

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 Eight Tangut Ideographs
Source: Andrew West
Status: Individual Contribution
Action: For consideration by JTC1/SC2/WG2 and UTC
Date: 2023-03-01

This document requests glyph corrections for the eight Tangut ideographs listed in the table below.

Table 1. Summary of Glyph Corrections for 8 Tangut Ideographs

Code Point	Current Glyph	New Glyph	Current Radical/ Strokes (V. 15.0)	Proposed New Radical/ Strokes
U+17105	𐰇	𐰇	11.10	11.9
U+172A4	𐰇	𐰇	36.8	36.8
U+17BD1	𐰇	𐰇	169.9	169.9
U+17BD2	𐰇	𐰇	169.12	169.12
U+17BD3	𐰇	𐰇	169.13	169.13
U+17EF9	𐰇	𐰇	260.17	260.17
U+18136	𐰇	𐰇	316.10	316.10
U+18171	𐰇	𐰇	316.15	316.15

Three of the proposed changes (17BD1..17BD3) show a structural repositioning of the component next to the component 𠂇 which reflects the actual glyph forms found in the primary sources. The other five proposed changes relate to supplementary characters added to the 2008 edition of Lǐ Fànwén's Tangut-Chinese dictionary (pages 945–952, nos. 5995–6074) which were mostly sourced from 𠂇𠂇𠂇𠂇 *Joined Rimes of the Fives Sounds* (surviving in six partial manuscript versions) and some other manuscript texts which we had not previously carefully checked. The glyph forms for these eighty supplementary characters have now all been verified as best as we can given the poor quality of some of the hand-written character forms.

Tangut Data Changes

One of the proposed glyph corrections involve a change in stroke count, which if accepted would necessitate a change in the Tangut data files for the Unicode and ISO/IEC 10646 standards. The affected line in the data files is listed below.

Current (15.0):

U+17105	kRSTUnicode	11.10
---------	-------------	-------

New:

U+17105	kRSTUnicode	11.9
---------	-------------	------

Bibliography

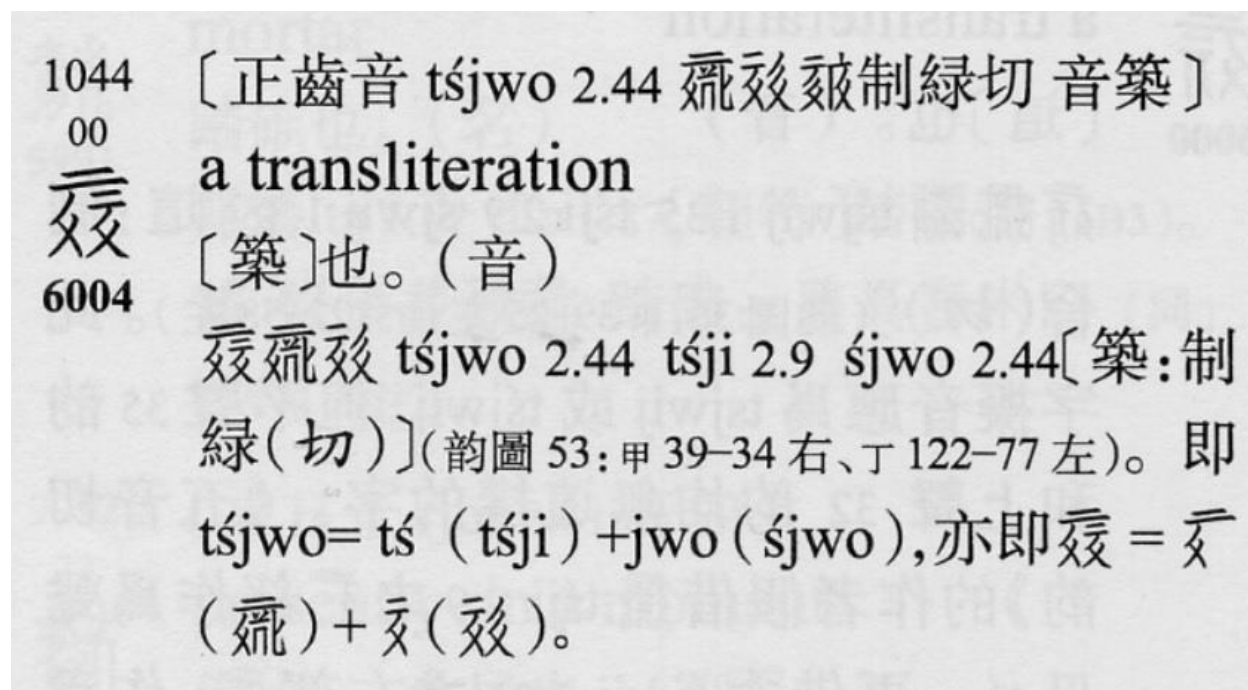
Lǐ Fànwén 李範文 (comp.) and Jiǎ Chángyè 賈常業 (rev. and exp.). 2008. *Xià-Hàn zìdiǎn* 夏漢字典 [Tangut-Chinese dictionary (revised ed.)]. Běijīng 北京: Zhōngguó shèhuì kēxué chūbǎnshè 中国社会科学出版社. ISBN 978-7-5004-2113-9

Evidence for Glyph Corrections

U+17105 𪛗

U+17105 𪛗 *tśjwo* is only attested in a single manuscript source, 𪛗𪛗𪛗𪛗 *Joined Rimes of the Fives Sounds*, and is entry #6004 in the 2008 edition of Lǐ Fànwén's Tangut-Chinese dictionary (see Fig. 1).

Fig. 1. Lǐ Fànwén, *Xià-Hàn zìdiǎn* (2008) p. 946 entry for #6004



Lǐ Fànwén explains the character as being constructed from the top and left of 𪛗 *tśji*, and the left side of 𪛗 *śjwo* [sic], with the *fǎnqiè* 反切 spelling of *tśjwo* from these two characters. However, Lǐ Fànwén here confuses the two graphically similar characters 𪛗 [𪛗 𪛗] *śjwo* and 𪛗 [𪛗 𪛗] *ljw*, (which together form the word 𪛗𪛗 *śjwo-ljw*), and the glyph for the second source character (*śjwo*) should actually be 𪛗 not 𪛗; thus the bottom right component of U+17105 should be 𪛗 not 𪛗.

This analysis is confirmed by the extant manuscript versions of *Joined Rimes of the Fives Sounds* which do indeed show 𪛗 (see Figs. 2 and 3).

