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 two Sinological Latin letters to the UCS**
Source: **Michael Everson and Andrew West**
Status: **Individual Contribution**
Date: **2017-08-17**

This proposal requests the encoding of a casing pair of two Latin characters used in Sinological and Sino-Tibetanist phonetic transcription. If this proposal is accepted, the following characters will exist:

ꭦ   AB66   LATIN SMALL LETTER DZ DIGRAPH WITH RETROFLEX HOOK

• used in sinological transcription for a voiced retroflex affricate

ꭧ   AB67   LATIN SMALL LETTER TS DIGRAPH WITH HOOK

• used in sinological transcription for a voiceless retroflex affricate

Although the International Phonetic Alphabet deprecated the letters

U+02A3 ʣ LATIN SMALL LETTER DZ DIGRAPH
U+02A4 ʤ LATIN SMALL LETTER DEZH DIGRAPH
U+02A5 ʥ LATIN SMALL LETTER DZ DIGRAPH WITH CURL,
U+02A6 ʦ LATIN SMALL LETTER TS DIGRAPH
U+02A8 ʨ LATIN SMALL LETTER TC DIGRAPH WITH CURL

in favour of the sequences [dz], [dʒ], [dʑ], [ts], and [tɕ], in the Sinological phonetic tradition the digraphs are the norm and continue to be used. In the two source dictionaries for Naxi Dongba and Naxi Geba, some additional digraphs were discovered; these should be added to the UCS in order to support Sinological practice.

Unicode Character Properties. Character properties are proposed here.

```
AB66;LATIN SMALL LETTER DZ DIGRAPH WITH RETROFLEX HOOK;Ll;0;L;;;;;N;;;;;
AB67;LATIN SMALL LETTER TS DIGRAPH WITH HOOK;Ll;0;L;;;;;N;;;;;
```

Bibliography

Fāng Guóyú (方国瑜). 1995. 纳西象形文字谱 (*Nàxī xiàngxíng wénzì pǔ* 'A dictionary of Naxi pictographic characters'). Kūnmíng: Yúnnán Rénmín chūbǎnshè. ISBN 7-222-01751-8

Lǐ Líncàn (李霖灿). 2001. 纳西族象形标音文字字典 (*Nàxīzú xiàngxíng biāo yīn wénzì zìdiǎn* 'Naxi Pictographic Symbols Dictionary'). Kūnmíng: Yúnnán Mínzú Chūbǎnshè. ISBN 7-5367-2126-9

Figures.

Figure 1. Example from Fāng Guóyú 1995:447 showing ʣ̢ LATIN SMALL LETTER DZ DIGRAPH WITH RETROFLEX HOOK and ʦ LATIN SMALL LETTER TS DIGRAPH WITH HOOK.

Figure 2. Example from Fāng Guóyú 1995:444 showing ʦ LATIN SMALL LETTER TS DIGRAPH WITH HOOK alongside U+02A5 ʥ LATIN SMALL LETTER DZ DIGRAPH WITH CURL.

33 热

　　ʦʽʌ⁵⁵—　讲价

　　——　再热一下（饭）

　　nu⁵⁵ mɛ³³—　热心，动心

31 救上来，取消

　　ʈʂʽwɑ⁵⁵—　星名（昂宿）

ʦʽɯ

55 ⿰ヨ彡、⿰ヨ彡、⿰ヨ彡、⿰ヨ彡、⿰ヨ彡

　　山羊，跪下，以木管吸酒，捏，

　　建立

　　—bu³¹　公山羊

Figure 3. Example from Lǐ Líncàn 2001:486 showing ʈʂ LATIN SMALL LETTER TS DIGRAPH WITH HOOK alongside U+02A6 ʦ LATIN SMALL LETTER TS DIGRAPH.

ʈʂ

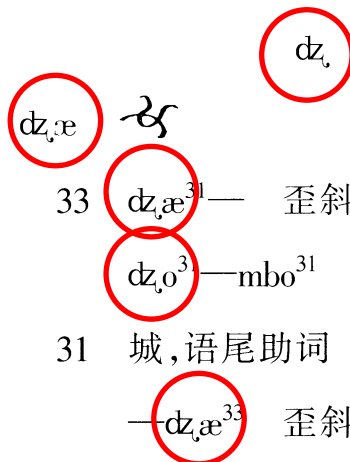ʈʂæ　⿰⿱⿰、⿰、⿰

　　55 舂熟米，放置

　　　　ndzʌ—　接木

　　　　—ʈʂæ³³　重叠

　　　　ndʐʌ³¹—　请神加被"威灵"

　　33 神像画

　　　　ʈʂæ⁵⁵—　重叠

Figure 4. Example from Lǐ Líncàn 2001:502 showing ʈʂ LATIN SMALL LETTER TS DIGRAPH WITH HOOK alongside ɖʐ LATIN SMALL LETTER DZ DIGRAPH WITH RETROFLEX HOOK. In this example the sequence "dz" is used where the digraph "ɖʐ" would be expected. Note the offset from the baseline of ʈʂ and ɖʐ, however, suggesting that this is simply a font matter. There are other font-related typos in this dictionary, such as "nɑʐo³¹" for "ndʐo³¹ on p. 40.

Figure 5. Example from Lǐ Líncàn 2001:506 showing ʣ LATIN SMALL LETTER DZ DIGRAPH WITH RETROFLEX HOOK.

Figure 6. Example from Lǐ Líncàn 2001:486 showing ʣ LATIN SMALL LETTER DZ DIGRAPH WITH RETROFLEX HOOK alongside U+02A5 ʥ LATIN SMALL LETTER DZ DIGRAPH WITH CURL and U+02A8 ʨ LATIN SMALL LETTER TC DIGRAPH WITH CURL

A. Administrative

1. Title
Proposal to add two Sinological Latin letters to the UCS
2. Requester's name
Michael Everson and Andrew West
3. Requester type (Member body/Liaison/Individual contribution)
Individual contribution.
4. Submission date
2017-08-17
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.
1b. Proposed name of script
1c. The proposal is for addition of character(s) to an existing block
Yes
1d. Name of the existing block
Latin Extended-E
2. Number of characters in proposal
2.
3. Proposed category (A-Contemporary; B.1-Specialized (small collection); 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)
Category A.
4a. Is a repertoire including character names provided?
Yes.
4b. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?
Yes.
4c. Are the character shapes attached in a legible form suitable for review?
Yes.
5a. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?
Michael Everson.
5b. If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:
Michael Everson, Fontographer.
6a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?
Yes.
6b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?
Yes.
7. 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)?
Yes.
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 Unicode Character Database http://www.unicode.org/ Public/UNIDATA/UnicodeCharacterDatabase.html and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.
See 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?
Marc Miyake.

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?

Sinologists, Sino-Tibetanists.

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

Common in linguistic texts.

4b. Reference

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

Yes.

5b. If YES, where?

Various publications.

6a. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?

Yes.

6b. If YES, is a rationale provided?

Yes.

6c. If YES, reference

Accordance with the Roadmap. Keep with other Latin characters.

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?

No.

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

No.

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)?

No.

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

11c. If YES, reference

11d. Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?

No.

11e. If YES, reference

12a. Does the proposal contain characters with any special properties such as control function or similar semantics?

No.

12b. If YES, describe in detail (include attachment if necessary)

13a. Does the proposal contain any Ideographic compatibility character(s)?

No.

13b. If YES, is the equivalent corresponding unified ideographic character(s) identified?