Description

The folio below (from Byambaa Ragchaagiin 2005: 97) shows usage of the headmark 凹凸 in Zanabazar Square text. This mark corresponds to the Tibetan ༄༅ head mark. Although the Tibetan mark may be interpreted as an atomic unit, it is represented in encoded text as a sequence of ༄ U+0F04 TIBETAN MARK INITIAL YIG MGO MDUN MA and ༅ U+0F05 TIBETAN MARK CLOSING YIG MGO SGAB MA. The Zanabazar Square 凹凸 may also be interpreted as an atomic unit, but its encoded representation should follow the Tibetan model as consisting of an 'initial' 凹 and 'closing' 卪 element.

In the proposal to encode the Zanabazar Square script, it was suggested that these 'double-lined' initial and closing head marks should be treated as glyphic variants of 凵 ZANABAZAR SQUARE INITIAL HEAD MARK and ⌐ ZANABAZAR SQUARE CLOSING HEAD MARK on account of their limited occurrence. However, as they are required for producing transliterations of Tibetan texts in Zanabazar Square, they should be encoded as separate characters so that they can be used concurrently with the 'single-lined' head marks, eg. 凵卪 and 凹凸.

Similar to ᥡ ᴢᴀɴᴀʙᴀᴢᴀʀ ꜱQᴜᴀʀᴇ ɪɴɪᴛɪᴀʟ ʜᴇᴀᴅ ᴍᴀʀᴋ, the proposed ᥡ ᴢᴀɴᴀʙᴀᴢᴀʀ ꜱQᴜᴀʀᴇ ɪɴɪᴛɪᴀʟ ᴅᴏᴜʙʟᴇ-ʟɪɴᴇᴅ ʜᴇᴀᴅ ᴍᴀʀᴋ can be used as a base for *candrabindu* and *anusvāra* signs:

    ᥡ̐   <ᥡ ɪɴɪᴛɪᴀʟ ᴅᴏᴜʙʟᴇ-ʟɪɴᴇᴅ ʜᴇᴀᴅ ᴍᴀʀᴋ, ◌̐ ꜱɪɢɴ ᴄᴀɴᴅʀᴀʙɪɴᴅᴜ>

    ᥡ̐   <ᥡ ɪɴɪᴛɪᴀʟ ᴅᴏᴜʙʟᴇ-ʟɪɴᴇᴅ ʜᴇᴀᴅ ᴍᴀʀᴋ, ◌̐ ꜱɪɢɴ ᴄᴀɴᴅʀᴀʙɪɴᴅᴜ ᴡɪᴛʜ ᴏʀɴᴀᴍᴇɴᴛ>

    ᥡ̆   <ᥡ ɪɴɪᴛɪᴀʟ ᴅᴏᴜʙʟᴇ-ʟɪɴᴇᴅ ʜᴇᴀᴅ ᴍᴀʀᴋ, ◌̆ ꜱɪɢɴ ᴄᴀɴᴅʀᴀ ᴡɪᴛʜ ᴏʀɴᴀᴍᴇɴᴛ>

    ᥡ̊   <ᥡ ɪɴɪᴛɪᴀʟ ᴅᴏᴜʙʟᴇ-ʟɪɴᴇᴅ ʜᴇᴀᴅ ᴍᴀʀᴋ, ◌̊ ꜱɪɢɴ ᴀɴᴜꜱᴠᴀʀᴀ>

When the proposed ᥡ ᴢᴀɴᴀʙᴀᴢᴀʀ ꜱQᴜᴀʀᴇ ᴄʟᴏꜱɪɴɢ ᴅᴏᴜʙʟᴇ-ʟɪɴᴇᴅ ʜᴇᴀᴅ ᴍᴀʀᴋ is used in conjunction with the above sequences, it is placed after the combining sign:

    ᥡ̐ᥡ   ᥡ ɪɴɪᴛɪᴀʟ ᴅᴏᴜʙʟᴇ-ʟɪɴᴇᴅ ʜᴇᴀᴅ ᴍᴀʀᴋ, ◌̐ ꜱɪɢɴ ᴄᴀɴᴅʀᴀʙɪɴᴅᴜ ᴡɪᴛʜ ᴏʀɴᴀᴍᴇɴᴛ,
          ᥡ ᴄʟᴏꜱɪɴɢ ᴅᴏᴜʙʟᴇ-ʟɪɴᴇᴅ ʜᴇᴀᴅ ᴍᴀʀᴋ