Fig. 2. IOM Tang. 22, Old Inv. No. 620 (𪛗𪛘𪛙𪛚 MS A) folio 34B

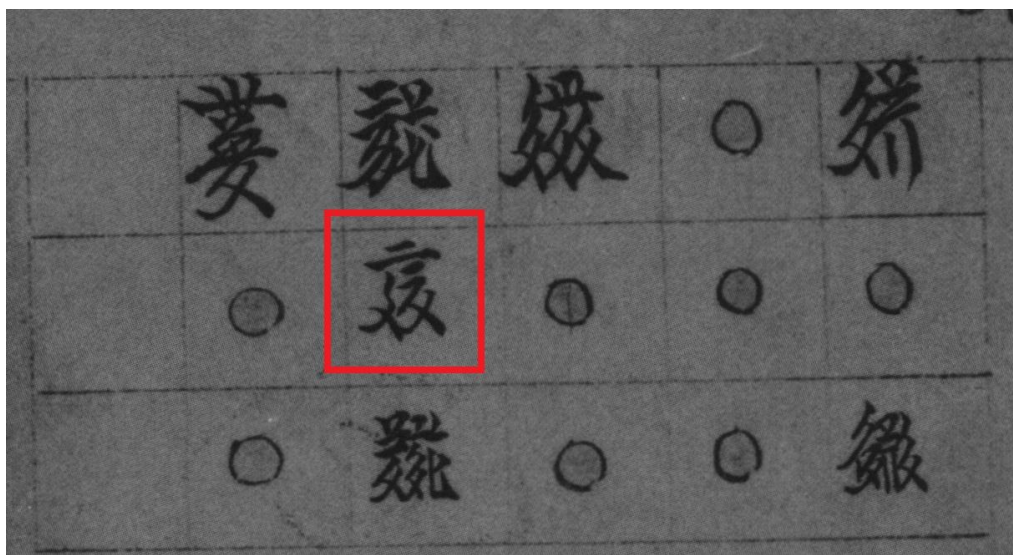


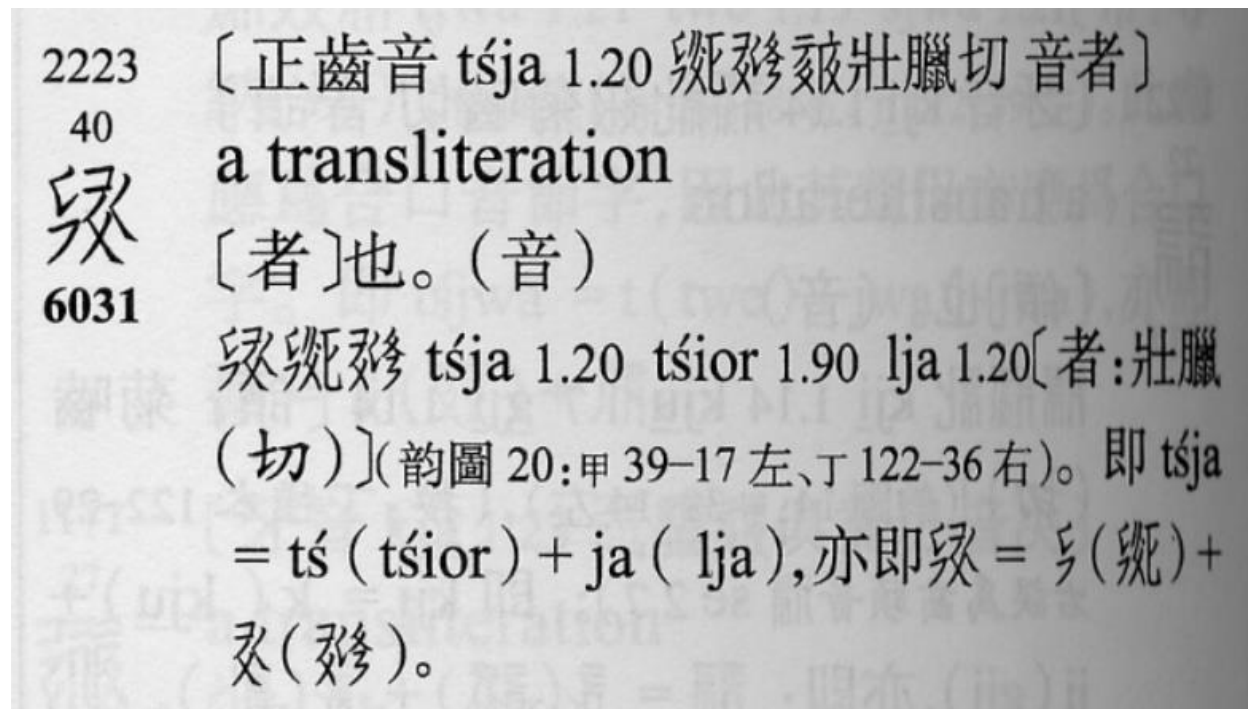
Fig. 3. IOM Tang. 22, Old Inv. No. 621 (𪛗𪛘𪛙𪛚 MS B) folio 14A



U+172A4 𪛗

U+172A4 𪛗 *tšja* is only attested in a single manuscript source, 𪛗𪛗𪛗𪛗 *Joined Rimes of the Fives Sounds*, and is entry #6031 in the 2008 edition of Lǐ Fànwén's Tangut-Chinese dictionary (see Fig. 4).

Fig. 4. Lǐ Fànwén, *Xià-Hàn zìdiǎn* (2008) p. 948 entry for #6031



Lǐ Fànwén explains the character as being constructed from the left side of 𪛗 *tšior* and the left side and middle part of 𪛗 *ljwa*, with the *fǎnqiè* 反切 spelling of *tšja* from these two characters. This implies that the right component of 𪛗 should be 𪛗 rather than 𪛗. This is confirmed by the manuscript versions of *Joined Rimes of the Fives Sounds* which all clearly show 𪛗 on the right side (see Figs. 5 and 6 for two of the five extant examples). This glyph shape contrasts with the glyph shape for 𪛗 seen in printed and manuscript sources (sometimes 3-shaped):

U+17B03 𪛗
(printed form)



Homophones 𪛗𪛗
Edition B5 32B47

U+17B03 𪛗
(manuscript form)



*Joined Rimes of the
Fives Sounds*
𪛗𪛗𪛗𪛗 MS A 12A

Fig. 5. IOM Tang. 22, Old Inv. No. 620 (𪛗𪛘𪛙𪛚 MS A) folio 18A

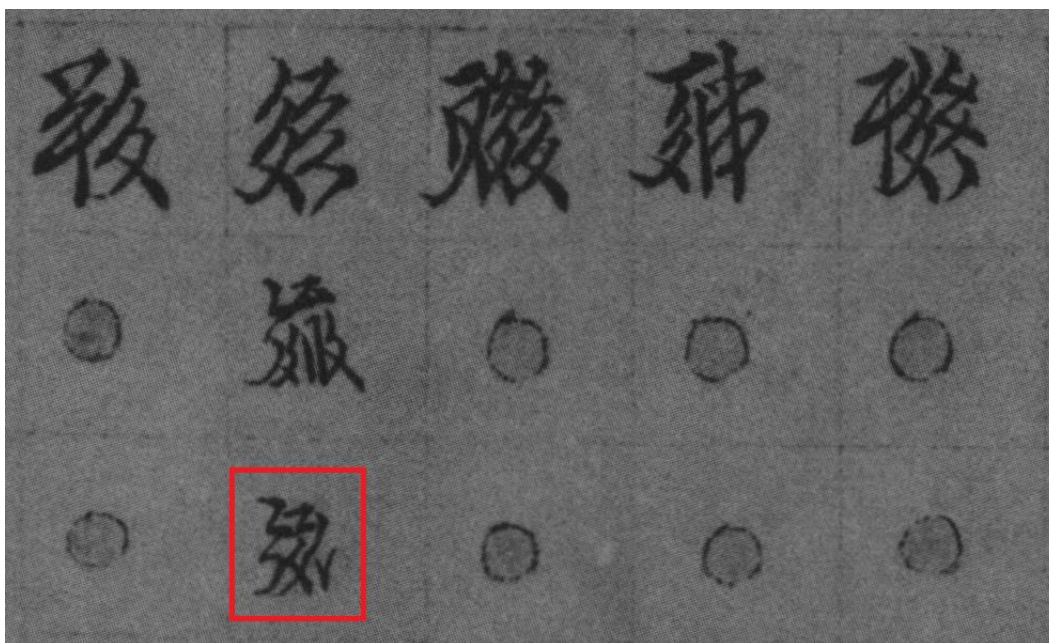
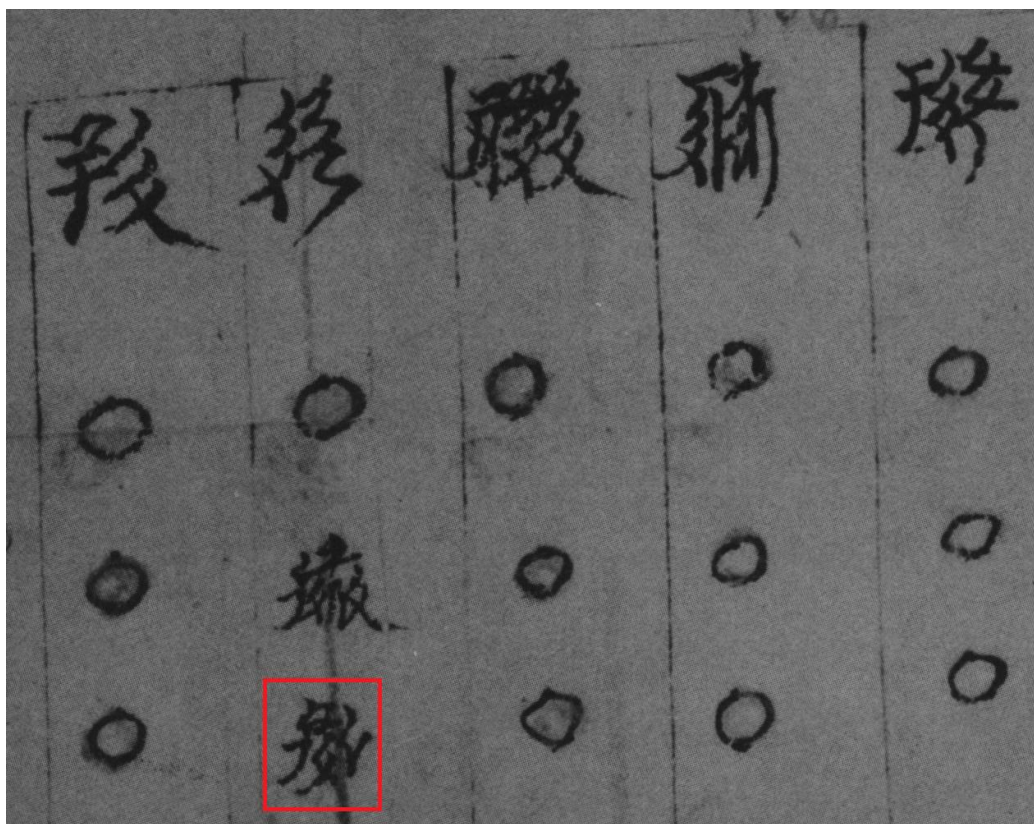


Fig. 6. IOM Tang. 22, Old Inv. No. 621 (𪛗𪛘𪛙𪛚 MS B) folio 16B



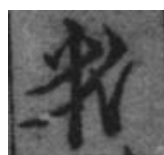
The only other character with 𠂔 on the right side is U+17BD4 𠂔, and the woodblock printed editions of *Homophones* confirm that the right side is indeed the same shape as seen for U+172A4 in the manuscript versions of *Joined Rimes of the Fives Sounds* (i.e. 𠂔):

U+17BD4 𠂔



Homophones 𠂔
Edition A 50A65

U+17BD4 𠂔



Homophones 𠂔
Edition B2 50B53

U+17BD1, U+17BD2, and U+17BD3 𪛗𪛘𪛙

In the 2008 edition of Lǐ Fànwén's Tangut-Chinese dictionary the component to the right of 𪛗 in the three characters U+17BD1, U+17BD2, and U+17BD3 is written full-height, i.e. 𪛗 𪛗, 𪛗 𪛗 𪛗 𪛗, and 𪛗 𪛗 𪛗 𪛗 respectively (see Figs. 7, 8 and 9).

Fig. 7. Lǐ Fànwén, *Xià-Hàn zìdiǎn* (2008) p. 921 entry for #5848

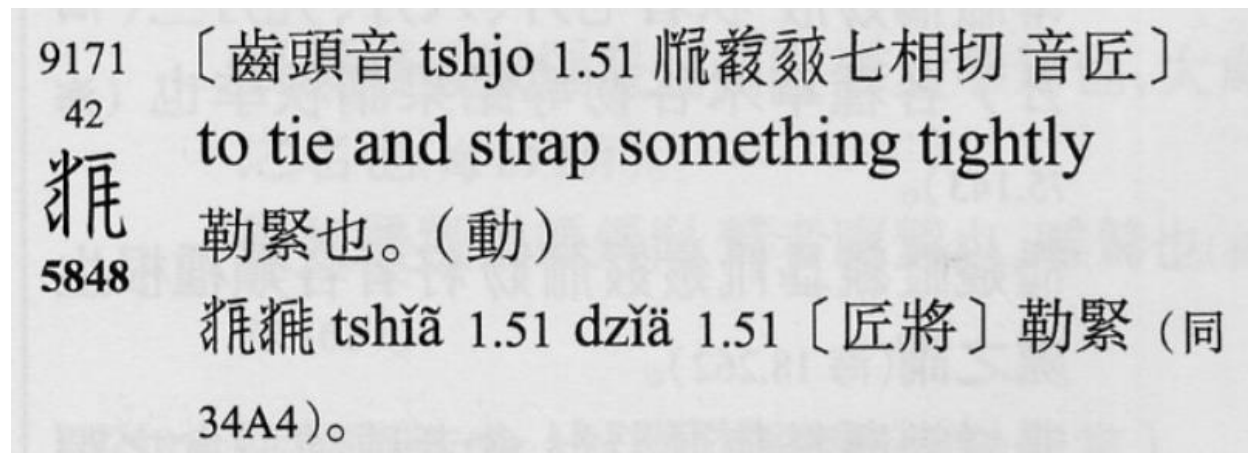


Fig. 8. Lǐ Fànwén, *Xià-Hàn zìdiǎn* (2008) p. 922 entry for #5854

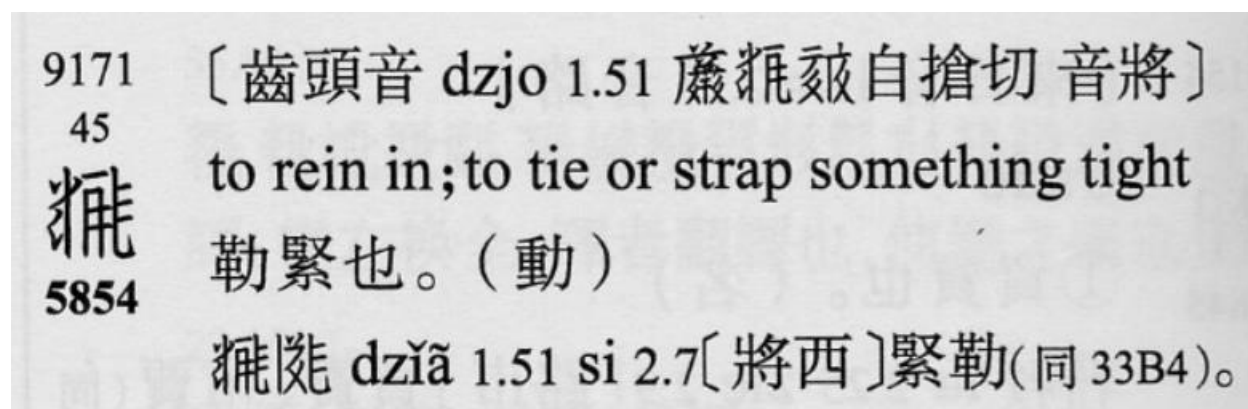
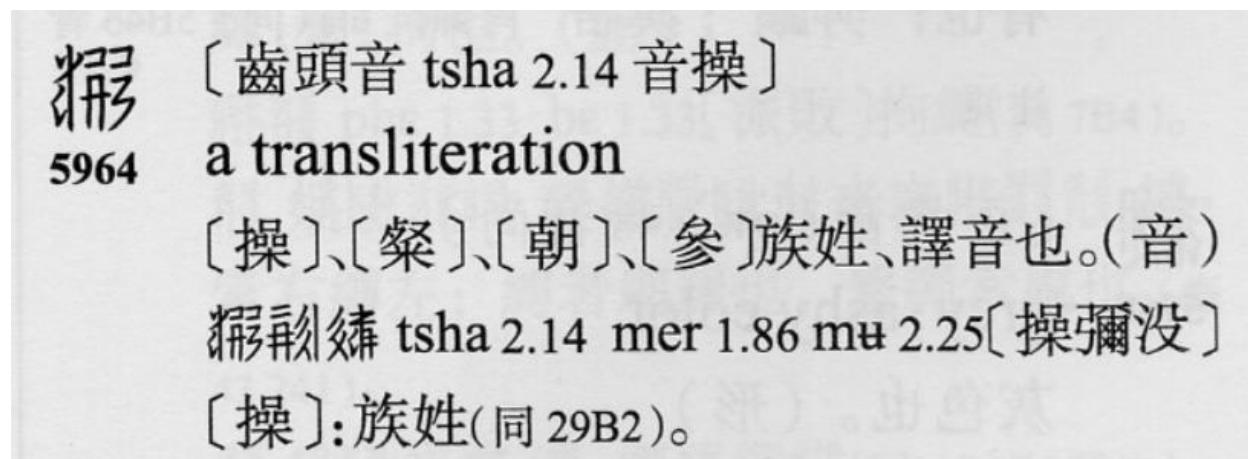
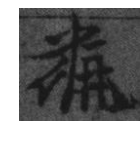


Fig. 9. Lǐ Fànwén, *Xià-Hàn zìdiǎn* (2008) p. 939 entry for #5964

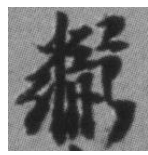


However, almost all primary sources show that the right component of U+17BD1 and the middle components of U+17BD2 and U+17BD3 are tucked into the gap of 𠂇, i.e. 𠂇𠂇𠂇, 𠂇𠂇𠂇𠂇, and 𠂇𠂇𠂇𠂇𠂇 respectively. Examples from a variety of sources are shown below. Only *Homophones* Edition A shows forms that are not completely tucked into 𠂇, but the A edition is generally less reliable with regard to glyph forms than the B edition of *Homophones*.

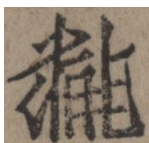
U+17BD1

<i>Sea of Writing</i> 1:58.161	<i>Homophones</i> A 34A45	<i>Homophones</i> B 34B44	<i>Pearl in the Palm</i> A 224E	<i>Joined Rimes of the Five Sounds</i> A 6B	<i>Joined Rimes of the Five Sounds</i> B 5A
					

U+17BD2

<i>Homophones</i> A 29B28	<i>Homophones</i> B 30A65	<i>Grains of Gold</i> 5A4	<i>Forest of Categories</i> 4:01A3	<i>Mixed Characters</i> 11B7
				

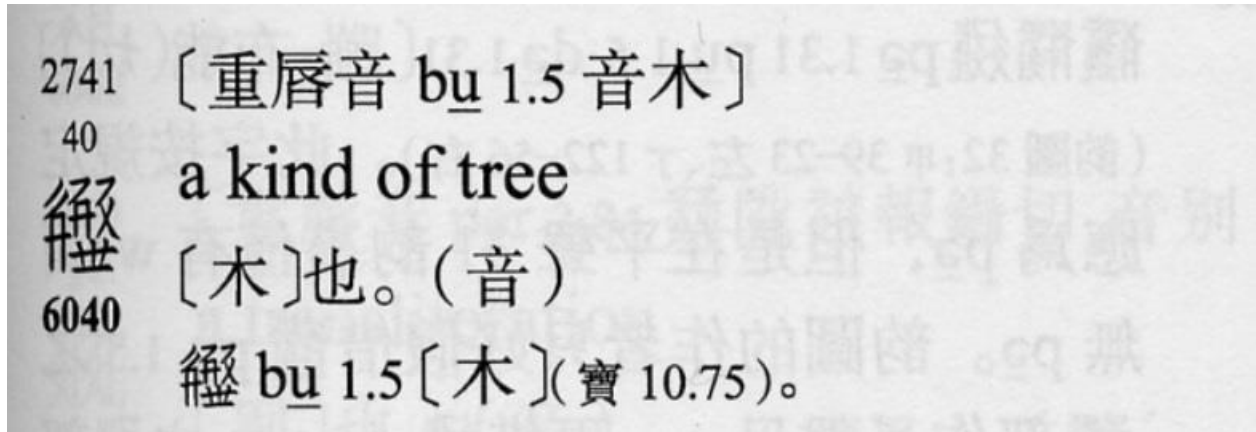
U+17BD3

<i>Sea of Writing</i> 3:5.143	<i>Homophones</i> A 33B47	<i>Homophones</i> B 34A57	<i>IOM Tang.27</i> 29B
			

U+17EF9 𣎵

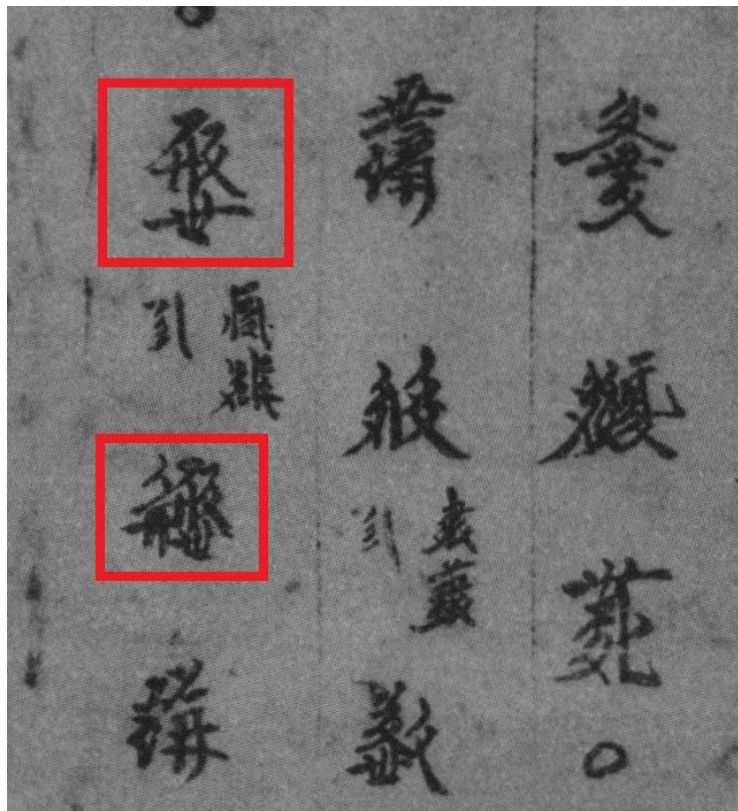
U+17EF9 𣎵 *bu* is written as 𣎵 on the left side and U+17D9F 𣎵 *bu* on the right. U+17D9F was originally written as 𣎵 in Unicode 9.0, but was corrected to 𣎵 in Unicode 10.0 (see WG2 N4723; L2/16-112). However, we forgot to correct the glyph for U+17EF9 at the same time.

Fig. 10. Lǐ Fànwén, *Xià-Hàn zìdiǎn* (2008) p. 949 entry for #6040



U+17EF9 only occurs in a single manuscript source, 𣎵 𣎵 𣎵 *Precious Rimes of the Sea of Writing*, where U+17D9F and U+17EF9 are indeed written as 𣎵 and 𣎵 respectively.

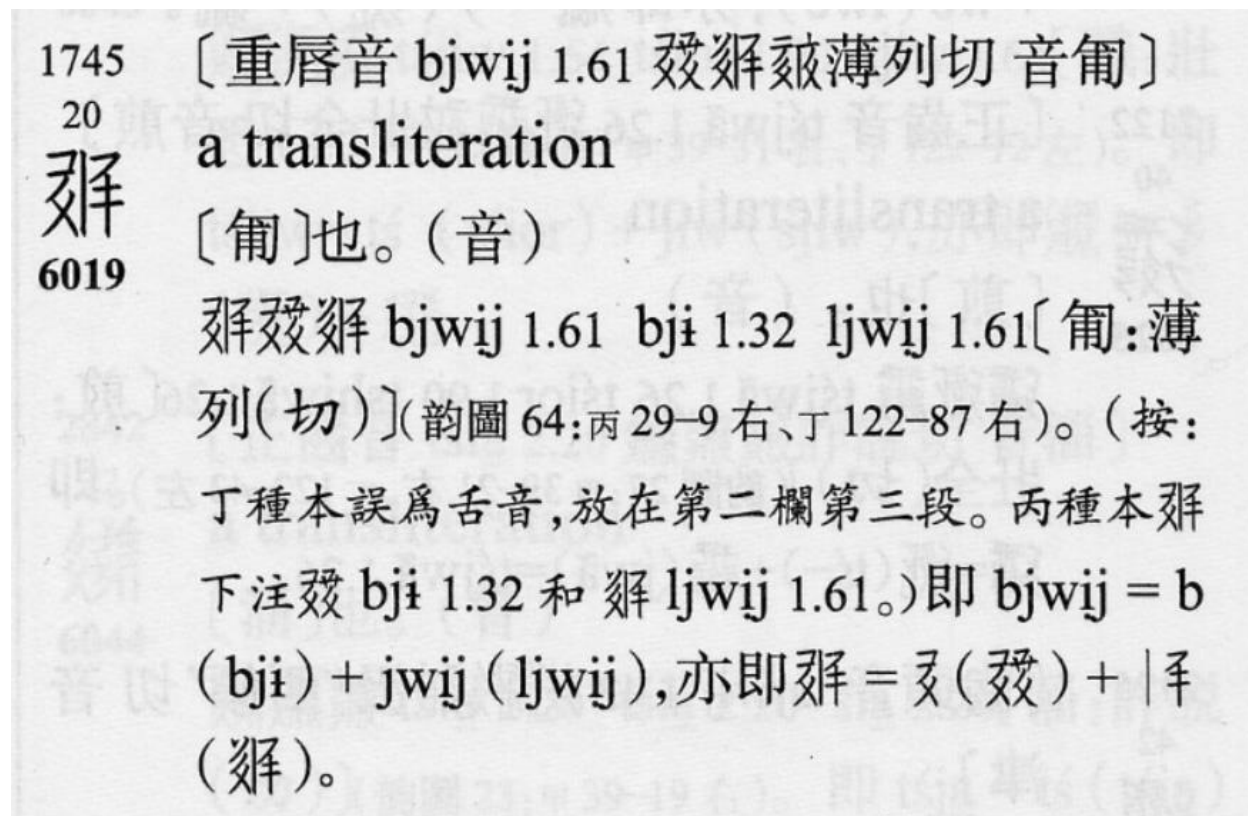
Fig. 11. IOM, Old Inv. No. 4154 8364 (𣎵 𣎵 𣎵) vol. 1 folio 6a



U+18136 𪛗

U+18136 𪛗 *bjwij* is only attested in a single manuscript source, 𪛗 𪛗 𪛗 𪛗 *Joined Rimes of the Fives Sounds*, and is entry #6019 in the 2008 edition of Lǐ Fànwén's Tangut-Chinese dictionary (see Fig. 12).

