

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 encode 2 Tangut components and 28 Tangut ideographs

Source: Andrew West

Status: Individual Contribution



Action: For consideration by JTC1/SC2/WG2 and UTC

Date: 2023-10-02

1. Introduction

This is a proposal to encode two Tangut components shown in Table 1, and twenty-eight Tangut ideographs shown in Table 2. The two Tangut components are used in original Tangut texts from the Western Xia period. A proposal to encode 114 additional Tangut components used for indexing purposes in the recently-published Tangut dictionary compiled by Hán Xiǎománg has been submitted as [WG2 N5218 = L2/23-150](#).

Table 1: Proposed Tangut Components

Ref.	C/P	Glyph	Character Name	Sources
3.1	18D80		TANGUT COMPONENT-769	<i>Combined Edition of Homophones and Sea of Writing 'C' manuscript</i>
3.2	18D81		TANGUT COMPONENT-770	<i>𪚩𪚪𪚫𪚬𪚭𪚮𪚯𪚰 (Homonyms)</i>

Some of the proposed Tangut ideographs are entirely new characters, but some are (or may be) variant forms of existing encoded characters. I have currently documented over a thousand variants of Tangut ideographs, and the vast majority of these show minor differences such as an additional or missing stroke, substitution of one graphically similar shape for another, or modification of the relative positioning of components (similar to CJK UCV #305 and #307). These can be considered to be unifiable variants, and are candidates for representation using IVS in a future IVD registration for Tangut ideographs. This document proposes for separate encoding a small number of variant characters where the difference in glyph forms is too great for them to be considered to be unifiable variants (i.e. where the supposed variant form has one or more completely different components compared with the regular form). It may be useful to document the principles for Tangut glyph unification in a future Unicode Technical Note.

Table 2: Proposed Tangut Ideographs

Ref.	C/P	Glyph	IDS	R/S	Sources
4.1	187F8	𐰇𐰏	𐰇𐰏𐰑 𐰒 𐰓 𐰔	206.19	𐰇𐰏𐰑𐰒𐰓𐰔 = 重譯孝經傳 (<i>Classic of Filial Piety with Commentary</i>)
4.2	187F9	𐰇𐰏	𐰇𐰏	267.9	𐰇𐰏 = 論語 (<i>Analects of Confucius</i>) vols. 6 and 8
4.3	187FA	𐰇𐰏	𐰇𐰏𐰑 𐰒 𐰓 𐰔	75.19	𐰇𐰏𐰑𐰒𐰓𐰔 = 金光明最勝王經 (<i>Suvarṇaprabhāsa Sūtra</i>) vol. 4 𐰇𐰏𐰑𐰒𐰓𐰔 (Homonyms)
4.4	187FB	𐰇𐰏	𐰇𐰏	490.12	𐰇𐰏𐰑𐰒𐰓𐰔 = 金光明最勝王經 (<i>Suvarṇaprabhāsa Sūtra</i>) vol. 4
4.5	187FC	𐰇𐰏	𐰇𐰏	114.9	𐰇𐰏𐰑𐰒𐰓𐰔 = 金光明最勝王經 (<i>Suvarṇaprabhāsa Sūtra</i>) vol. 6
4.6	187FD	𐰇𐰏	𐰇𐰏𐰑 𐰒	217.15	𐰇𐰏𐰑𐰒𐰓𐰔 = 大方廣佛華嚴經 (<i>Mahāvaiṣṭya Buddhāvataṃsaka Sūtra</i>) vol. 41
4.7	187FE	𐰇𐰏	𐰇𐰏	519.13	𐰇𐰏𐰑𐰒𐰓𐰔 = 佛頂心觀世音菩薩陀羅尼經 (<i>Dhāraṇī Sūtra of the Heart of the Buddha Crown of Bodhisattva Avalokiteśvara</i>) vol. 2
4.8	187FF	𐰇𐰏	𐰇𐰏𐰑 𐰒	736.18	𐰇𐰏𐰑𐰒𐰓𐰔 = 過去莊嚴劫千佛名經 (<i>Sūtra of the Names of the Thousand Buddhas of the Past Kalpa</i>)
4.9	18D09	𐰇𐰏	𐰇𐰏	590.14	𐰇𐰏𐰑𐰒𐰓𐰔 = 過去莊嚴劫千佛名經 (<i>Sūtra of the Names of the Thousand Buddhas of the Past Kalpa</i>)
4.10	18D0A	𐰇𐰏	𐰇𐰏𐰑 𐰒 𐰓 𐰔 𐰕	267.18	𐰇𐰏𐰑𐰒𐰓𐰔 = 妙法蓮華經 (<i>Saddharma Puṇḍarīka</i>)
4.11	18D0B	𐰇𐰏	𐰇𐰏	267.10	𐰇𐰏𐰑𐰒𐰓𐰔 (Homonyms)
4.12	18D0C	𐰇𐰏	𐰇𐰏	655.13	𐰇𐰏𐰑𐰒𐰓𐰔 (Homonyms)
4.13	18D0D	𐰇𐰏	𐰇𐰏	456.14	𐰇𐰏𐰑𐰒𐰓𐰔 (Homonyms)
4.14	18D0E	𐰇𐰏	𐰇𐰏	273.9	𐰇𐰏𐰑𐰒𐰓𐰔 (Homonyms)

Ref.	C/P	Glyph	IDS	R/S	Sources
4.15	18D0F	𦉑	𦉑 𦉑 𦉑	278.11	𦉑𦉑𦉑𦉑𦉑𦉑 = 番漢合時掌中珠 (<i>Pearl in the Palm</i>)
4.16	18D10	𦉒	𦉒 𦉒 𦉒	106.13	𦉑𦉑𦉑𦉑𦉑𦉑 = 番漢合時掌中珠 (<i>Pearl in the Palm</i>)
4.17	18D11	𦉓	𦉓 𦉓 𦉓	75.8	𦉑𦉑𦉑𦉑𦉑𦉑 = 番漢合時掌中珠 (<i>Pearl in the Palm</i>)
4.18	18D12	𦉔	𦉔 𦉔	17.7	𦉑𦉑 (Homophones)
4.19	18D13	𦉕	𦉕 𦉕 𦉕	106.15	𦉑𦉑 = 雜字 (<i>Assorted Characters</i>)
4.20	18D14	𦉖	𦉖 𦉖 𦉖	167.15	𦉑𦉑 = 雜字 (<i>Assorted Characters</i>)
4.21	18D15	𦉗	𦉗 𦉗	462.11	𦉑𦉑𦉑𦉑 = 五音切韻 (<i>Joined Rimes of the Five Sounds</i>)
4.22	18D16	𦉘	𦉘 𦉘 𦉘	579.17	𦉑𦉑𦉑𦉑 = 五音切韻 (<i>Joined Rimes of the Five Sounds</i>)
4.23	18D17	𦉙	𦉙 𦉙 𦉙	210.13	𦉑𦉑𦉑𦉑 = 五音切韻 (<i>Joined Rimes of the Five Sounds</i>)
4.24	18D18	𦉚	𦉚 𦉚 𦉚	278.13	𦉑𦉑𦉑𦉑 = 五音切韻 (<i>Joined Rimes of the Five Sounds</i>)
4.25	18D19	𦉛	𦉛 𦉛 𦉛	141.19	𦉑𦉑𦉑𦉑 = 五音切韻 (<i>Joined Rimes of the Five Sounds</i>)
4.26	18D1A	𦉜	𦉜 𦉜 𦉜	75.13	𦉑𦉑𦉑𦉑 (Newly Assembled <i>Precious Dual Maxims</i>)
4.27	18D1B	𦉝	𦉝 𦉝	36.7	𦉑𦉑𦉑𦉑𦉑𦉑 (Collected Writings of <i>the Shining Speech of Three Generations</i>)
4.28	18D1C	𦉞	𦉞 𦉞 𦉞	141.9	<i>Stele of Sulaiman</i>

2. Code Point Allocation and Character Properties

2.1 Tangut Ideographs

This document follows the recommendation of the Script Ad Hoc group to UTC #176 (L2/23-164) in assigning the twenty-eight proposed Tangut ideographs to the ranges 187F8..187FF at the end of the Tangut block (eight characters) and 18D09..18D1C in the Tangut Supplement block (twenty characters).

2.2 Tangut Components

This document follows the recommendation of the Script Ad Hoc group to UTC #176 (L2/23-164) in assigning the two proposed Tangut ideographs to the start of a new Tangut Components Supplement block in the range 18D80..18DFF.

2.3 Unicode Character Properties

The entries for Tangut ideographs in the UnicodeData.txt file should be amended to:

```
17000;<Tangut Ideograph, First>;Lo;0;L;;;;N;;;;;
187FF;<Tangut Ideograph, Last>;Lo;0;L;;;;N;;;;;
18D00;<Tangut Ideograph Supplement, First>;Lo;0;L;;;;N;;;;;
18D1C;<Tangut Ideograph Supplement, Last>;Lo;0;L;;;;N;;;;;
```

The following two entries for Tangut components should be added to the UnicodeData.txt file:

```
18D80;TANGUT COMPONENT-769;Lo;0;L;;;;N;;;;;
18D81;TANGUT COMPONENT-770;Lo;0;L;;;;N;;;;;
```

2.4 Tangut Sources

The TangutSources.txt file should be amended to add a new source reference for the 28 proposed additional Tangut ideographs. The source reference can be the WG2 number of the proposal document, i.e. “N5217”, followed by a two-digit sequential number (i.e. N5217-01 through N5217-28):

```
# N5217 = Andrew West, Proposal to encode 2 Tangut components
and 28 Tangut ideographs,
#           WG2 N5217 = L2/23-149. 2023-10-02.
```

The following entries should be added to the TangutSources.txt file:

U+187F8	kTGT_MergedSrc	N5217-01
U+187F8	kRSTUnicode	206.19
U+187F9	kTGT_MergedSrc	N5217-02
U+187F9	kRSTUnicode	267.9
U+187FA	kTGT_MergedSrc	N5217-03
U+187FA	kRSTUnicode	75.19
U+187FB	kTGT_MergedSrc	N5217-04
U+187FB	kRSTUnicode	490.12
U+187FC	kTGT_MergedSrc	N5217-05
U+187FC	kRSTUnicode	114.9
U+187FD	kTGT_MergedSrc	N5217-06
U+187FD	kRSTUnicode	217.15
U+187FE	kTGT_MergedSrc	N5217-07
U+187FE	kRSTUnicode	519.13
U+187FF	kTGT_MergedSrc	N5217-08
U+187FF	kRSTUnicode	736.18
U+18D09	kTGT_MergedSrc	N5217-09
U+18D09	kRSTUnicode	590.14
U+18D0A	kTGT_MergedSrc	N5217-10
U+18D0A	kRSTUnicode	267.18
U+18D0B	kTGT_MergedSrc	N5217-11
U+18D0B	kRSTUnicode	267.10
U+18D0C	kTGT_MergedSrc	N5217-12
U+18D0C	kRSTUnicode	655.13
U+18D0D	kTGT_MergedSrc	N5217-13
U+18D0D	kRSTUnicode	456.14
U+18D0E	kTGT_MergedSrc	N5217-14
U+18D0E	kRSTUnicode	273.9
U+18D0F	kTGT_MergedSrc	N5217-15
U+18D0F	kRSTUnicode	278.11
U+18D10	kTGT_MergedSrc	N5217-16
U+18D10	kRSTUnicode	106.13
U+18D11	kTGT_MergedSrc	N5217-17
U+18D11	kRSTUnicode	75.8
U+18D12	kTGT_MergedSrc	N5217-18
U+18D12	kRSTUnicode	17.7
U+18D13	kTGT_MergedSrc	N5217-19
U+18D13	kRSTUnicode	106.15
U+18D14	kTGT_MergedSrc	N5217-20
U+18D14	kRSTUnicode	167.15
U+18D15	kTGT_MergedSrc	N5217-21
U+18D15	kRSTUnicode	462.11
U+18D16	kTGT_MergedSrc	N5217-22
U+18D16	kRSTUnicode	579.17
U+18D17	kTGT_MergedSrc	N5217-23
U+18D17	kRSTUnicode	210.13

U+18D18	kTGT_MergedSrc	N5217-24
U+18D18	kRSTUnicode	278.13
U+18D19	kTGT_MergedSrc	N5217-25
U+18D19	kRSTUnicode	141.19
U+18D1A	kTGT_MergedSrc	N5217-26
U+18D1A	kRSTUnicode	75.13
U+18D1B	kTGT_MergedSrc	N5217-27
U+18D1B	kRSTUnicode	36.7
U+18D1C	kTGT_MergedSrc	N5217-28
U+18D1C	kRSTUnicode	141.9

