







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

Doc Type: Working Group Document
Title: Glyph Corrections for 3 Tangut ideographs
Source: Andrew West, Viacheslav Zaytsev, Sun Bojun
Status: Individual Contribution
Action: For consideration by JTC1/SC2/WG2 and UTC
Date: 2016-04-21

1. Introduction

Document WG2 N4588 identified many Tangut characters with incorrectly written glyphs in modern sources. This document identifies a further three Tangut characters that are incorrectly written in Li Fanwen's 2008 *Tangut-Chinese Dictionary*. As the font for the Unicode and ISO/IEC 10646 code charts is based on the font used in this dictionary, the characters in the code charts are also incorrectly written, and need to be corrected, as shown in Table 1.

Table 1. Summary of Glyph Corrections

Character	Current Glyph	Proposed New Glyph	Old Stroke Count	New Stroke Count
U+17013			9	10
U+17712			13	14
U+17D9F			12	12



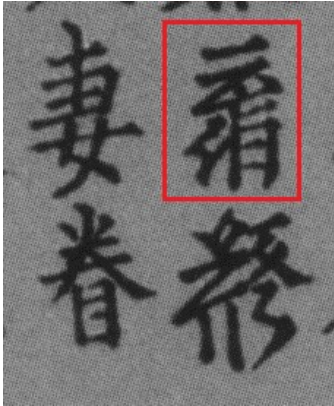
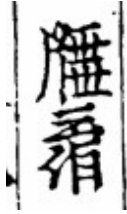

2. Glyph Corrections

2.1 U+17013


In modern sources it is not always clear exactly how U+17013 𠬞 ‘to request’ is written. In the font used for Li Fanwen’s 2008 *Tangut-Chinese Dictionary* (and for the ISO/IEC 10646 and Unicode code charts), this character is written with a single dot on the right of the Z-shaped stroke, and the central vertical stroke extends up into the Z-shaped stroke.

Nishida 1966	Sofronov 1968	Li Fanwen 1986	Han Xiaomang 2004	Kychanov 2006	Li Fanwen 2008	Proposed New Glyph
 35-061	 1124	 2273	 1992/2070	 1514	 0147	



However, original printed Tangut sources show that this character written with an extra dot in the centre of the Z-shaped stroke, and with the central vertical stroke not extending into the Z-shaped stroke. Only the example from the *Odes* shows a form that possibly matches the current code chart glyph (it is not clear).




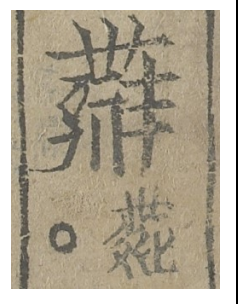
Tongyin A 40B58	Tongyin B2 41A34	Zhangzhongzhu A 343B	Zazi 15B	Odes No. 4 folio 8b
				

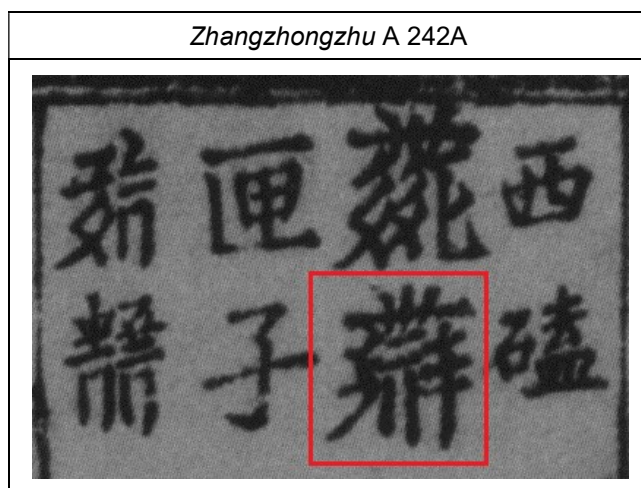
2.2 U+17712

In most modern sources U+17712 𪗇 ‘casket’ *kha* is written as , as shown below. This is also the form used in the Unicode and ISO/IEC 10646 code charts.




Nishida 1966	Sofronov 1968	Li Fanwen 1986	Han Xiaomang 2004	Kychanov 2006	Li Fanwen 2008	Proposed New Glyph
						
17-09e	0908	2409	1614/1711	0454	4201	

U+17712 is a rare character, and we only know of two attestations in original Tangut texts, once in *Tongyin* (‘Homophones’) and once in *Zhangzhongzhu* (‘Pearl in the Palm’). In these texts the character is written as , with a different lower middle component. The only source which shows the  component is the A edition of *Tongyin*.

<i>Tongyin</i> A 27A14	<i>Tongyin</i> B1 27B36	<i>Tongyin</i> B2 27B36	<i>Tongyin</i> D 27B36
			









U+17712 is evidently constructed using the left and middle components of U+1765D 𪛗 'sound' *kha* as a phonetic. Both characters are read *kha*, and U+1765D is the character immediately preceding U+17712 in *Tongyin* ('Homophones'). In the A edition both U+1765D and U+17712 are written with the 𪛗 component, whereas in the B edition both characters are written with the 𪛗 component, indicating that the same component occurs in both characters.

Tongyin A 27A13-14	Tongyin B1 27B35-36	Tongyin B2 27B35-36	Tongyin D27B35-36
			





U+1765D 𪛗 'sound' *kha* is constructed using the left side of U+1764F 𪛗 'sound' as a semantic, and the middle and right components of U+186CB 𪛗 'bitter lettuce' *kha* as a phonetic, which indicates that the left side component of U+1765D must be 𪛗, as shown in the B and D editions of *Tongyin*. As U+17712 uses the left and middle components of U+1765D 𪛗 as its phonetic, this clearly demonstrates that the lower middle component of U+17712 𪛗 must also be 𪛗 rather than 𪛗.


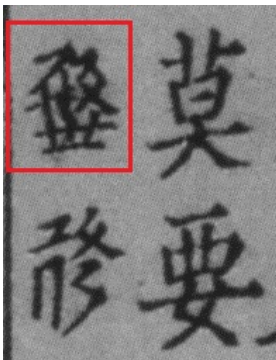

2.3 U+17D9F

In modern sources U+17D9F 𪚩 ‘ceremony’ is written either as 𪚩 𪚩 𪚩 or 𪚩 𪚩 𪚩, as shown below. The former form is used in the Unicode and ISO/IEC 10646 code charts.

Nishida 1966	Sofronov 1968	Li Fanwen 1986	Han Xiaomang 2004	Kychanov 2006	Li Fanwen 2008	Proposed New Glyph
 143-081	 4826	 4221	 3286/3406	 0049	 0906	

However, original Tangut sources all use the 𪚩 component. Compare U+180A8 𪚪 which has U+17D9F as its right hand component, but is correctly written with 𪚩 in the code charts.

Tongyin A 7B47	Tongyin B1 10A37	Tongyin B2 10A37	Tongyin D 10A37
			

Wenhai 1.11.121	Zhangzhongzhu A 291B	Odes No. 5 folio 2A
		

3. Tangut Data

The following entries in TangutSrc.txt (ISO/IEC 10646) and TangutSources.txt (Unicode) need to be modified as shown to account for changes in stroke count for two of the three characters:

```
U+17013    kTGT_MergedSrc L2008-0147
U+17013    kRSTUnicode    1.10
U+17712    kTGT_MergedSrc L2008-4201
U+17712    kRSTUnicode    106.14
```