Fig. 12. Lǐ Fànwén, *Xià-Hàn zìdiǎn* (2008) p. 947 entry for #6019



U+18136 originally had the glyph form 𪛗 (with Component 316 𪛗 on the left) in Unicode 9.0, but in Unicode 14.0 the glyph was changed to 𪛗 (with Component 768 𪛗 on the left). This change was based on Lǐ Fànwén's explanation that the character is constructed from the left side of U+18156 𪛗 *bji* (with Component 768 𪛗 on the left) and the right side of U+180AD 𪛗 *ljwij*, with the *fǎnqiè* 反切 spelling of *bjwij* from these two characters.

The evidence for this construction and spelling comes from the C manuscript of *Joined Rimes of the Fives Sounds* which gives 𪛗 (with left side component unclear) in the first column (for labial initials), below which are two characters giving its *fǎnqiè* spelling (see Fig. 13). These two *fǎnqiè* characters are not well-written, but Lǐ Fànwén reads them as 𪛗 *bji* and 𪛗 *ljwij*. On the other hand, the D and E manuscripts put 𪛗 in the second column (for dental initials), and the glyphs clearly show Component 316 𪛗 on the left (see Figs. 14 and 15). It is very possible that the C manuscript has placed 𪛗 in the wrong column, and the first *fǎnqiè* character (below right of 𪛗) should actually be read as U+18147 𪛗 *tji* (with Component 316 𪛗 on the left) giving the *fǎnqiè* reading of *tjwij*. Regardless of which interpretation is correct, the actual glyph forms shown in Figs. 14 and 15 write the left side component as 𪛗, and so the glyph for U+18136 should be reverted to 𪛗.

Fig. 13. IOM Tang. 22, Old Inv. No. 622 (𪛗龍𪛗𪛗 MS C) folio 9B

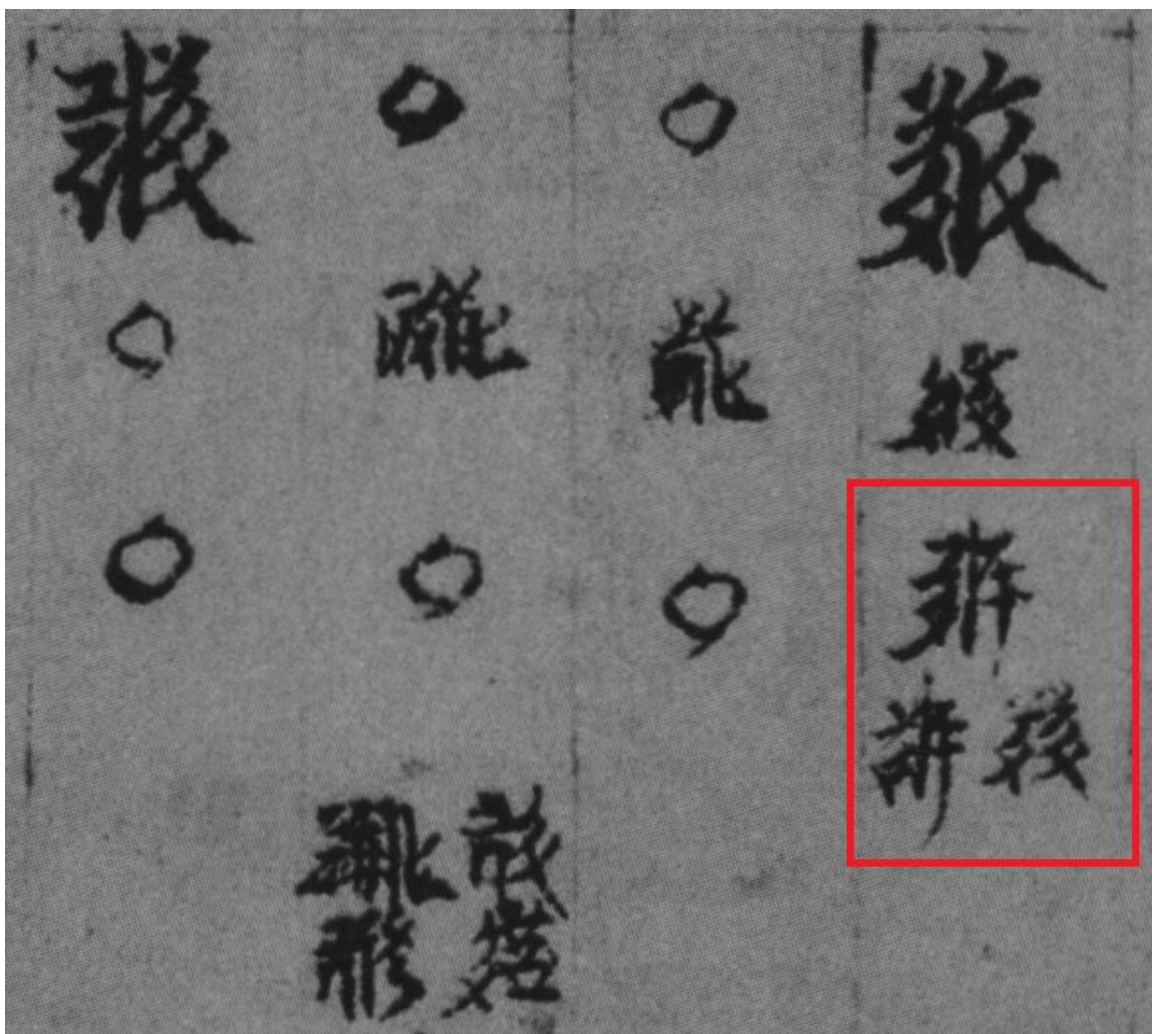


Fig. 14. IOM Tang. 22, Old Inv. No. 623 (𪛗𪛘𪛙𪛚 MS D) folio 68B

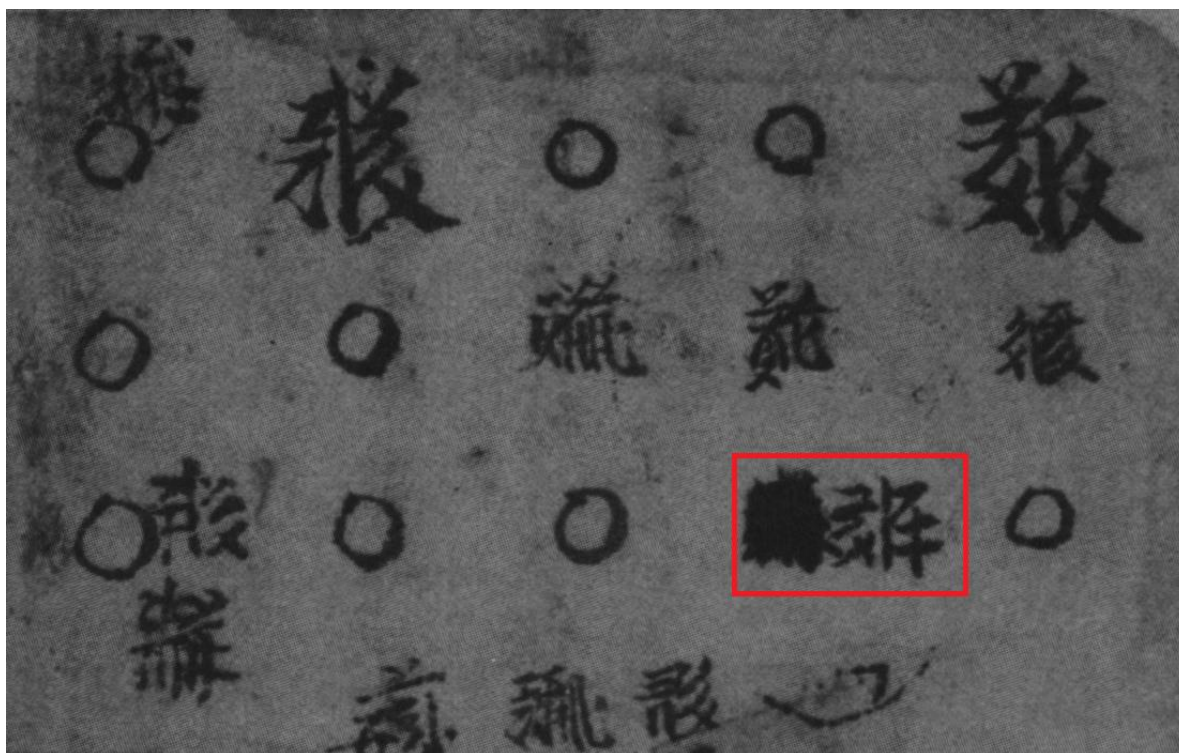
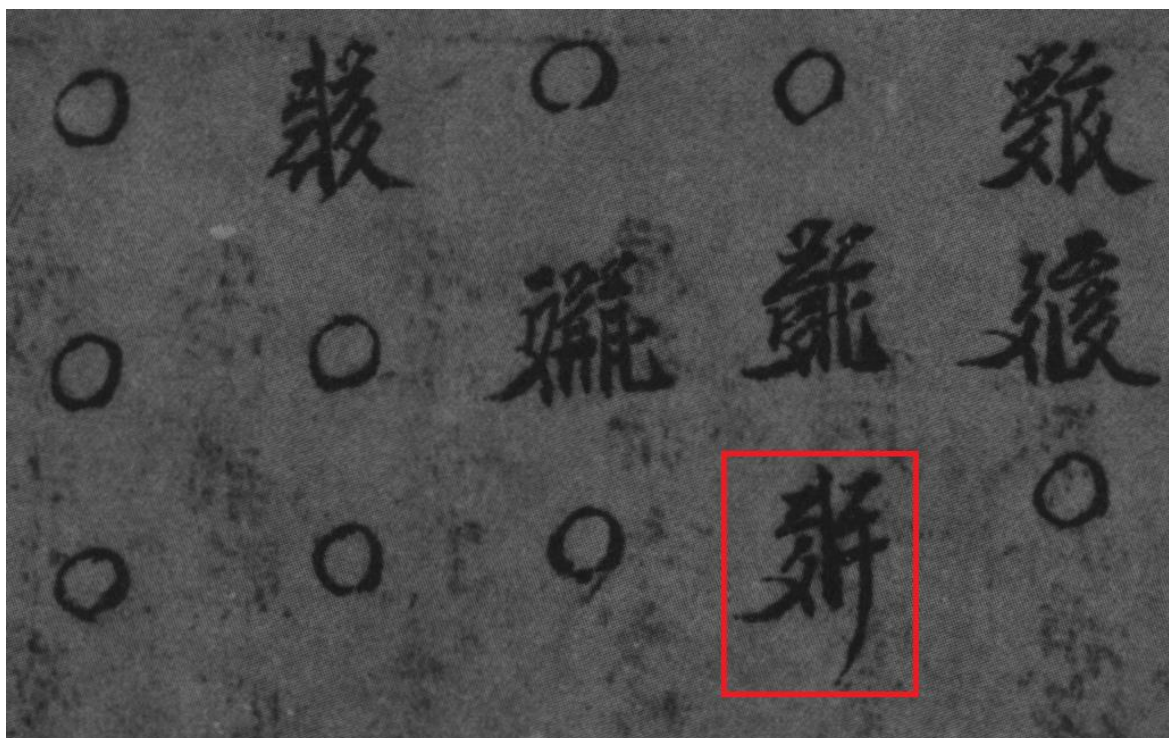


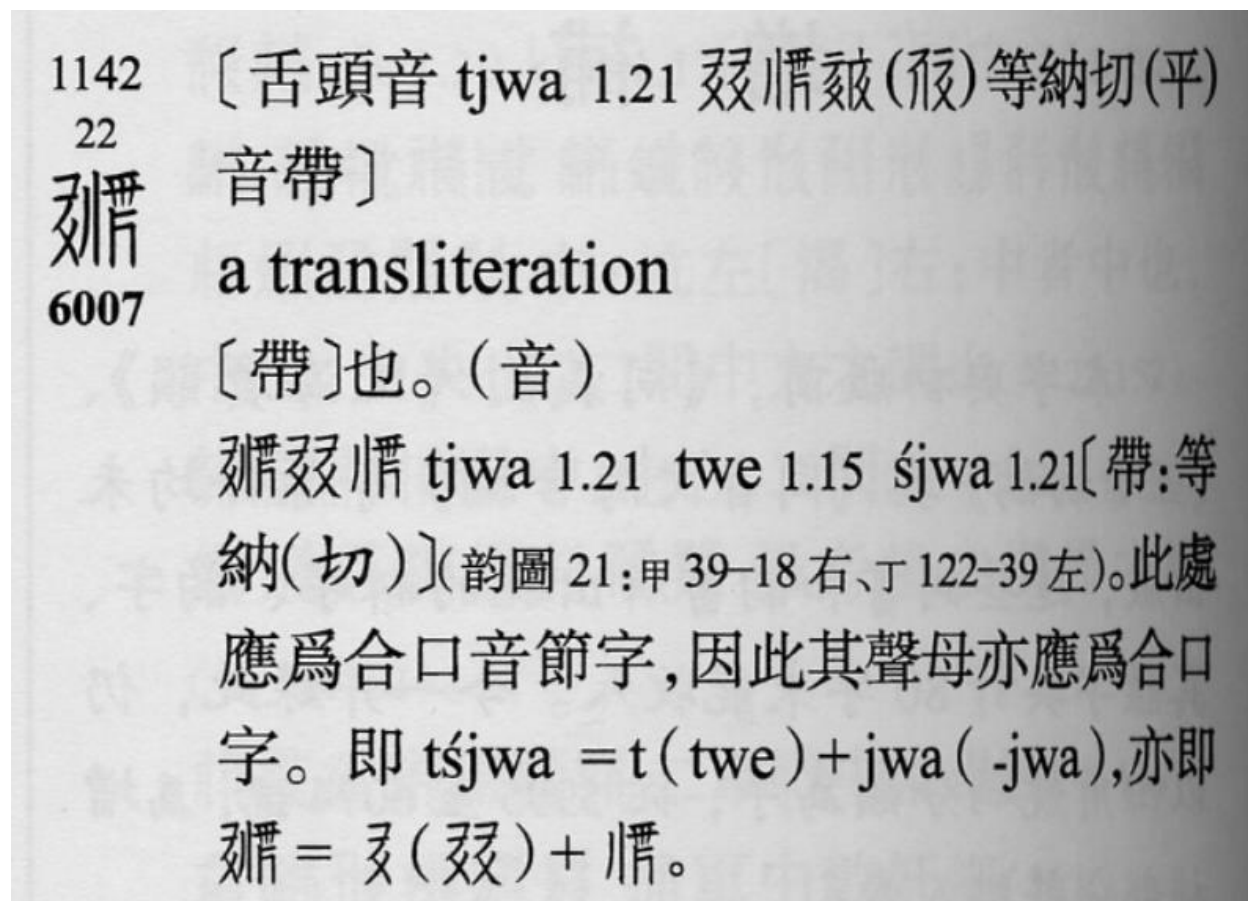
Fig. 15. IOM Tang. 22, Old Inv. No. 624 (𪛗𪛘𪛙𪛚 MS E) folio 37



U+18171 𪛗

U+18171 𪛗 *tjwa* is only attested in a single manuscript source, 𪛗𪛗𪛗𪛗 *Joined Rimes of the Fives Sounds*, and is entry #6007 in the 2008 edition of Lǐ Fànwén's Tangut-Chinese dictionary (see Fig. 16).

Fig. 16. Lǐ Fànwén, *Xià-Hàn zìdiǎn* (2008) p. 946 entry for #6007



This character is explained by Lǐ Fànwén as being constructed from the left side of U+18139 𪛗 *twe* and the whole of U+17154 𪛗 *sjwa*, with the *fǎnqiè* 反切 spelling of *tjwa* from these two characters.

U+18171 𪛗 only occurs in three manuscript versions of 𪛗𪛗𪛗𪛗 *Joined Rimes of the Fives Sounds* (see Figs. 17, 18, and 19), where the actual glyph form shows that the left middle vertical stroke is placed underneath the top part of the left side component.

Additionally, the radical for U+18171 was changed from 316 𪛗 to 768 𪛗 in Unicode 14.0 (see WG2 N5134; L2/20-166), but this was a mistake as can be seen from the actual glyph form in the original texts, as well as its presumed construction from the left side of U+18139 𪛗.

Fig. 17. IOM Tang. 22, Old Inv. No. 620 (𪛗𪛘𪛙𪛚 MS A) folio 18b

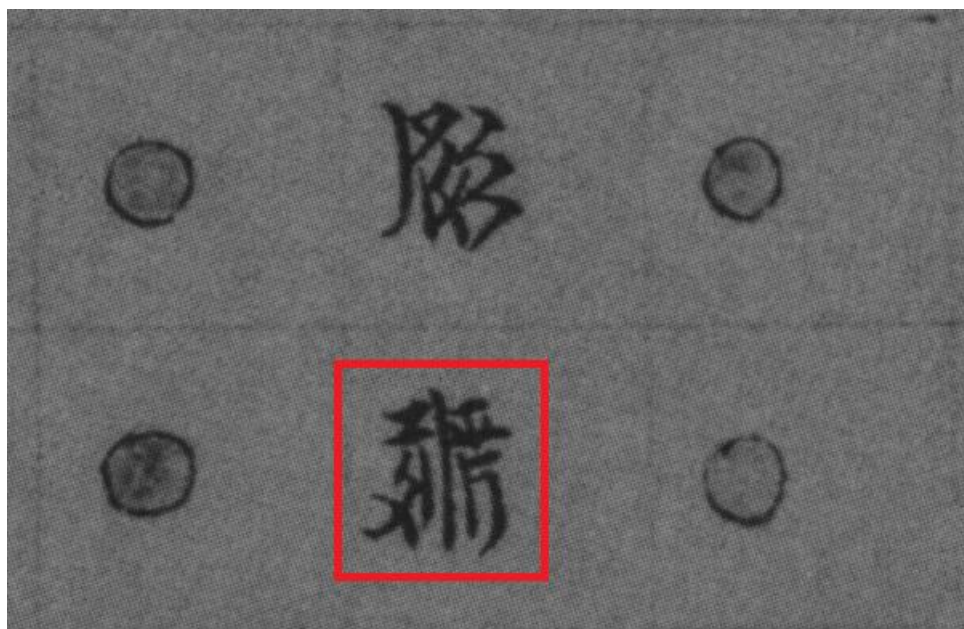


Fig. 18. IOM Tang. 22, Old Inv. No. 623 (𪛗𪛘𪛙𪛚 MS D) folio 24a

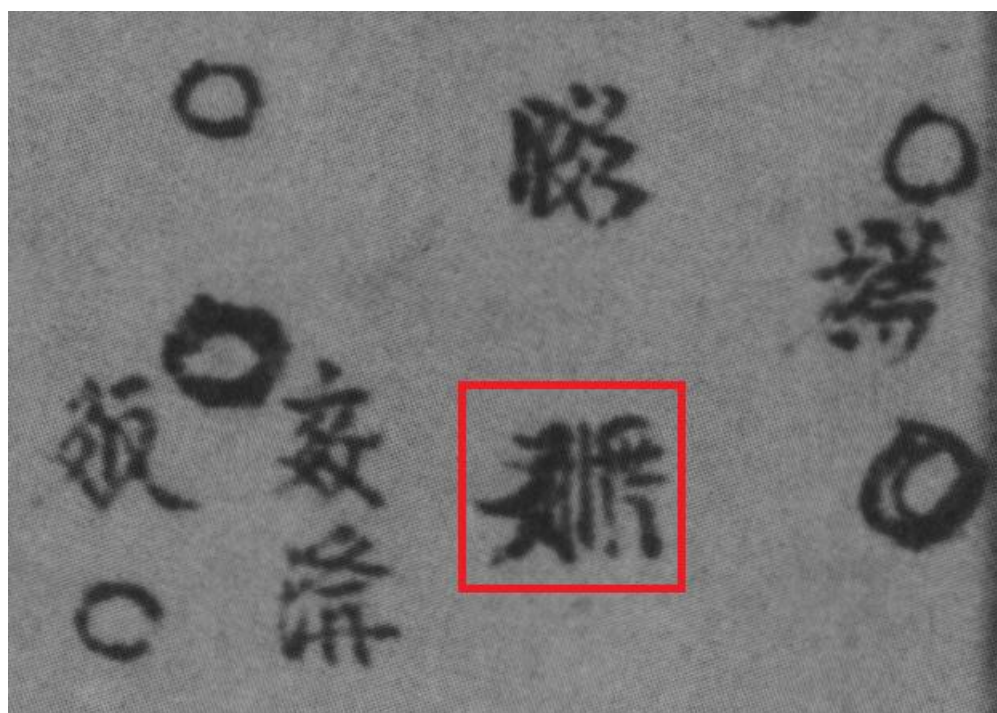


Fig. 19. IOM Tang. 22, Old Inv. No. 624 (𪔐𪔐𪔐𪔐 MS E) folio 4