3. Evidence and Notes for Tangut Components

3.1 Component 1 𠵹

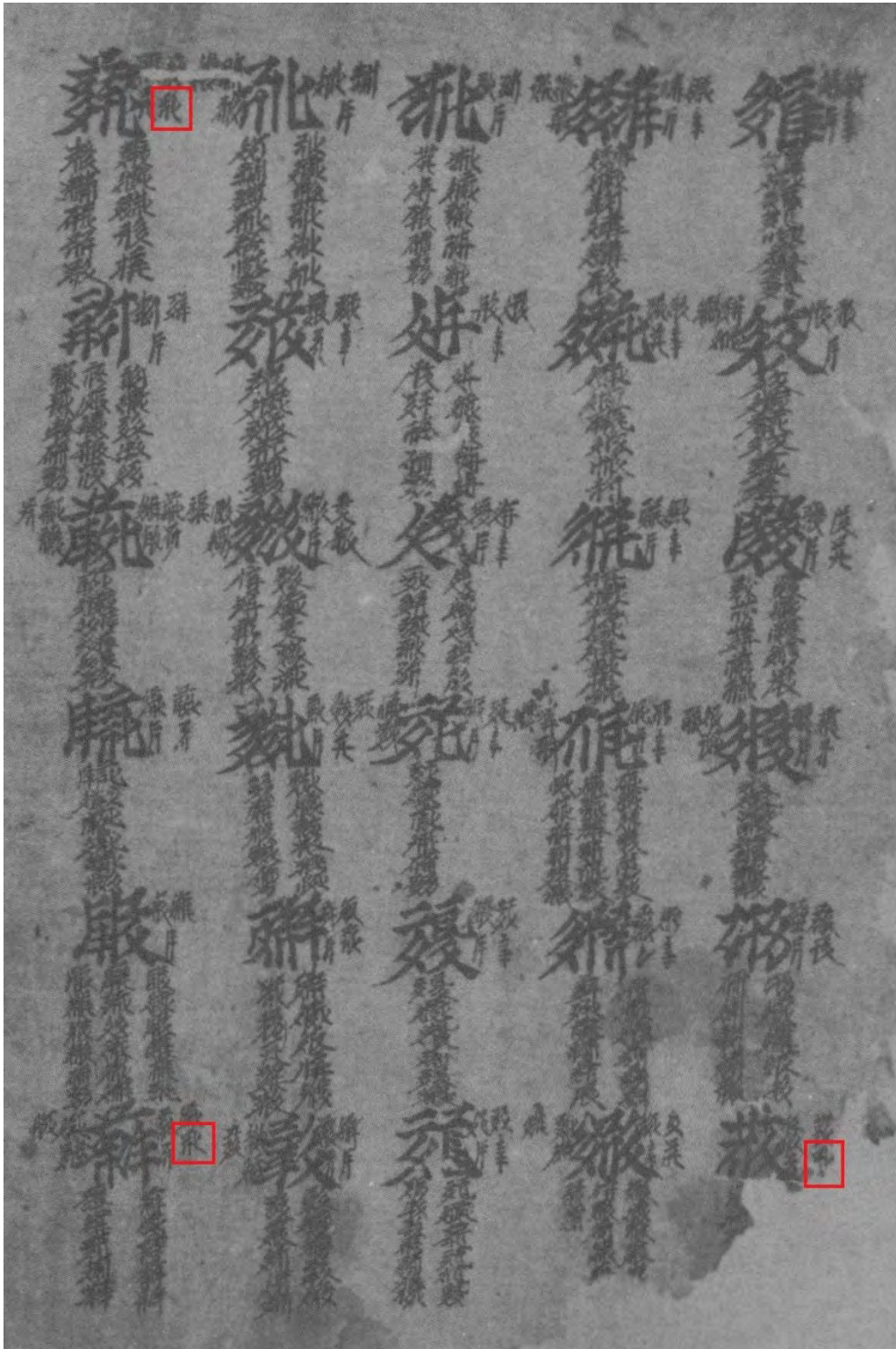


Fig. 1: IOM Inv. No. 4152 folio A ¹

¹ 'Tóngyīn Wénhǎi Bǎoyùn Hébiān' Zhènglǐ yǔ Yánjiū plate 62.

The 'C' manuscript of the *Combined Edition of Homophones and Sea of Writing* uses Tangut components as abbreviations for the seven Tangut ideographs used elsewhere to indicate structural positioning when describing the construction of Tangut characters (a four character formula that acts as a form of ideographic description).

Table 3: Abbreviations for positional characters

Character	Reading	Meaning	Abbreviation	Component
𗀃	<i>pha</i> ¹	left side	𗀃	394
𗀄	<i>bjir</i> ²	right side	𗀄	118
𗀅	<i>yu</i> ¹	head (top)	𗀅	
𗀆	<i>tshjij</i> ¹	bottom	𗀆	89
𗀇	<i>io</i> ¹	surrounding	𗀇	428
𗀈	<i>njij</i> ¹	middle	𗀈	9
𗀉	<i>nowr</i> ²	whole	𗀉	408

Six of the seven abbreviations are representable using existing encoded Tangut components, but the abbreviation for 𗀅 'head' (𗀅) does not exist either as a Tangut ideograph or as a Tangut component. This abbreviation is used five times in the only two surviving pages of the 'C' manuscript, which are listed in Table 4.

Table 4: Character constructions with abbreviation for 'head'

Entry	Character	Construction	Description
A16	𗀃	𗀃 𗀅 𗀆 𗀇	Top of 𗀃 and bottom of 𗀆
A51	𗀅	𗀅 𗀆 𗀇 𗀉	Top of 𗀅 and whole of 𗀉
A56	𗀅	𗀅 𗀆 𗀇 𗀉	Top [right] of 𗀅 and bottom of 𗀉
B14	𗀅	𗀅 𗀆 𗀇 𗀉	Top of 𗀅 and whole of 𗀉
B55	𗀅	𗀅 𗀆 𗀇 𗀉	Top of 𗀅 and whole of 𗀉

3.2 Component 2 [≡]

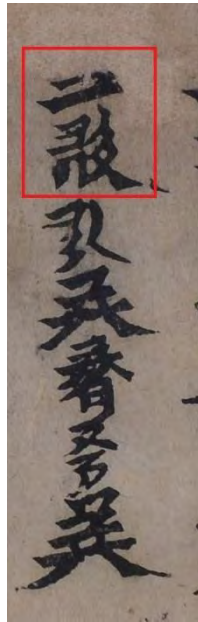


Fig. 2: 《𐰇𐰏𐰐𐰑𐰒𐰓𐰔𐰕》 folio 4A

In WG2 N4957 = L2/18-194 I proposed seven Tangut components used in *Homonyms* for encoding, and these were added to Unicode 13.0 as 18AF3..18AF9. However, I misidentified one component on folio 4A of *Homonyms* because it was damaged and missing the top part (Fig. 2). I misidentified the sample character below the component as U+18375 𐰇, and assumed the damaged component above it was Component 172 [≡] (see N4957 Table 1 Row 27). However, a more careful inspection of the sample character shows that it is actually U+180F8 𐰇, and the damaged component is actually [≡].

4. Evidence and Notes for Tangut Ideographs

4.1 Classic of Filial Piety with Commentary

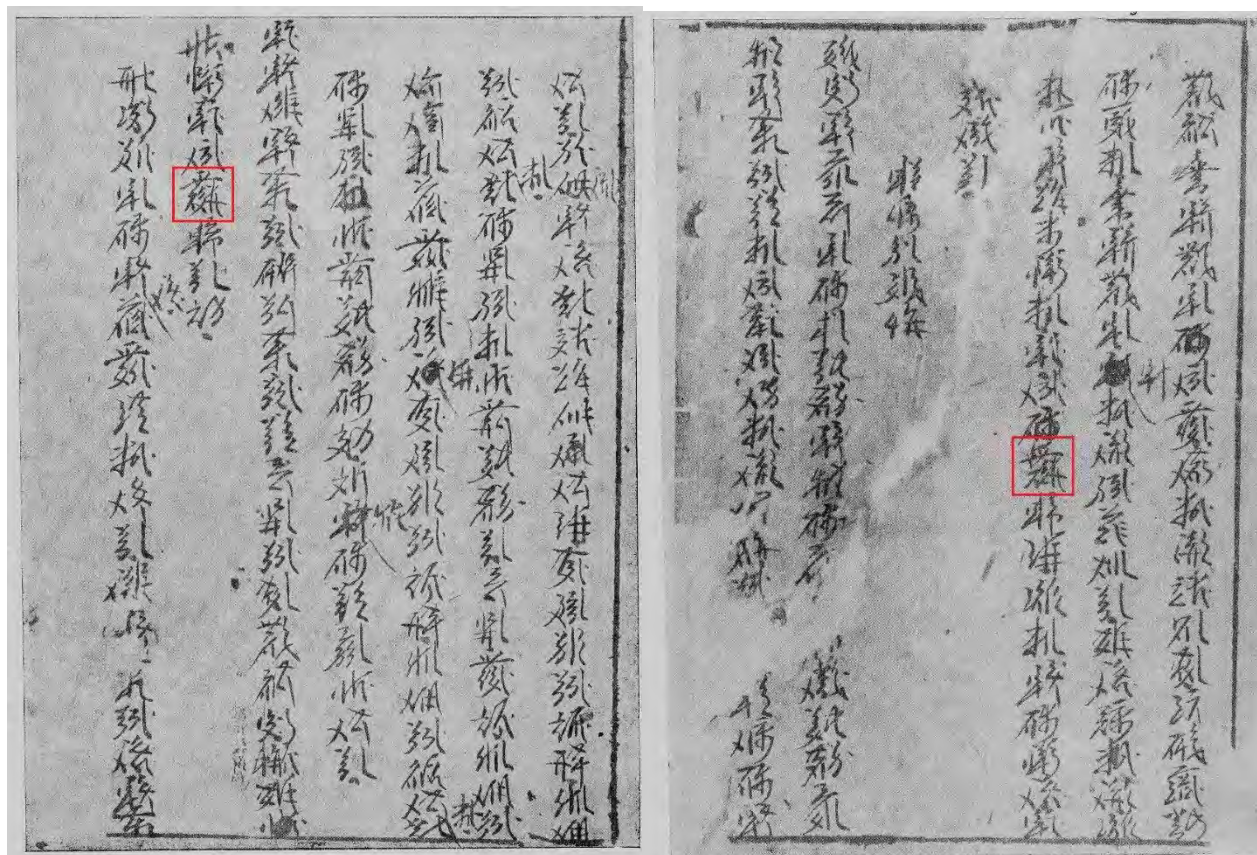


Fig. 3: IOM Tang. 1 Inv. No. 2627: 《總攬粹駁》 folios 6A and 7A ²

This character occurs three times in the Tangut translation of the *Classic of Filial Piety* (two examples shown above). Unfortunately, this text only survives as a single manuscript written in a cursive hand, so it is a little difficult to determine the correct regular script form of the character. The character is included in Jiǎ Chángyè's *Xìxiàwén Zìdiǎn* (2019) and in Hán Xiǎománg's *Xìxiàwén Cídiǎn* (2021), where it is regularized as 𐞑 by both authors. Jiǎ Chángyè gives the character construction as the top and bottom left-middle of 𐞑 *tśja*¹ 'road' and the right side of 𐞑 *twɛ*¹ 'to sew' (although there is no external evidence to support this construction). However, in my opinion, the manuscript shows that the bottom middle component is 𐞑 (Component 141) rather than 𐞑 (Component 767) or 𐞑 (Component 278), which would suggest a regularised form of 𐞑. Nevertheless, without further textual evidence for the correct glyph form, it is best to encode the character form as given by Jiǎ Chángyè (Fig. 4) and Hán Xiǎománg (Fig. 5).

² Kitajskaja klassika v tangutskom perevode pp. 145 and 147; also one further example on p. 137.

4.2 Analects of Confucius

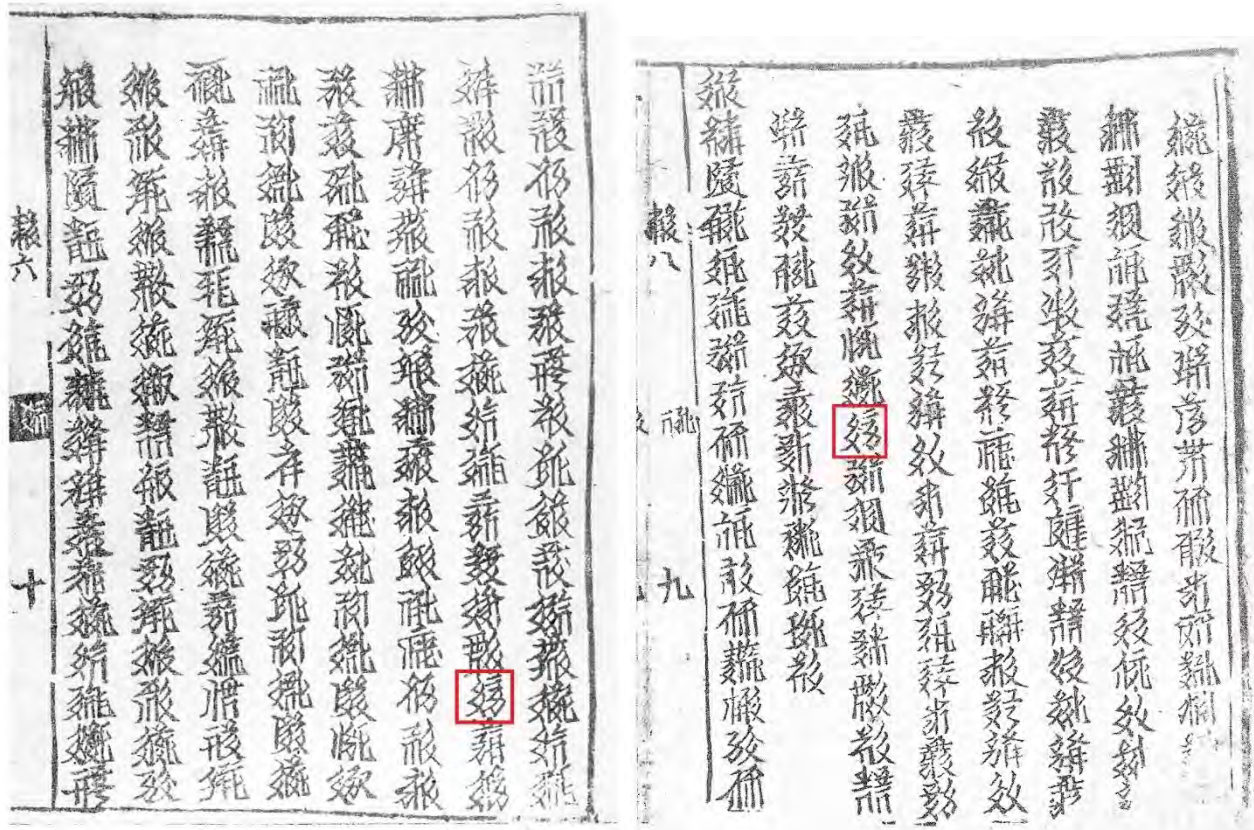


Fig. 6: IOM Tang. 1 Inv. No. 2627: 《𐰇𐰏》 6:10A and 8:9A ³

This character occurs three times in the woodblock printed edition of the Tangut translation of the *Analects of Confucius* (two examples shown above). It is included in Hán Xiǎománg's *Xìxiàwén Cídiǎn* (2021) where it is regularized as 𐰇 (Fig. 7). However, the original woodblock printed edition shows that the left side is Component 766 (𐰇) not Component 267 (𐰇), so the correct form should be 𐰇.

³ Kitajskaja klassika v tangutskom perevode pp. 23 and 36; also one further example on p. 41.

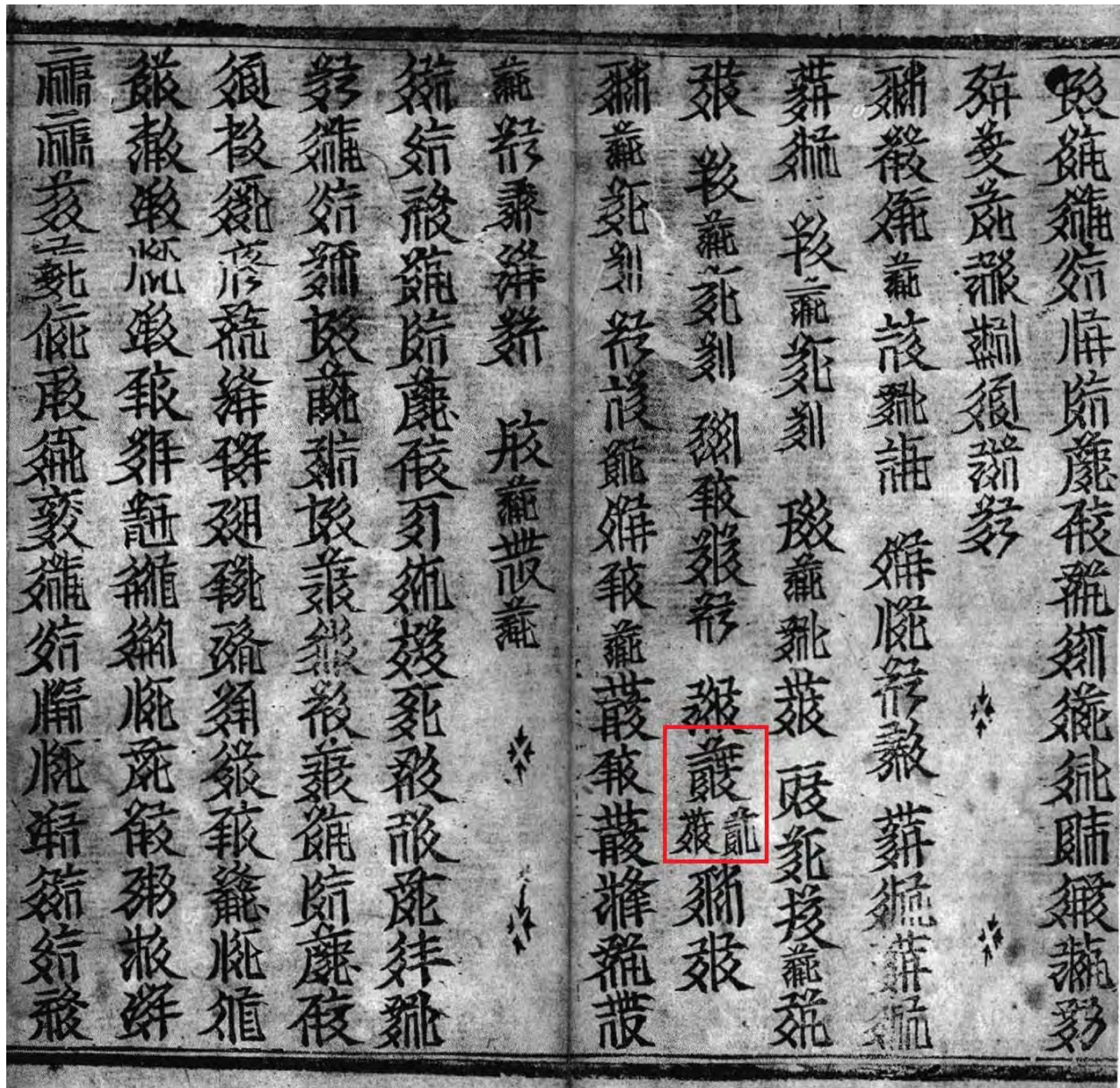


Fig. 8: National Library of China B11-028: 《釋迦傳編禪席𑖀𑖀》𑖀𑖀𑖀⁴

The character 𑖀 is a transcription character for Sanskrit used in a *dhāraṇī* incantation, and is glossed graphically and phonetically with the two characters 𑖀 *dji*¹ and 𑖀 *jo*² which give the *fanqie* of reading *djo*². In this *dhāraṇī* the Tangut word 𑖀𑖀𑖀 *wu djo tja tji* corresponds to the Sanskrit word *cukuti* (!) (*dau*⁴*dāntē* in the Pali version, and 調⁴但底 in the Chinese version).

This character also occurs in *Homonyms* under the homophone group 𑖀 *dwu*² ‘similar’, together with the characters 𑖀 *dwu*² and 𑖀 *twu*¹ (Fig. 9), and so should have the reading *dwu* or *twu*, which is similar to the reading *djo*² given in the *dhāraṇī* incantation.

⁴ *Zhōngguó cáng Xìxià Wénxiàn* vol. 3 p. 155.

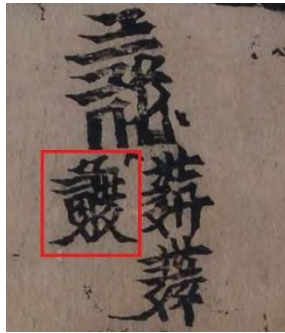


Fig. 9: 《𦵑𦵑𦵑𦵑𦵑𦵑𦵑𦵑》 folio 12B

The character is written slightly differently here, as 蕨 rather than 蕨, with one stroke missing (dot on middle right). However, the character 芟 *tšji*¹ 'branch' is written as 芟 in *Homonyms* folio 8A, with the same missing stroke, which suggests that 蕨 is a variant form of 蕨. The difference between the two character forms is minor, and 蕨 can be considered to be a unifiable variant; therefore, only 蕨 is proposed for encoding.

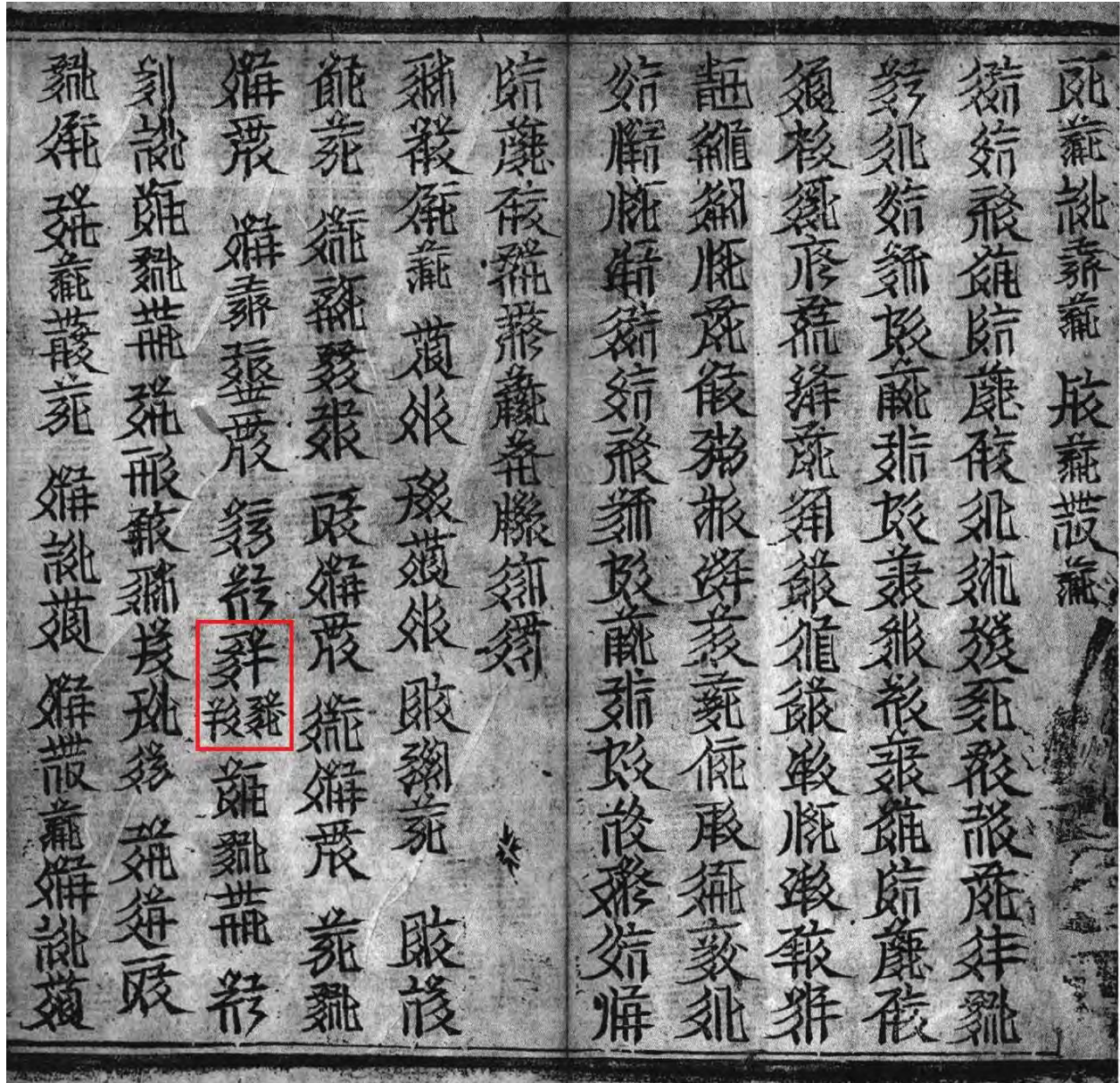


Fig. 10: National Library of China B11-028: 《釋迦傳經新集》 翻經錄⁵

The character 𑖦 is a transcription character for Sanskrit used in a *dhāraṇī* incantation, and is glossed graphically and phonetically with the two characters 𑖦 *njj*¹ and 𑖦 *ja*² which give the *fanqie* reading of *nja*². In this *dhāraṇī* the Tangut word 𑖦𑖦𑖦𑖦𑖦 *xji rjar nja gja r(jir) bjij* corresponds to the Sanskrit word *hiraṇyagarbhe* (𑖦𑖦𑖦𑖦𑖦 𑖦𑖦 in the Chinese version).

⁵ Zhōngguó cáng Xìxià Wénxiàn vol. 3 p. 162.

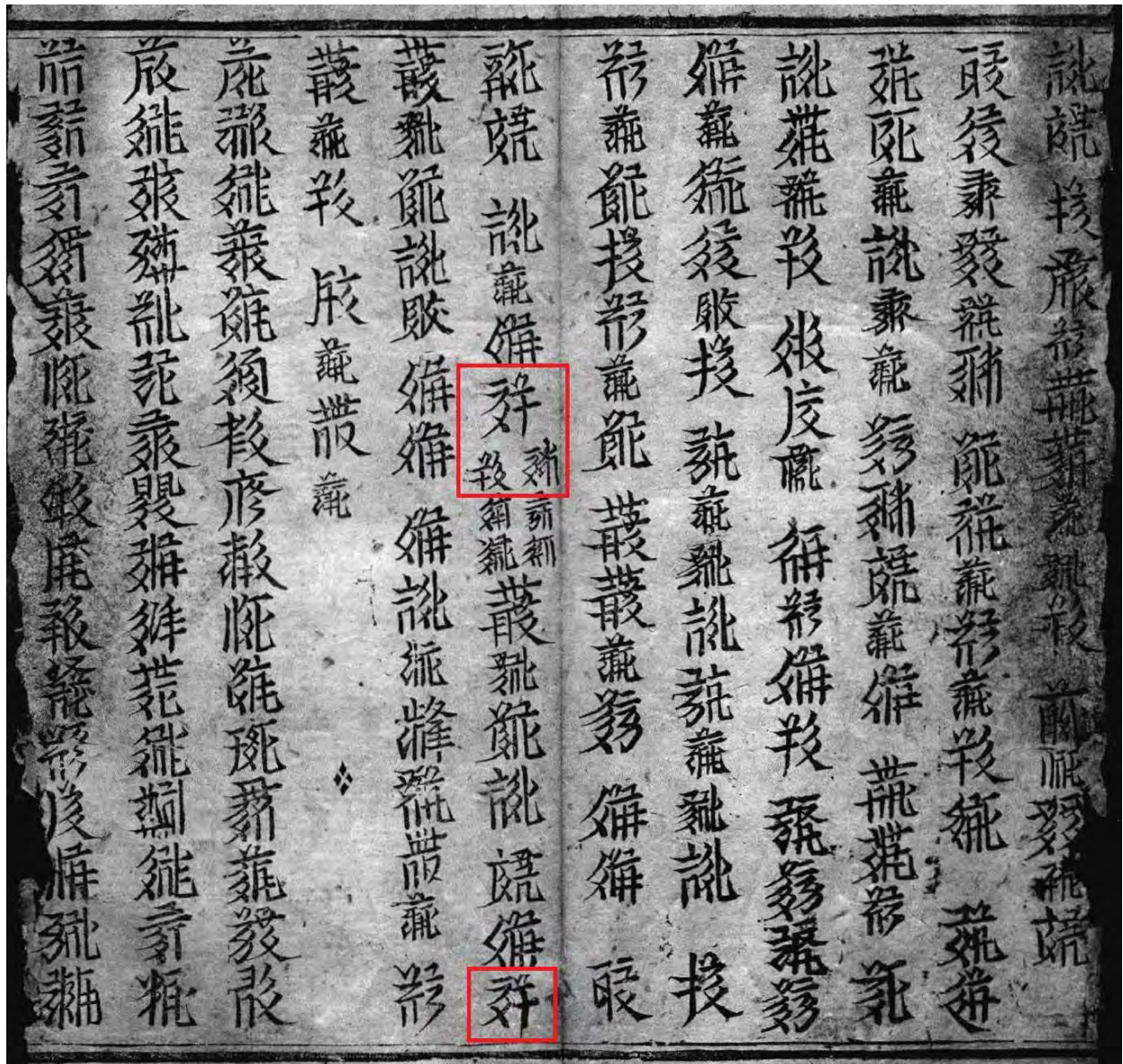


Fig. 11: National Library of China B11-028: 《繆殿傷癒祈席瓊茲》𑖑𑖑𑖑⁶

The character 𑖑 is a transcription character for Sanskrit used in a *dhāraṇī* incantation, and is glossed graphically and phonetically with the two characters 𑖑 *sjj*¹ and 𑖑 *ja*² which give the *fanqie* reading of *sjja*². In this *dhāraṇī* the Tangut words 𑖑𑖑𑖑 𑖑𑖑𑖑 𑖑𑖑𑖑 𑖑𑖑𑖑 *ja mu kja nja: mja sja* and 𑖑𑖑𑖑 𑖑𑖑𑖑 𑖑𑖑𑖑 𑖑𑖑𑖑 *dja r(j)r śja nja kja mja sja* correspond to the Sanskrit words *amakanamaya* (!) and *darśanakāmasya* (阿目迦那末寫 and 達哩設那迦末寫 in the Chinese version).

⁶ Zhōngguó cáng Xìxià Wénxiàn vol. 3 p. 324; there is also another copy of the same edition in vol. 3 p. 285.



Fig. 14: National Library of China B11-052: 《維摩經疏》 colophon ⁹

⁹ Zhōngguó cáng Xìxià Wénxiàn vol. 6 p. 56.

The 1312 colophon for the *Sūtra of the Names of the Thousand Buddhas of the Past Kalpa* includes the character 𪛗 in the phrase 𪛗𪛗𪛗𪛗 which Shǐ Jīnbō translates as 漢孝明帝 'Emperor Xiaoming of Han'. Therefore it should be a variant form or mistake for U+187D1 𪛗 *xā*² with the common right hand component 𪛗 (occurs in 18 characters) instead of the expected right side component (only occurs in 𪛗). This character is also given in Nevsky's draft manuscript Tangut dictionary (Fig. 16).

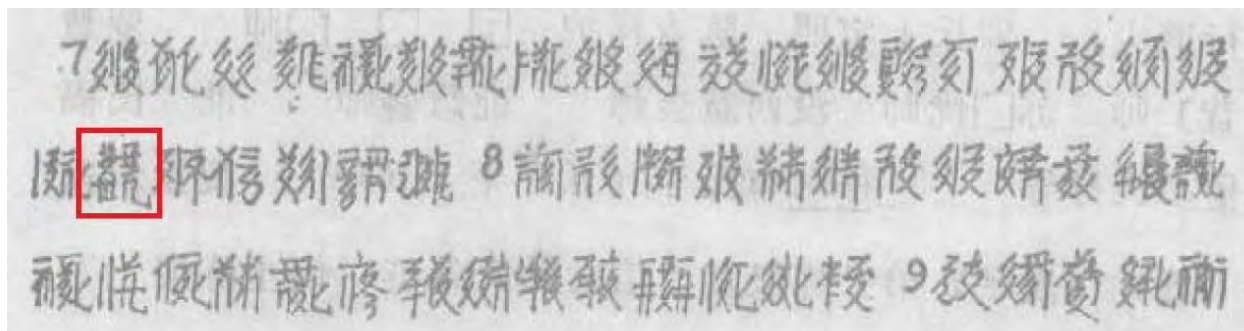


Fig. 15: *Xixia Fójiào Shilüe* p. 316

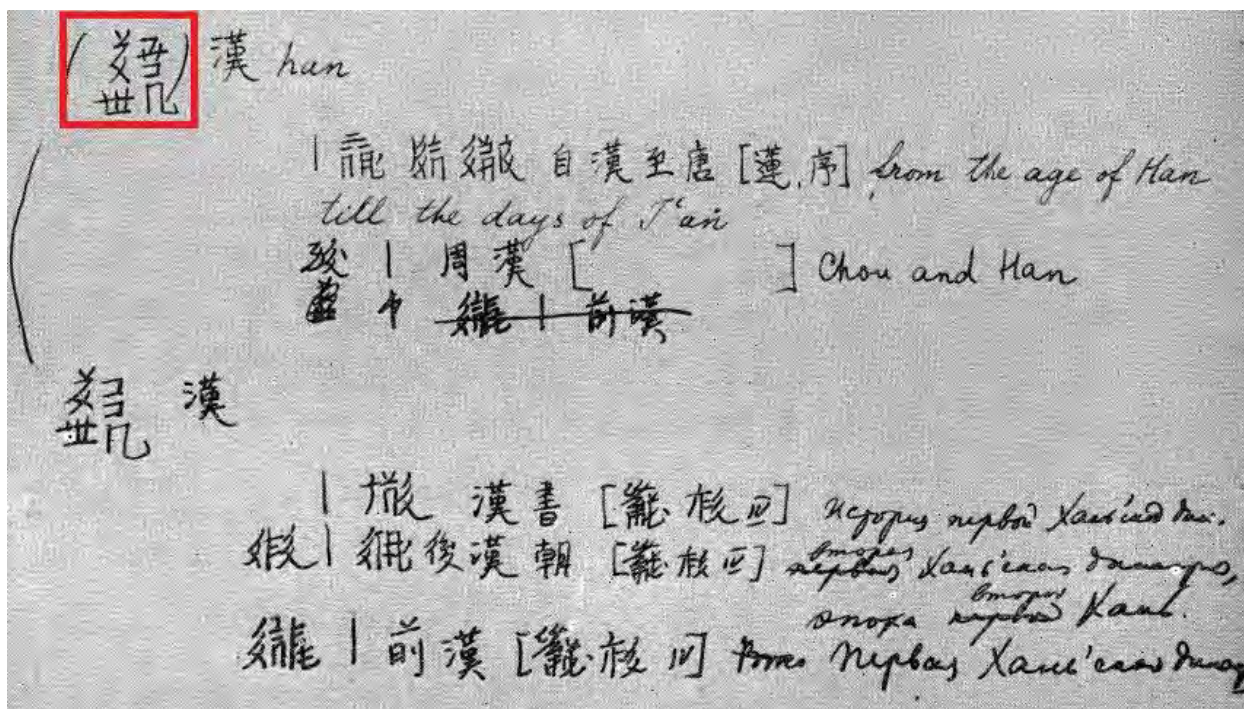


Fig. 16: *Tangutskaya filologiya* vol. II p. 655

4.9 Sūtra of the Names of the Thousand Buddhas of the Past Kalpa



Fig. 17: National Library of China B11-052: 《維摩詰經》 colophon ¹⁰

¹⁰ Zhōngguó cáng Xìxià Wénxiàn vol. 6 p. 57.

The 1312 colophon for the *Sūtra of the Names of the Thousand Buddhas of the Past Kalpa* includes the character 𪛗 in the phrase 𪛗𪛗𪛗𪛗 which Shǐ Jīnbō translates as 令依蕃譯 ‘caused to be translated according to Tangut [language]’. In this case, 𪛗 should be a variant form or mistake for U+1869C 𪛗 *lhjwij*¹ ‘Tangut’.

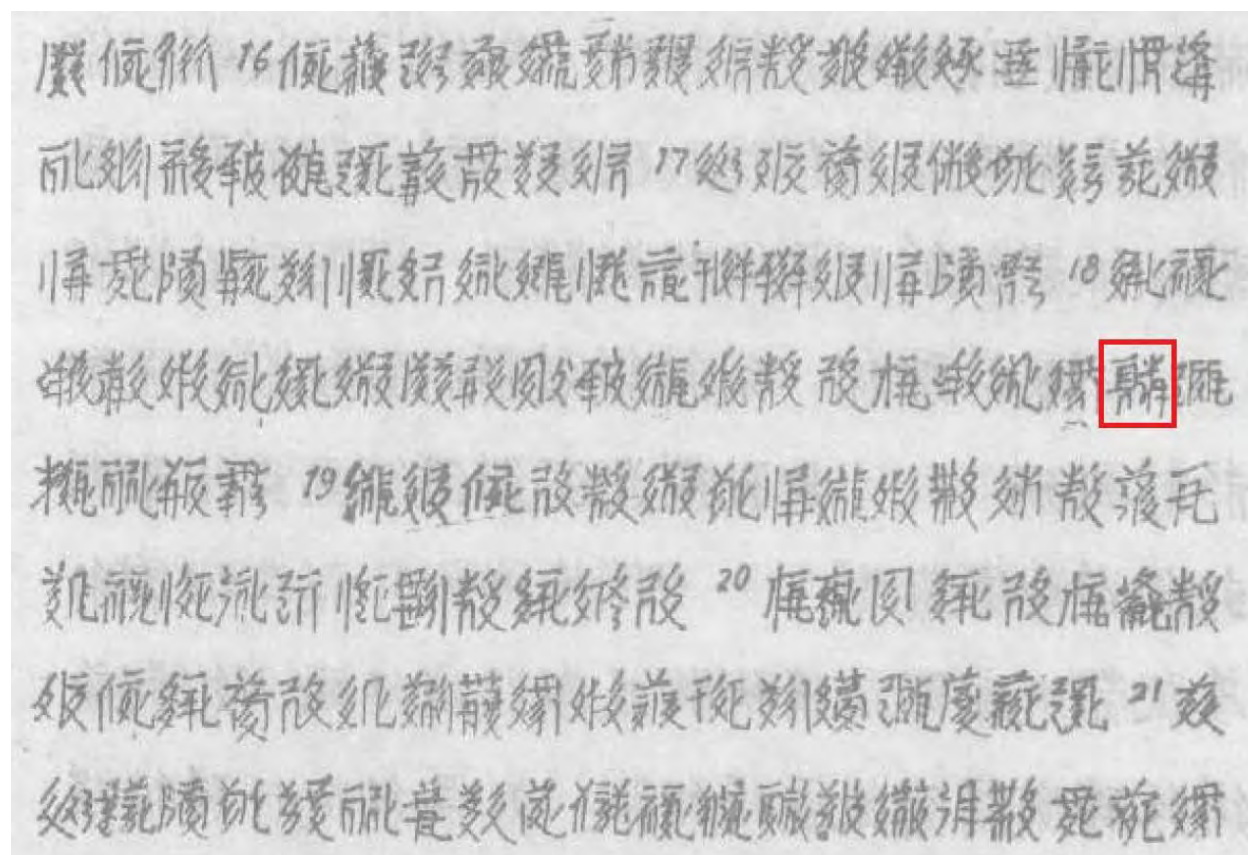


Fig. 18: *Xixia Fójiào Shǐlüè* p. 317

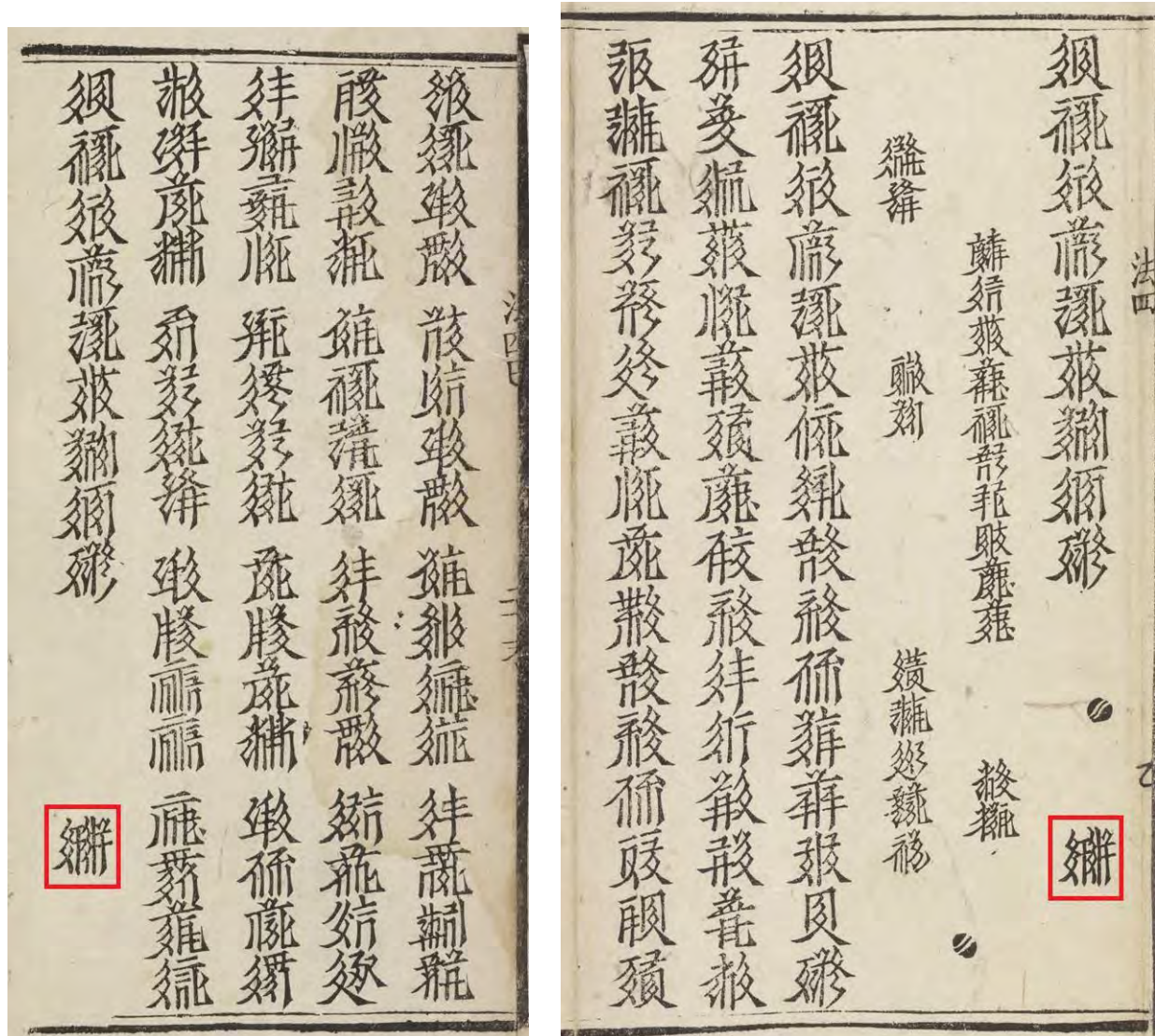


Fig. 19: Princeton BQ2053.T3 T38: 《網禪經》¹¹

The character 辯 occurs as the sutra index character at the start and end of a woodblock edition of the *Saddharma Puṇḍarīka* (*Lotus Sutra*) held at Princeton University Library. It is probably a variant of U+17FF8 辯 *lhjwij*¹ ‘title’ with the components rearranged (Arakawa Shintarō transcribes it as 辯), but as there is no context for its pronunciation or meaning it is not impossible that it is a separate character. Certainly it cannot be a one-off error as it occurs in the same form at the start and end of the sutra, and multiple times in the pagination throughout the volume.

¹¹ Tangut Version of the Lotus Sutra in the Collection of Princeton University Library, *Facsimile, Text and Linguistic Studies* pp. V and LIII.

4.11 Homonyms



Fig. 20: 《玃玃玃玃玃玃玃玃》 folio 13A

In *Homonyms* the character U+17FE0 玃 *tsjwiw*² ‘amber’ is written as 玃, without the top right component (玃). This is probably a mistake influenced by the following character 玃 *nji*² ‘pearl’. However, this is not a unifiable variant.

4.12 Homonyms

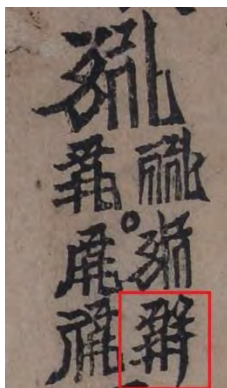


Fig. 21: 《𪗇𪗈𪗉𪗊𪗋𪗌𪗍𪗎𪗏𪗐》 folio 8A

The character 解 occurs in *Homonyms* under the homophone group 𪗌 *phji*¹ ‘idea’, together with the characters 𪗌 *phji*¹, 𪗍 *phji*², 𪗎 *phji*¹, 𪗏 *phji*¹, and 𪗐 *phji*¹, and so should have the reading *phji*¹ or *phji*². The character is perhaps a mistake for U+170C1 𪗑 *phja*², with the first stroke omitted, but the reading does not match for this homophone group.

4.13 Homonyms

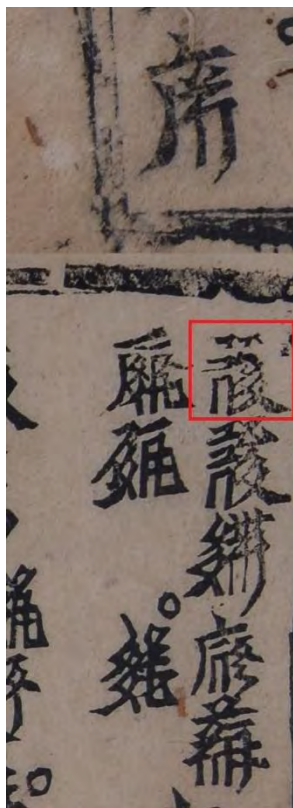


Fig. 22: 《麗麗麗麗麗麗麗麗》 folio 12B-13A

The character 麗 occurs in *Homonyms* under the homophone group 席 lu^2 'seat', together with the characters 麗 lu^1 , 麗 lu^2 , 麗 lu^2 , 麗 lu^2 , 麗 lu^2 , 麗 lu^1 , and 麗 lu^2 , and so should have the reading lu^1 or lu^2 . The character looks almost the same as the following character 麗, but the right side is written as 麗 with has an additional stroke compared with 麗. The component 麗 also occurs in the characters 麗 bju^1 , 麗 ka^1 , 麗 bju^1 , and 麗 ka^1 .

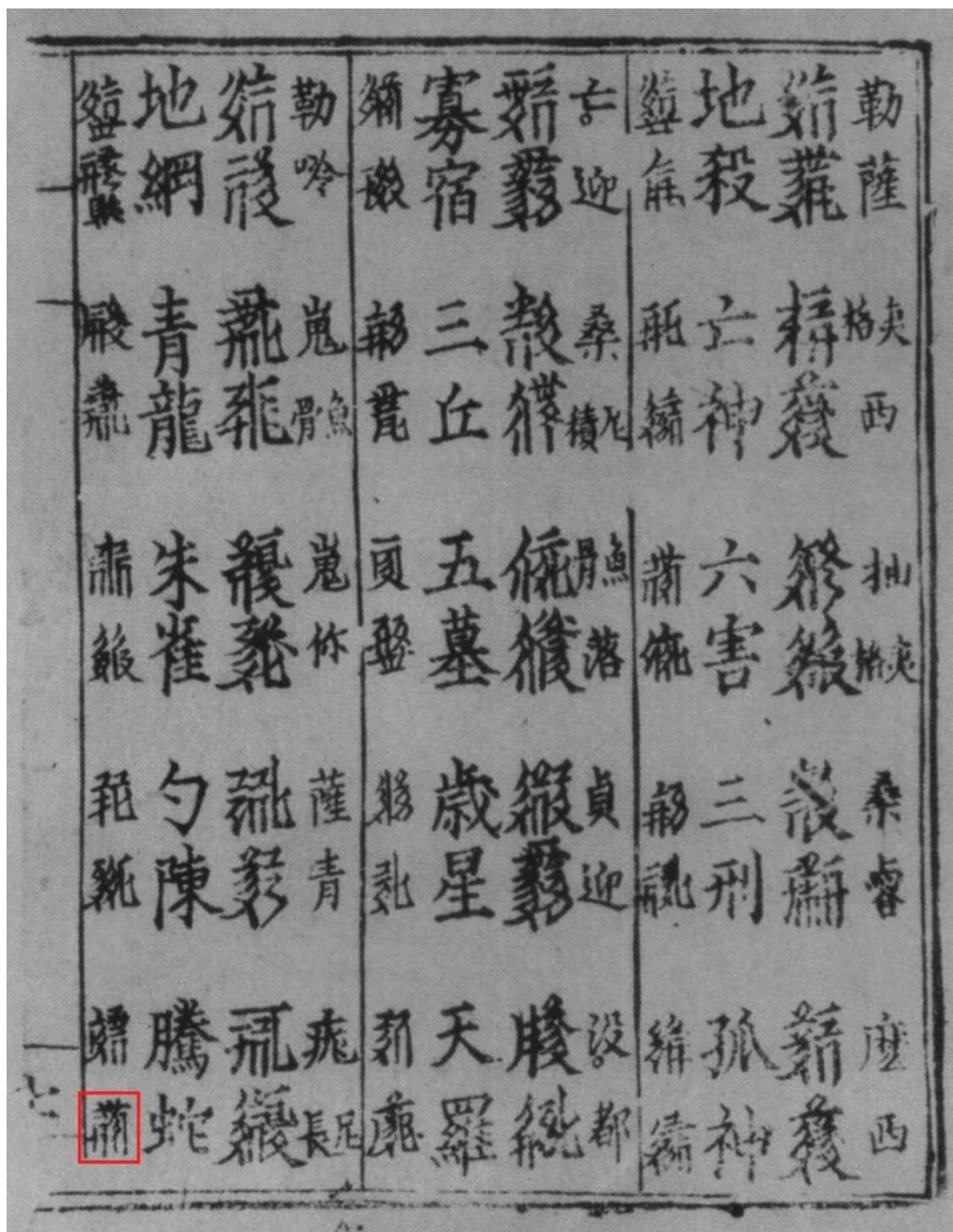


Fig. 25: IOM Tang.13: 《級殺形織蕪花辭絞》 Ed. B folio 7B¹²

The character 蕪 occurs in the phonetic gloss 蕪蕪 *thē² (síj¹)* for Chinese 騰蛇 *téng shé* 'soaring snake'. It is likely to be a non-unifiable variant of U+17D46 蕪 *síj¹*. This is the only example I have ever encountered where a character with Component 206 𠂔 is written with Component 106 𠂔.

¹² This entry also occurs in the A edition (folio 6B), but the gloss character is lost due to damage.

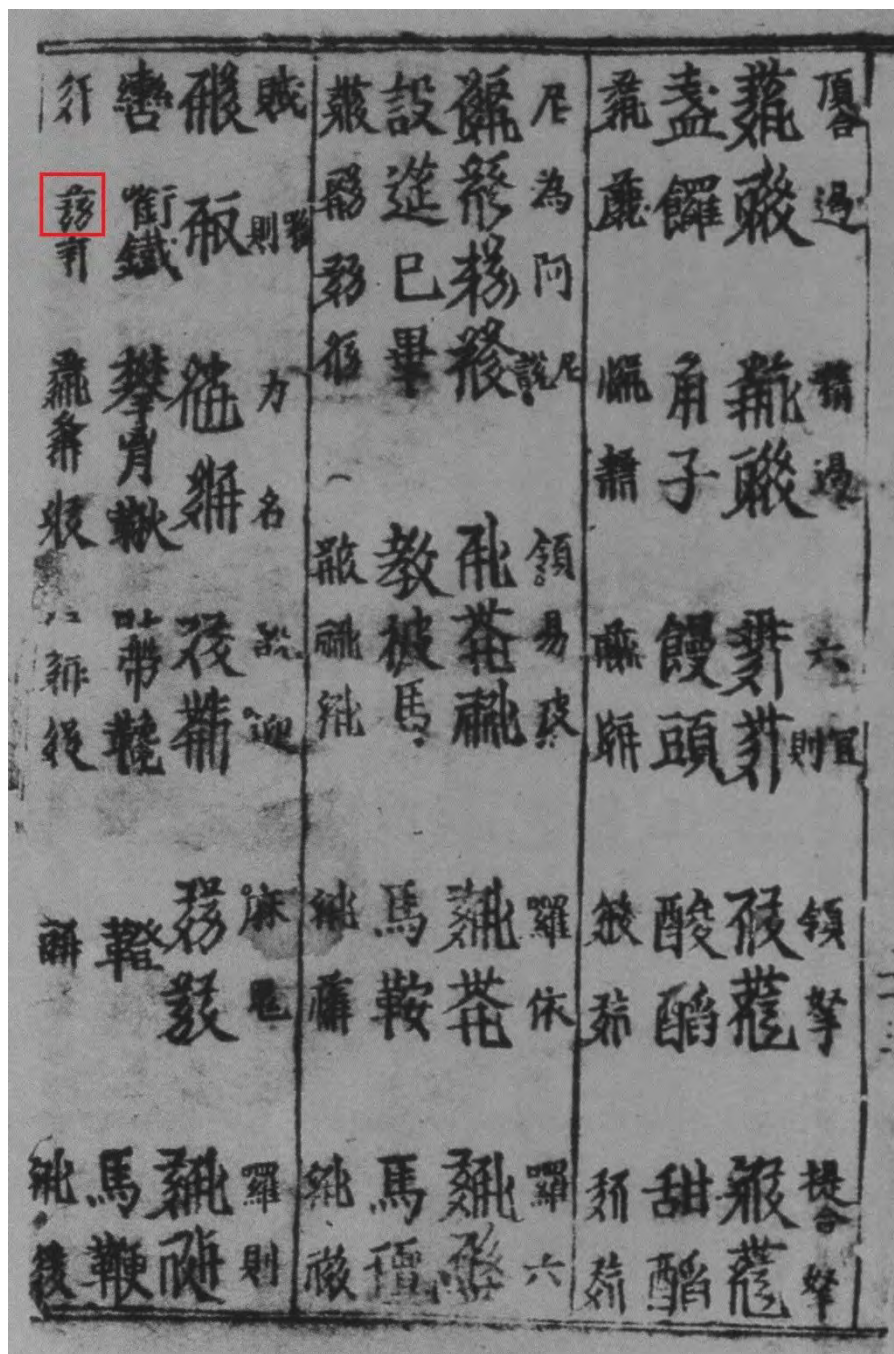


Fig. 26: IOM Tang.13: 《𐰇𐰏𐰑𐰒𐰓𐰔𐰕𐰖𐰗𐰘𐰙𐰚𐰛𐰜𐰝𐰞𐰟𐰠𐰡𐰢𐰣𐰤𐰥𐰦𐰧𐰨𐰩𐰪𐰫𐰬𐰭𐰮𐰯𐰰𐰱𐰲𐰳𐰴𐰵𐰶𐰷𐰸𐰹𐰺𐰻𐰼𐰽𐰾𐰿𐰠𐰡𐰢𐰣𐰤𐰥𐰦𐰧𐰨𐰩𐰪𐰫𐰬𐰭𐰮𐰯𐰰𐰱𐰲𐰳𐰴𐰵𐰶𐰷𐰸𐰹𐰺𐰻𐰼𐰽𐰾𐰿》 Ed. A folio 33B

The character 𐰇 occurs in the phonetic gloss 𐰇𐰏 (xā¹) thij² for Chinese 嚼鐵 xián tiě 'horse's bit'. It is likely to be a non-unifiable variant of U+17C54 𐰇 xā¹. This character is also given in Nevsky's draft manuscript Tangut dictionary (Fig. 27). This is the only example I have ever encountered where a character with Component 185 𐰇 is written with Component 75 𐰇.

| 翰 (厄卒訛) 御史 [珠, 28]
 罷 康 瓶 菠 補 瓠 瓠 恠 | 卍 諸王 統理 民物 不以
 節度 [七佛]

𪛗 𪛗 hien [珠, 33]

𪛗 宜會 R 上 30 Velars

𪛗 垂 c'ui R 平 10 (w 333) Palato-Alveolars
 齒。[壽。年齡]年

| 𪛗 𪛗 𪛗 (重六吃噉) 身齒已衰 [珠, 34]
 𪛗 𪛗 𪛗 年小弟子 [蓮, 14]

Fig. 27: *Tangutskaya filologiya* vol. I p. 292

4.18 Homophones



Fig. 28: IOM Tang.18: 《龍龍》 Ed. A folio 52B

The character 𠂔 occurs as the gloss for 𠂔 *ljwi*² ‘colt’ in the ‘A’ edition of *Homophones*. In the ‘B’ edition the gloss character is U+1733B 𠂔 *jar*² ‘to tame’, so 𠂔 is evidently a variant or mistake for 𠂔, but as it has a completely different left side component it cannot be considered to be a unifiable variant.

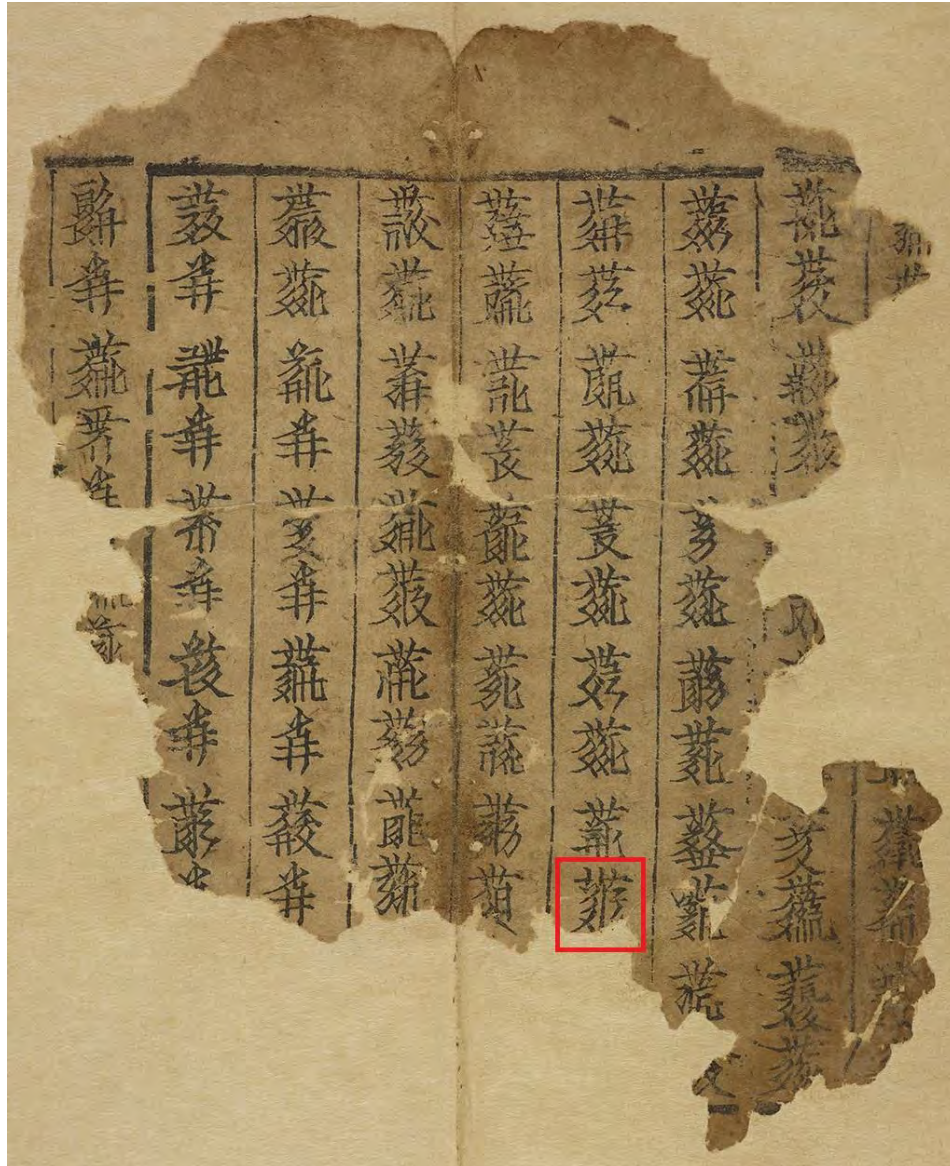


Fig. 29: British Library Or.12380/1843 (G): 《佩衮》

The character 蕞 occurs in the word 蕞蕞 (a type of tree or bush) in a woodblock printed edition of *Assorted Characters* held at the British Library. In the edition of *Assorted Characters* held at the IOM in Russia, this word is written as 蕞蕞 (Tang.19 folio 7A), so 蕞 should be a variant form of U+17768 蕞. However, it cannot be considered to be a unifiable variant because it uses a completely different bottom left component.

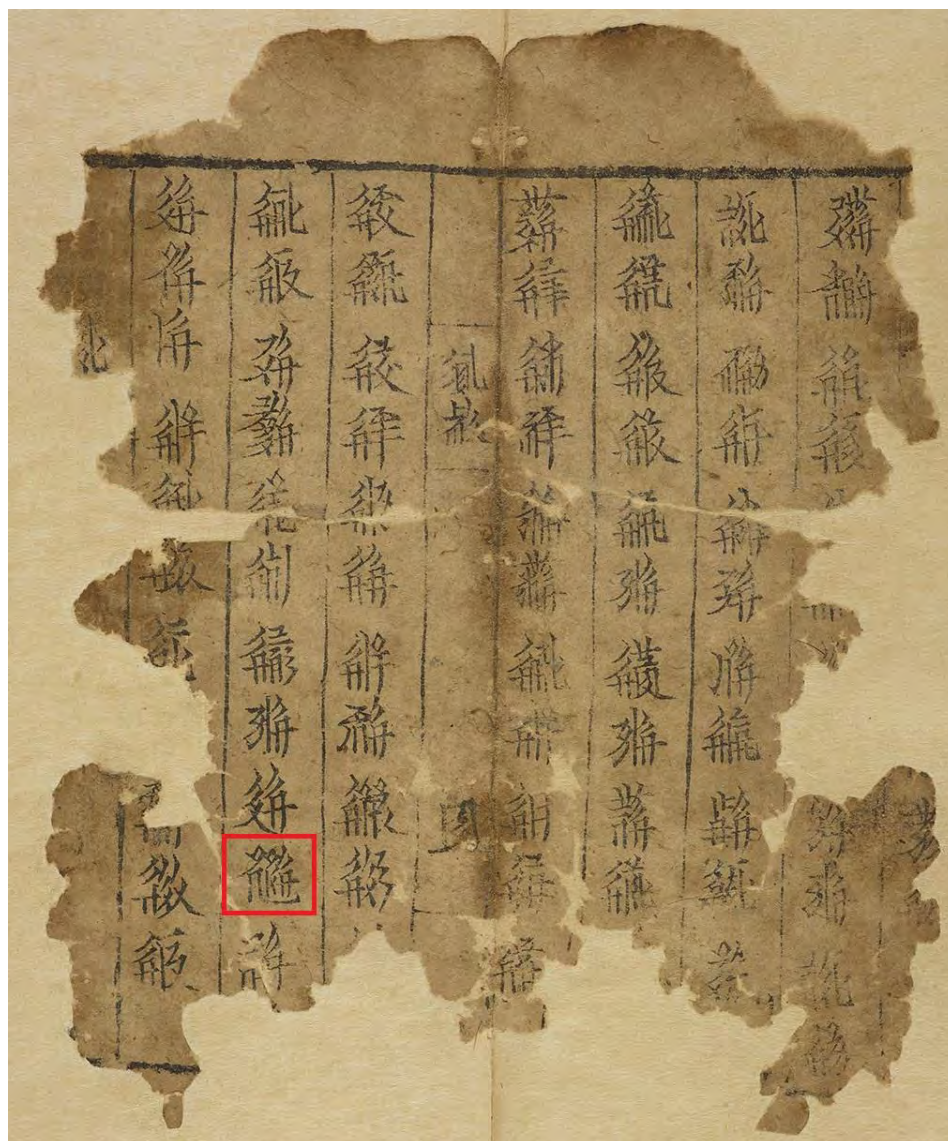


Fig. 30: British Library Or.12380/1843 (H): 《夙敫》

The character 繼 occurs in the word 筵繼 (a type of plant) in a woodblock printed edition of *Assorted Characters* held at the British Library. In the edition of *Assorted Characters* held at the IOM in Russia, this word is written as 筵繼 (Tang.19 folio 8A), so 繼 should be a variant form of U+17BC0 繼. However, it cannot be considered to be a unifiable variant because it uses a completely different middle component (Component 419 𠄎 instead of Component 112 𠄎).

4.21 Joined Rimes of the Five Sounds

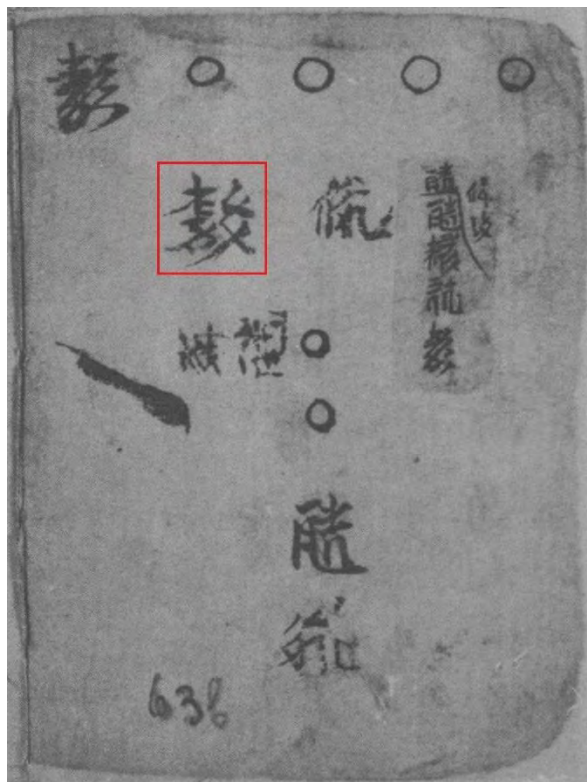


Fig. 31: IOM Tang.22: 《𪛗𪛘𪛙𪛚》 MS D folio 63B

The character 𪛗 occurs in the 'D' manuscript version of the Tangut rime tables, under the fourth column (for dental affricate initials) of Rime Table 60 𪛘 *tšjow*² (this character is not given in any of the other surviving manuscript versions of Rime Table 60). There are two small characters beneath it which should give the *fanqie* reading of the character (*fanqie* readings are only given for novel characters), but unfortunately they are illegible here. Lǐ Fànwén interprets this character as a mistake for U+184AB 𪛙 *zew*² (see Fig. 32), which should be under Rime Table 48, but none of the manuscripts give 𪛙 under Rime Table 48 or elsewhere, so his suggestion does not seem convincing.

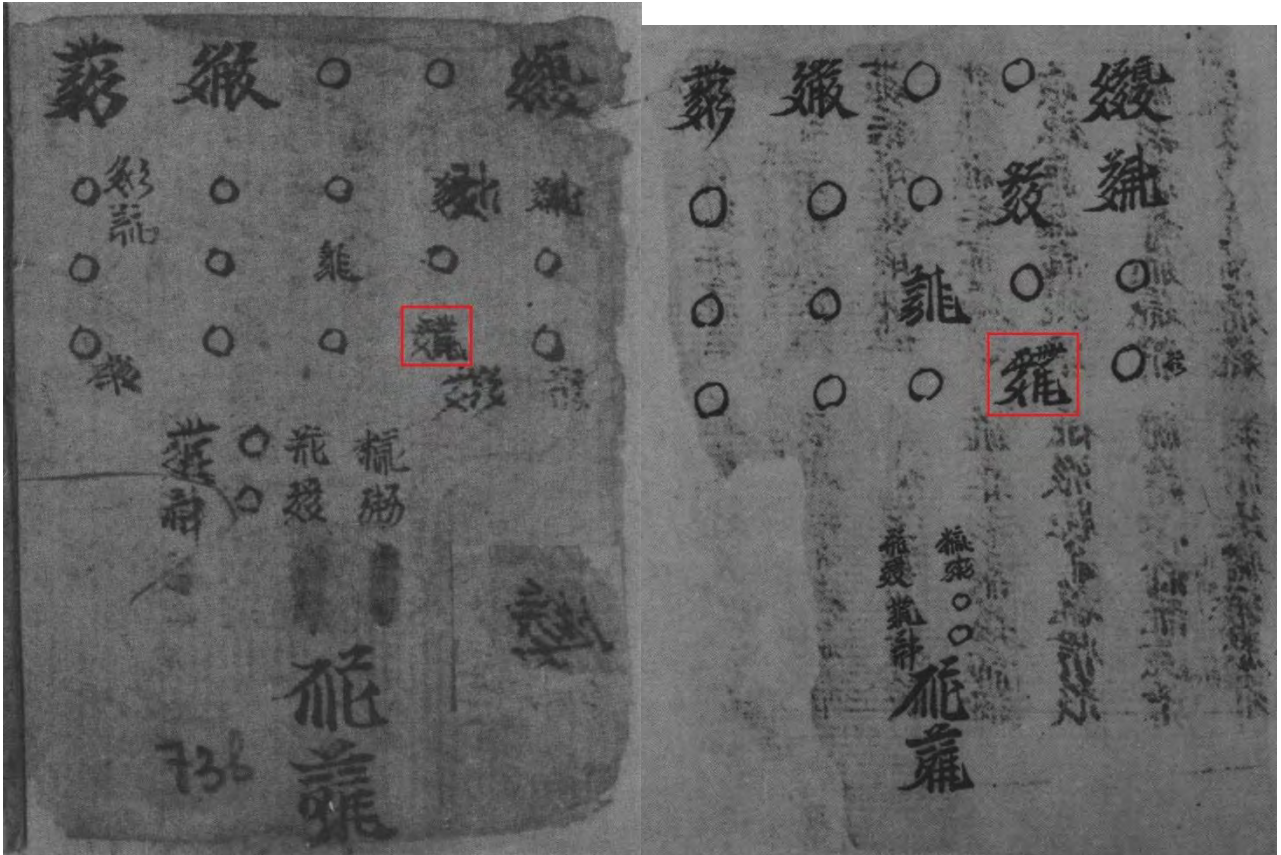
韵图 60 (2.50), 甲种本与丁种本是相同的。图上共 3 字, 除一字表示上声韵外, 只有两字。即第一段喉音 𪛗·jow 2.5 和上声代表字 𪛗 tsjow 2.50 外, 无其它字。丁种本上的 𪛗 笔误 (应为 𪛗 zew 2.41 遣送)、𪛗 me 2.7 巽 (八卦名) 二字并非第 60 (2.50) 韵图上的声韵代表字。韵图第二栏舌音下, 贴有上声 50 韵的四个韵字。译释如下:

① 𪛗 tsjow 2.50 喂养 ② 五 𪛗 ③ 𪛗 十 ④ 𪛗 tsjow 2.50 喂养 ⑤ 𪛗 tsjow 2.50 奉、随 ⑥ 𪛗 tsjow 2.50 无 ⑦ 𪛗·jow 2.50 炉。

《五音切韵》与《文海宝韵》各收 4 字。尚有 ① 𪛗·jow 2.50 爱 ② 𪛗 dziow 2.50 胜 ③ 𪛗 dziow 2.50 冷寒 3 字未收, 后 2 字在杂类中。

Fig. 32: 'Wǔyīn qièyùn' yǔ 'Wénhǎi bǎoyùn' Bǐjiào Yánjiū p. 183

4.22 Joined Rimes of the Five Sounds



4.23 *Joined Rimes of the Five Sounds*

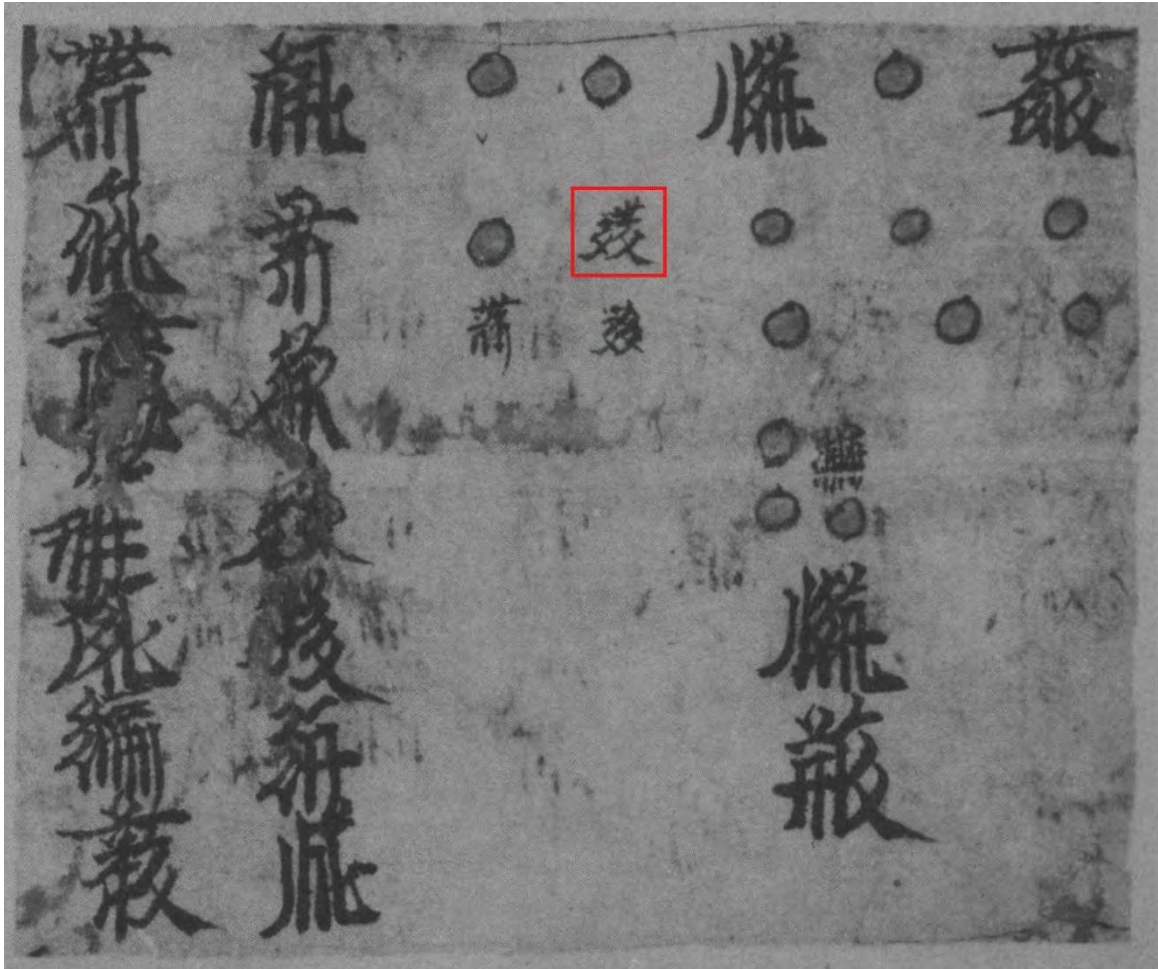


Fig. 34: IOM Tang.22: 《𪛗能𪛗𪛗𪛗》 MS F page 25

The 'A', 'D', and 'E' manuscripts of the Tangut rime tables all have U+17D91 𪛗 *tšiw*² in the second position of the fourth column (for dental affricate initials) of Rime Table 52 𪛗 *kio*¹ / 𪛗 *kio*². However, MS 'F' gives the unencoded character 𪛗 at this position. This is likely to be an error for 𪛗, but as the bottom right side is completely different it cannot be considered to be a unifiable variant.

4.24 *Joined Rimes of the Five Sounds*

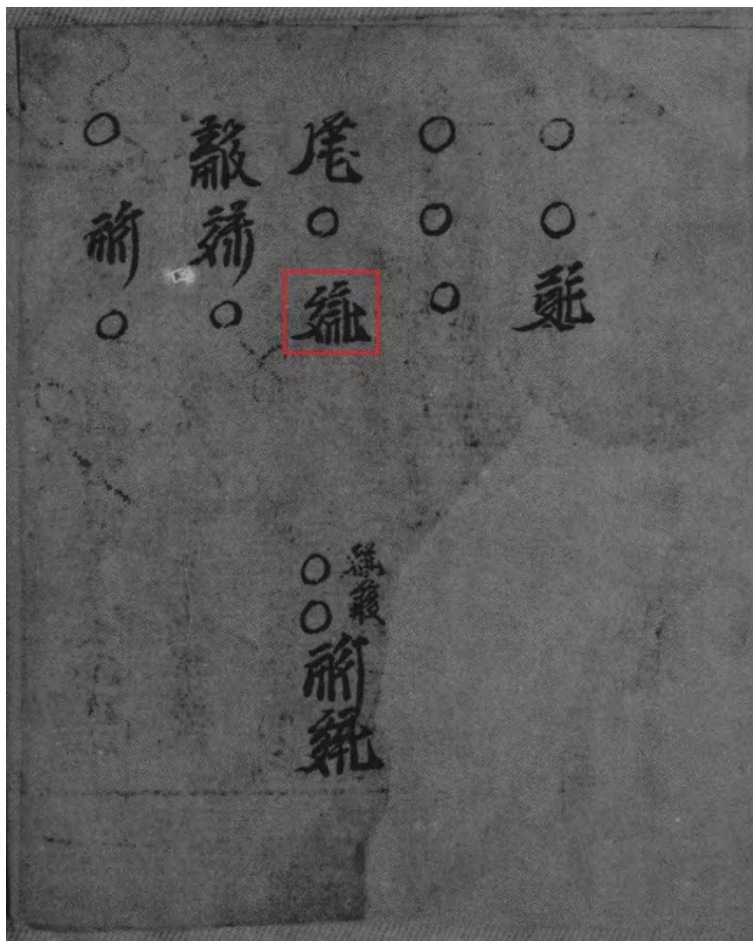


Fig. 35: IOM Tang.22: 《儀能被韻》 MS E page 52

The 'C' and 'D' manuscripts of the Tangut rime tables have U+18040 𪛗 *kjir*¹ in the third position of the third column of Rime Table 84 𪛗 *jir*¹ / 𪛗 *jir*². However, MS 'E' gives the unencoded character 𪛗 at this position. This is likely to be an error for 𪛗, but as it omits the 彡 component it cannot be considered to be a unifiable variant.

4.25 *Joined Rimes of the Five Sounds*

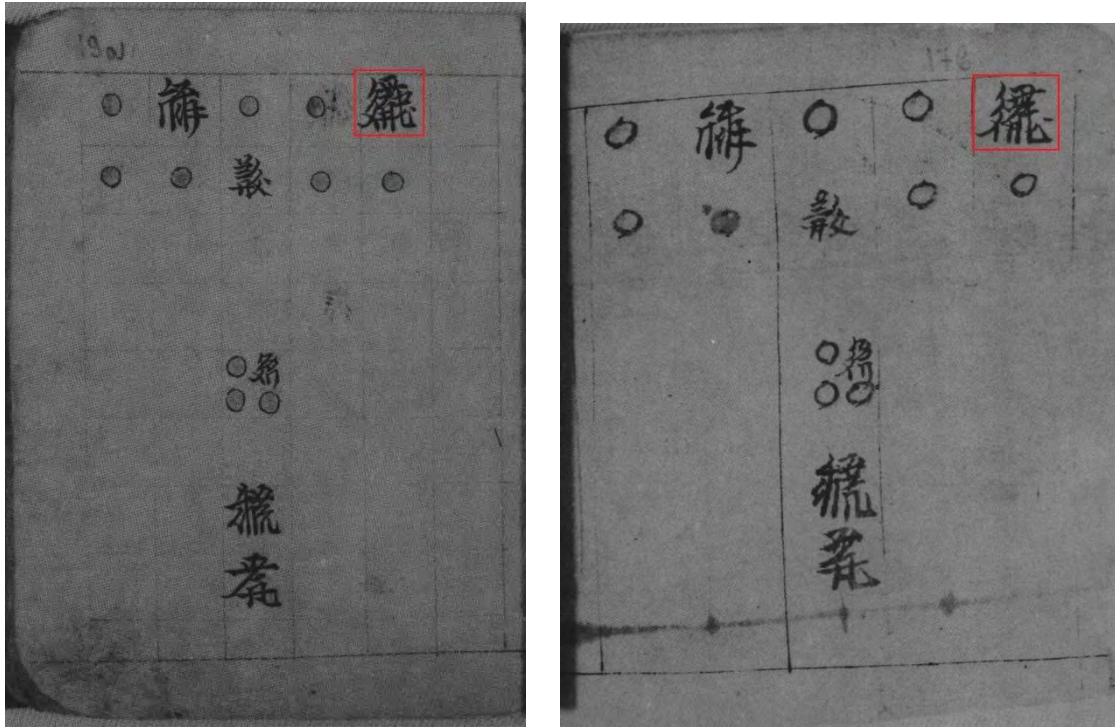


Fig. 36: IOM Tang.22: 《禡禡禡禡》 MS A folio 19A and MS B folio 17B

The 'E' manuscript of the Tangut rime tables has U+17AE3 禡 pa^1 in the first position of the first column of Rime Table 22 禡 ga^1 / 禡 ga^2 . However, the 'A' and 'B' manuscripts both have 禡 in this position. This may be a variant or mistake for 禡, but it cannot be considered to be a unifiable variant.

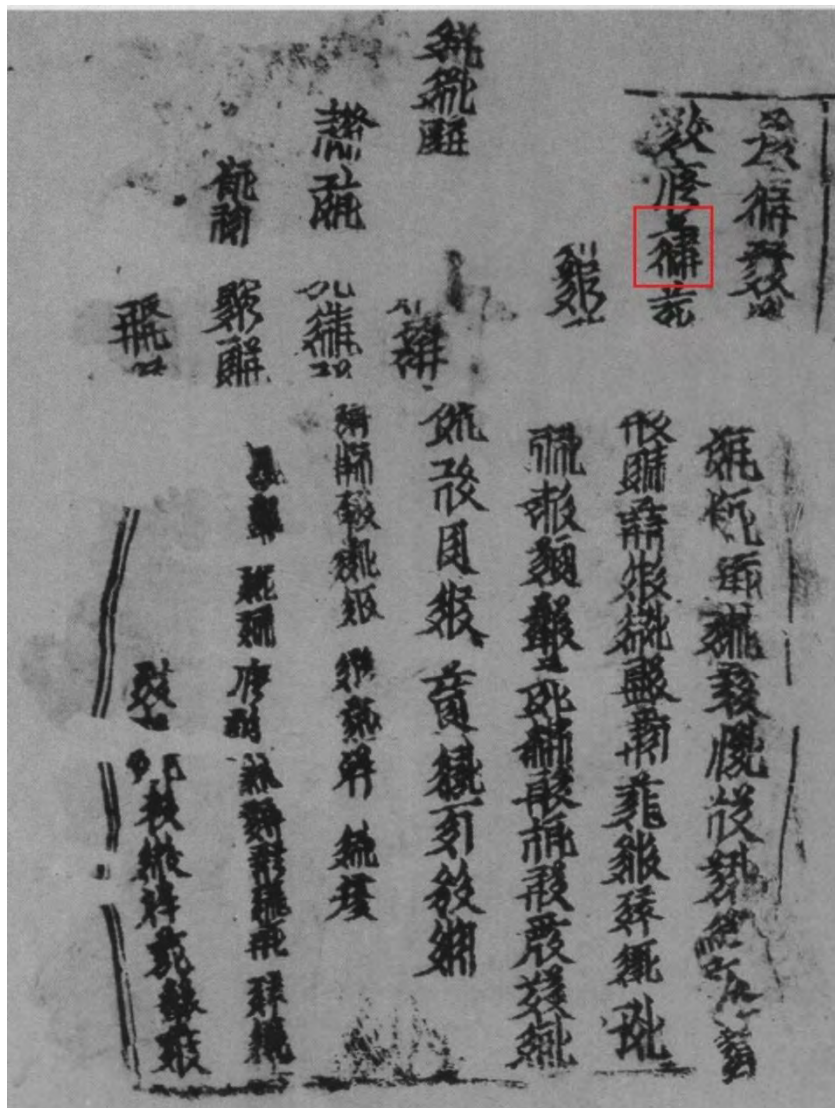


Fig. 37: IOM Tang. 35/1: 《疑祖會影疑疑》 folio 31B¹³

The character 禱 occurs in the phrase 效禱禱禱 in the colophon on the last page of *Newly Assembled Precious Dual Maxims*. Kychanov translates the phrase uncertainly as “те, кто открыли (?)” = “those who opened (?)” which would be a reasonable translation for 效禱禱, but does not assign any meaning to 禱.¹⁴ Niè Hóngyīn transcribes the character as U+174A9 禱 *nju*¹, but he omits the phrase with this character from his translation.¹⁵ 禱 is a rare character which only occurs in the family name 禱効 *nju*¹ *rjijr*², and does not make any sense here, so it is not convincing that 禱 is a variant form of 禱.

¹³ *Vnov' sobrannye dragotsennye parnye izrechenija* p. 212.

¹⁴ *Vnov' sobrannye dragotsennye parnye izrechenija* p. 125.

¹⁵ “Xìxià yíwén lù” p. 158.

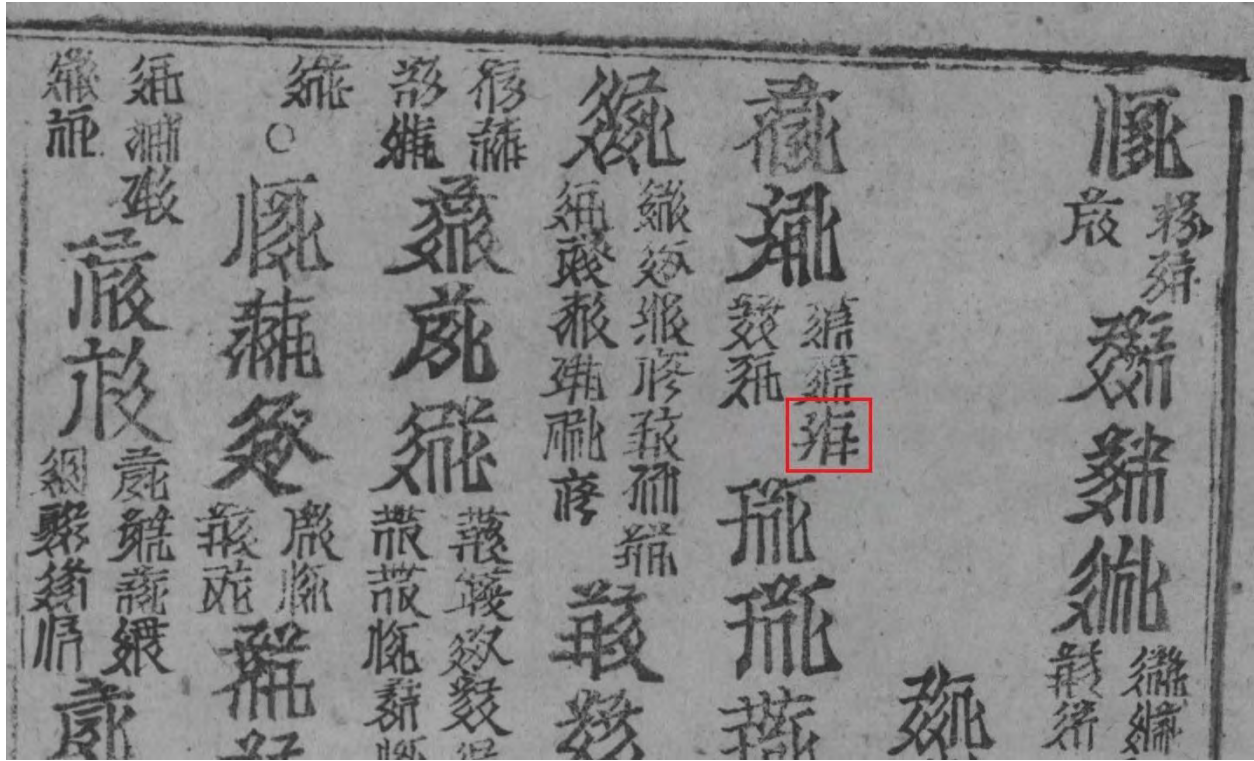


Fig. 38: IOM Tang.27: 《彘蕪蕪彘彘》 folio 9A

The character 彘 occurs in the phrase 蕪蕪彘彘, but it is not clear what it means here, and there is no obvious character that it could be a mistake or variant form of.

5. Bibliography

- Fǎcáng Dūnhuáng Xīxiàwén Wénxiàn* 法藏敦煌西夏文文獻 = *Documents XiXia de Dunhuang conservés à la Bibliothèque nationale de France*. Shanghai: Shanghai guji chubanshe, 2007.
- Kitajskaja klassika v tangutskom perevode. Kıtայская классика в тангутском переводе* (Лунь юй, Мэн цзы, Сяо цзин). Факсимиле текстов [Chinese classics in Tangut translation (*Lun yu, Meng zi, Xiao jing*). Facsimile of the texts]. B. C. Колоколов [V. S. Kolokolov] and E. И. Кычанов [E. I. Kuchanov] comps. Moscow: Nauka, 1966.
- Riběncáng Xīxiàwén Wénxiàn* 日本藏西夏文文獻 = *Tangut Manuscripts Collected in Japan*. Beijing: Zhonghua shuju, 2011.
- Tangut Version of the Lotus Sutra in the Collection of Princeton University Library, Facsimile, Text and Linguistic Studies*. プリンストン大学図書館所蔵西夏文妙法蓮華經 — 写真版及びテキストの研究. Arakawa Shintarō 荒川慎太郎 ed. *Lotus Sutra Manuscript Series* 16. Tōkyō: Soka Gakkai, Institute of Oriental Philosophy, 2018.
- Tangutskaya filologiya. Тангутская филология. Исследования и словарь* [Tangut philology. Research and dictionary]. Н. А. Невский [N. A. Nevskii]. Moscow, Oriental Literature Publishing House, 1960.
- ‘Tóngyīn Wénhǎi Bǎoyùn Hébiān’ Zhěnglǐ yǔ Yánjiū* 《同音文海宝韵合编》整理与研究 [Collation and study of *Tóngyīn Wénhǎi Bǎoyùn Hébiān*]. Hán Xiǎománg 韓小忙. Beijing: Zhongguo shehui kexue chubanshe, 2008.
- Vnov’ sobrannye dragotsennye parnye izrechenija. Вновь собранные драгоценные парные парные изречения* [The Newly-Assembled Precious Dual Maxims]. E. I. Kuchanov (E. И. Кычанов) ed. Moscow: Nauka, 1974.
- ‘Wǔyīn qièyùn’ yǔ ‘Wénhǎi bǎoyùn’ Bǐjiào Yánjiū* 《五音切韵》与《文海宝韵》比较研究 [Comparative study of *Wǔyīn qièyùn* and *Wénhǎi bǎoyùn*]. Lǐ Fànwén 李范文. *Xīxià Yánjiū* 西夏研究 [Tangut Studies] No. 2. Beijing: Zhongguo shehui kexue chubanshe, 2006.
- Xīxià Fójiào Shǐlùè* 西夏佛教史略. Shǐ Jīnbō 史金波. Yinchuan: Ningxia renmin chubanshe, 1988.
- Xīxiàwén Cídiǎn* 西夏文词典. Hán Xiǎománg 韓小忙 comp. 9 vols. Beijing: Zhongguo Shehui kexue chubanshe, 2021.
- Xīxiàwén Zìdiǎn* 西夏文字典. Jiǎ Chángyè 贾常业 comp. Lanzhou: Gansu wenhua chubanshe, 2019.
- “Xīxià yíwén lù” 西夏遗文录 [Record of documents from the Western Xià]. Niè Hóngyīn 聂鸿音. In Dù Jiàn lù 杜建录 (ed.), *Xīxià Xué* 西夏学 = *Xixia Studies* No. 2 pp. 134–166. Yinchuan: Ningxia renmin chubanshe, 2007.
- Zhōngguó cáng Xīxià Wénxiàn* 中國藏西夏文獻 = *Tangut Manuscripts Collected in China*. 20 vols. Lanzhou: Gansu People’s Publishing House and Dunhuang Literature and Art Publishing House, 2005.

6. Acknowledgments

My thanks Viacheslav Zaytsev, Romain Lefebvre, Sūn Bójūn 孫伯君, and 'hqsjg' for bringing some of the proposed characters to my attention, and for helpful discussion. Thanks to Jerry You for comments on the first version of the document.

7. Proposal Summary Form

**ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646¹⁶**

Please fill all the sections A, B and C below.

Please read Principles and Procedures Document (P & P) from <http://www.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://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html>.

See also <http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html> for latest Roadmaps.

A. Administrative

1. Title:	<i>Proposal to encode 2 Tangut components and 29 Tangut ideographs</i>
2. Requester's name:	<i>Andrew West</i>
3. Requester type (Member body/Liaison/Individual contribution):	<i>Individual contribution</i>
4. Submission date:	<i>2023-10-02</i>
5. Requester's reference (if applicable):	
6. Choose one of the following:	
This is a complete proposal:	<input checked="" type="checkbox"/> YES
(or) More information will be provided later:	<input type="checkbox"/>

B. Technical – General

1. Choose one of the following:	
a. This proposal is for a new script (set of characters):	<input checked="" type="checkbox"/> NO
Proposed name of script:	
b. The proposal is for addition of character(s) to an existing block:	<input checked="" type="checkbox"/> YES
Name of the existing block:	<i>Tangut Supplement</i>
2. Number of characters in proposal:	<i>30</i>
3. Proposed category (select one from below - see section 2.2 of P&P document):	
A-Contemporary <input type="checkbox"/> B.1-Specialized (small collection) <input type="checkbox"/> B.2-Specialized (large collection) <input type="checkbox"/>	
C-Major extinct <input type="checkbox"/> D-Attested extinct <input type="checkbox"/> E-Minor extinct <input type="checkbox"/>	
F-Archaic Hieroglyphic or Ideographic <input checked="" type="checkbox"/> G-Obscure or questionable usage symbols <input type="checkbox"/>	
4. Is a repertoire including character names provided?	<input checked="" type="checkbox"/> YES
a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?	<input checked="" type="checkbox"/> YES
b. Are the character shapes attached in a legible form suitable for review?	<input checked="" type="checkbox"/> YES
5. Fonts related:	
a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?	<i>Andrew West</i>
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):	<i>Andrew West</i>
6. References:	
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?	<input checked="" type="checkbox"/> YES
b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?	<input checked="" type="checkbox"/> 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)?	<input type="checkbox"/> NO

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.

¹⁶ Form number: N4102-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11, 2005-01, 2005-09, 2005-10, 2007-03, 2008-05, 2009-11, 2011-03, 2012-01)

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before? If YES explain	NO
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.)? If YES, with whom? If YES, available relevant documents:	NO
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Reference:	NO
4. The context of use for the proposed characters (type of use; common or rare) Reference:	rare
5. Are the proposed characters in current use by the user community? If YES, where? Reference:	YES
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? If YES, is a rationale provided? If YES, reference:	NO
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? If YES, is a rationale for its inclusion provided? If YES, reference:	NO
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? If YES, is a rationale for its inclusion provided? If YES, reference:	NO
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? If YES, is a rationale for its inclusion provided? If YES, reference:	NO
11. Does the proposal include use of combining characters and/or use of composite sequences? 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:	NO
12. Does the proposal contain characters with any special properties such as control function or similar semantics? If YES, describe in detail (include attachment if necessary)	NO
13. Does the proposal contain any Ideographic compatibility characters? If YES, are the equivalent corresponding unified ideographic characters identified? If YES, reference:	NO