As shown in the above folio, the proposed marks may be followed by ‖ ᴅᴏᴜʙʟᴇ ꜱʜᴀᴅ, and theoretically, by | ꜱʜᴀᴅ as well:

    ᥡᥡ‖   ᥡ ɪɴɪᴛɪᴀʟ ᴅᴏᴜʙʟᴇ-ʟɪɴᴇᴅ ʜᴇᴀᴅ ᴍᴀʀᴋ,
          ᥡ ᴄʟᴏꜱɪɴɢ ᴅᴏᴜʙʟᴇ-ʟɪɴᴇᴅ ʜᴇᴀᴅ ᴍᴀʀᴋ,
          ‖ ᴅᴏᴜʙʟᴇ ꜱʜᴀᴅ

Although not attested, the ᴄʟᴏꜱɪɴɢ ᴅᴏᴜʙʟᴇ-ʟɪɴᴇᴅ ʜᴇᴀᴅ ᴍᴀʀᴋ may theoretically be used for producing extended head marks, such as ᥡᥡᥡ, similar to Tibetan ༄༅༅

    ᥡᥡᥡ   ᥡ ɪɴɪᴛɪᴀʟ ᴅᴏᴜʙʟᴇ-ʟɪɴᴇᴅ ʜᴇᴀᴅ ᴍᴀʀᴋ,
          ᥡ ᴄʟᴏꜱɪɴɢ ᴅᴏᴜʙʟᴇ-ʟɪɴᴇᴅ ʜᴇᴀᴅ ᴍᴀʀᴋ,
          ᥡ ᴄʟᴏꜱɪɴɢ ᴅᴏᴜʙʟᴇ-ʟɪɴᴇᴅ ʜᴇᴀᴅ ᴍᴀʀᴋ

3   Character Data

3.1   Character properties

In the format of `UnicodeData.txt`:

```
11A45;ZANABAZAR SQUARE INITIAL DOUBLE-LINED HEAD MARK;Po;0;ON;;;;;N;;;;;
11A46;ZANABAZAR SQUARE CLOSING DOUBLE-LINED HEAD MARK;Po;0;ON;;;;;N;;;;;
```

3.2   Linebreaking

In the format of `LineBreak.txt`:

```
11A45..11A46; BB    # [2] ZANABAZAR SQUARE INITIAL DOUBLE-LINED HEAD MARK ..
                          ZANABAZAR SQUARE CLOSING DOUBLE-LINED HEAD MARK
```

4   Acknowledgments

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646[1]
Please fill all the sections A, B and C below.
Please read Principles and Procedures Document (P & P) from http://std.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://std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html**.**
See also http://std.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html **for latest *Roadmaps*.**

A. Administrative

1. **Title:** *Proposal to encode additional head marks for Zanabazar Square*
2. Requester's name: *Script Encoding Initiative (SEI) / Anshuman Pandey (pandey@umich.edu)*
3. Requester type (Member body/Liaison/Individual contribution): *Liaison contribution*
4. Submission date: *2015-12-30*
5. Requester's reference (if applicable):
6. Choose one of the following:
   This is a complete proposal: *x*
   (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: *x*
      Name of the existing block: *Zanabazar Square*
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"
      in Annex L of P&P document? *yes*
   b. Are the character shapes attached in a legible form suitable for review? *yes*
5. Fonts related:
   a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?
      Anshuman Pandey
   b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):
      Anshuman Pandey (pandey@umich.edu)
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 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)?
   (General category properties and line-breaking properties are included)

8. Additional Information:
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 Unicode Character Database (http://www.unicode.org/reports/tr44/) 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

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? _____Agata Bareja-Starzyńska (University of Warsaw, Poland)_____
   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? _See L2/15-337_
   Reference:

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

5. Are the proposed characters in current use by the user community? _____yes_____
   If YES, where? Reference: _____By scholars of Mongolian culture, history, and linguistics_____

6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely
   in the BMP? _____n/a_____
   If YES, is a rationale provided?
   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, or could be confused with, 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? _____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 characters? _____no_____
    If YES, are the equivalent corresponding unified ideographic characters identified?
    If YES, reference: