JTC1/SC2/WG2 N3817
2010-04-15

Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Organisation Internationale de Normalisation
Международная организация по стандартизации

| | |
|---|---|
| **Doc Type:** | **Working Group Document** |
| **Title:** | **Review of Revised Proposal to Encode Jurchen (N3788)** |
| **Source:** | **Andrew West** |
| **Status:** | **Individual Contribution** |
| **Action:** | **For consideration by JTC1/SC2/WG2 and UTC** |
| **Date:** | **2010-04-15** |

---

# 1. Character Names

## A. Errors in the Source Dictionary

The following corrections of errors in the phonetic transcriptions in the source dictionary should be reflected in the character names:

- 012-02 X-0051 𠂤 ( 湯古) *tanŋgu* should be *taŋgu*
- 021-05 X-0108 𠂤 ( 湯古) *tanŋgu* should be *taŋgu*
- 023-01 X-0114 孟 (希兒) *çïr* should be *çir*
- 041-05 X-0194 𠂢 (只里) *ʤiri* should be *ʤïri*
- 045-03 X-0216 𠂡 (哈称) *xatʃi* should be *xatʃï*
- 084-07 X-0405 𢀛 (間) *giyen* should be *gijen*
- 118-01 X-0557 𡜐 (先) *ʃiɛn* should be *çiɛn*
- 138-02 X-0639 𢆶 (謙) *ʧiɛn* should be *ʨiɛn*
- 168-02 X-0766 匆 (申) *ʃin* should be *ʃïn*
- 169-03 X-0771 𢘑 *futici*? should be *futʃïçi*
- 178-01 X-0809 𢘑 *futici*? should be *futʃïçi*
- 178-02a X-0810a 𢘑 *futici*? should be *futʃïçi*
- 207-01 X-0952 𠂢 (只速) *ʤisu* should be *ʤïsu*
- 223-01 X-1040 𠂢 (厥) *kiye* should be *kije*

## B. Simplification of the Phonetic Transcription

The phonetic reconstructions used in the source dictionary are strictly phonetic, and so use different phonetic symbols to represent the same phoneme in different contexts.

The following phonetic symbols only occur before back and mid vowels:

- *ʃ* before *a*, *o*, *u*, *ə*, *ɨ*
- *ʧ* before *a*, *o*, *u*, *ə*, *ɨ*
- *ʤ* before *a*, *o*, *u*, *ə*, *ɨ*
- *ʒ* before *u*

Whereas the following corresponding phonetic symbols only occur before front vowels:

- *ç* before *i*, *y*
- *ʨ* before *i*
- *ʣ* before *i*, *y*

The vowels *e* and *ə* are also in complementary distribution:

- *e* only occurs as *ei*, *ie* and *je*
- *ə* never occurs before/after *i* or after *j*

We believe that there is no need to preserve such non-phonemic distinctions in the character names, and suggest representing pairs of phonetic symbols that represent the same phoneme and that are in complementary distribution (i.e. *ʃ*/*ç*, *ʧ*/*ʨ*, *ʤ*/*ʣ* and *e*/*ə*) using the same ASCII letter or letters, for example using 'E' for both *e* and *ə*.

### C. Use of Accented Letters in Character Names

The character names proposed in N3788 include extended characters (e.g. Ś, Š, Ź, Ž, È), which are disallowed according to the character naming rules. We propose the following scheme for translating the phonetic transcription in the source dictionary to legal character names (all unlisted characters are unchanged):

- *e*/*ə* = E
- *ɛ* = AE
- *ŋ* = NG
- *ʣ*/*ʤ* = J
- *ʨ*/*ʧ* = C
- *ç*/*ʃ* = SH
- *ʒ* = ZH
- *j* = Y (in complementary distribution with 'Y' representing the vowel [y])

We prefer "J" and "C" to "DZH" and "TSH" as Manchu romanization uses "j" and "c", so the resultant Jurchen names appear much closer to their Manchu cognates.

## 2. Wrongly Ordered Characters

The following characters appear to be misordered:

- X-0663b (Radical 7 / 4 strokes) should be moved to after X-0218
- X-0894a (Radical 32 / 6 strokes) should be moved to after X-0931d
- X-0454 (Radical 22 / 8 strokes) and X-0455 (Radical 21 / 8 strokes) should be swapped
- X-0993 (214-08) and X-0994 (214-09) should be swapped

## 3. Duplicate Characters

N3788 proposes encoding the character in entry 048-03 (X-0231a and X-0231b) twice as it is a variant form of two different characters (048-02 and 051-01). We believe that this character should not be encoded twice.

## 4. Radical 21

The glyph form of Radical 21 艹 does not exactly match the shape of the radical in the characters under radical 21. The radical glyph has crossing horizontal strokes, whereas the characters under this radical do not:

- 092-05 X-0452 芙
- 092-06 X-0453 芊
- 093-02 X-0455 芮

We believe that although the difference is minor, it would be best to change the glyph for Radical 21 to match its actual appearance so that it is more distinct from Radical 20:

| Radical 20 | Radical 21 |
|---|---|
| 艹 | 艹 |

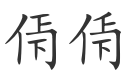
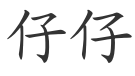
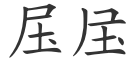
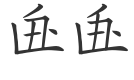
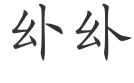
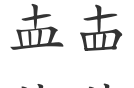
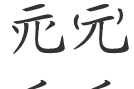
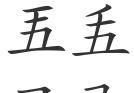
## 5. Glyph Variants

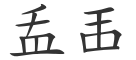
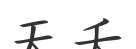
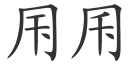
The revised proposal (N3788) proposes 1,430 characters for encoding, of which:

- 523 characters have no variant forms (including nearly a hundred characters with unknown readings that may in fact be variants of other characters proposed for encoding)
- 368 characters are the primary form of a character with multiple glyph variants
- 539 characters are glyph variants

Thus, almost 40% of the proposed Jurchen characters are glyph variants, which is a considerably higher proportion than for any other script encoded or proposed for encoding in the UCS. Moreover, very many of the glyph variants proposed for encoding show very insignificant differences, often just reflecting a slightly different way of writing the same stroke in different manuscript sources, as for example:

| | |
|---|---|
| AI | 帀 帀 帀 |
| AMBA | 夆 夆 |
| AN | 米 米 |
| AXU | 尅 尅 |

| | |
|---|---|
| BIRA |  |
| BUXA |  |
| CII |  |
| CUEN |  |
| DAI |  |
| ESE |  |
| FUN |  |
| GE |  |
| FI |  |
| FO |  |
| GE |  |
| GU |  |
| I |  |
| INDA |  |
| IU |  |
| JAL |  |
| JO |  |
| JU |  |
| KI |  |
| MA |  |
| MI |  |
| MINGGAN |  |
| MO |  |
| MUA |  |

| | |
|---|---|
| NADAN |  |
| NAN |  |
| O |  |
| SHI |  |
| SHIA |  |
| SHII |  |
| SHIIR |  |
| TAIYI |  |
| TU |  |
| U |  |
| UJE |  |
| XA |  |
| YA |  |

It may be appropriate to separately encode some significant character variants, but it may be more appropriate to represent insignificant glyph variants by means of variation sequences.

In several respects, the nature of the Jurchen character variants is significantly different to that of Tangut character variants, which means that although variation sequences are not appropriate in the case of Tangut, they may be the most appropriate solution for Jurchen:

| Tangut | Jurchen |
|---|---|
| Relatively few variant characters (about 120 out of the 6,054 characters proposed for the main Tangut block) | Relatively many variant characters (539 out of 1,430 proposed characters) |
| Most variants are attested in more than one modern Tangut dictionary | Most variants are only used in Jin Qicong's dictionary |
| Most Tangut variants are distinguished by a different structural composition (one or more different component elements, or the same component elements arranged differently) | Most Jurchen variants are distinguished only by a different way of writing the same stroke (e.g. the angle or length of a stroke) or a different placement of a stroke or a difference in stroke count |