

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 29 Tangut ideographs

Source: Andrew West

Status: Individual Contribution

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

Date: 2023-05-31

1. Introduction

This is a proposal to encode two Tangut components shown in Table 1, and twenty-nine 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.

Table 1: Proposed Tangut Components

Ref.	Glyph	Sources
3.1	𐰪	<i>Combined Edition of Homophones and Sea of Writing 'C' manuscript</i>
3.2	𐰪	𐰪 𐰪 𐰪 𐰪 𐰪 𐰪 𐰪 (Homonyms)

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.	Glyph	IDS	R/S	Sources
4.1	𡇗	𡇗 𡇗 𡇗 𡇗 𡇗 𡇗	206.19	經攏𡇗𡇗𡇗𡇗 = 重譯孝經傳 (<i>Classic of Filial Piety with Commentary</i>)
4.2	𡇘	𡇘 𡇘 𡇘 𡇘 𡇘 𡇘	236.15	經攏𡇘𡇘𡇘𡇘𡇘 = 重譯孝經傳 (<i>Classic of Filial Piety with Commentary</i>)
4.3	𡇙	𡇙 𡇙 𡇙	267.9	𡇙𡇙 = 論語 (<i>Analects of Confucius</i>) vols. 6 and 8
4.4	𡇚	𡇚 𡇚 𡇚 𡇚 𡇚 𡇚	75.19	𡇚𡇚𡇚𡇚𡇚𡇚𡇚𡇚 = 金光明最勝王經 (<i>Suvarṇaprabhāsa Sūtra</i>) vol. 4 𡇚𡇚𡇚𡇚𡇚𡇚𡇚𡇚 (<i>Homonyms</i>)
4.5	𡇛	𡇛 𡇛 𡇛	490.12	𡇛𡇛𡇛𡇛𡇛𡇛𡇛𡇛 = 金光明最勝王經 (<i>Suvarṇaprabhāsa Sūtra</i>) vol. 4
4.6	𡇜	𡇜 𡇜 𡇜	114.9	𡇜𡇜𡇜𡇜𡇜𡇜𡇜𡇜 = 金光明最勝王經 (<i>Suvarṇaprabhāsa Sūtra</i>) vol. 6
4.7	𡇝	𡇝 𡇝 𡇝 𡇝 𡇝 𡇝	217.15	𡇝𡇝𡇝𡇝𡇝𡇝𡇝𡇝 = 大方廣佛華嚴經 (<i>Mahāvaiṣṭvya Buddhāvataṃsaka Sūtra</i>) vol. 41
4.8	𡇞	𡇞 𡇞 𡇞	519.13	𡇞𡇞𡇞𡇞𡇞𡇞𡇞𡇞 = 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
4.9	𡇟	𡇟 𡇟 𡇟 𡇟 𡇟 𡇟	736.18	𡇟𡇟𡇟𡇟𡇟𡇟𡇟𡇟 = 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	𡇠	𡇠 𡇠 𡇠	590.14	𡇠𡇠𡇠𡇠𡇠𡇠𡇠𡇠 = 過去莊嚴劫千佛名經 (<i>Sūtra of the Names of the Thousand Buddhas of the Past Kalpa</i>)
4.11	𡇡	𡇡 𡇡 𡇡 𡇡 𡇡 𡇡	267.18	𡇡𡇡𡇡𡇡𡇡𡇡𡇡𡇡 = 妙法蓮華經 (<i>Saddharma Puṇḍarīka</i>)
4.12	𡇢	𡇢 𡇢 𡇢	267.10	𡇢𡇢𡇢𡇢𡇢𡇢𡇢𡇢 (<i>Homonyms</i>)
4.13	𡇣	𡇣 𡇣 𡇣	655.13	𡇣𡇣𡇣𡇣𡇣𡇣𡇣𡇣 (<i>Homonyms</i>)
4.14	𡇤	𡇤 𡇤 𡇤	456.14	𡇤𡇤𡇤𡇤𡇤𡇤𡇤𡇤 (<i>Homonyms</i>)

Ref.	Glyph	IDS	R/S	Sources
4.15	𪛗	𪛗 𪛗 𪛗	273.9	𪛗 𪛗 𪛗 𪛗 𪛗 𪛗 𪛗 (Homonyms)
4.16	𪛘	𪛘 𪛘 𪛘	278.11	𪛘 𪛘 𪛘 𪛘 𪛘 𪛘 𪛘 = 番漢合時掌中珠 (<i>Pearl in the Palm</i>)
4.17	𪛙	𪛙 𪛙 𪛙	106.13	𪛘 𪛘 𪛘 𪛘 𪛘 𪛘 𪛘 = 番漢合時掌中珠 (<i>Pearl in the Palm</i>)
4.18	𪛚	𪛚 𪛚 𪛚	75.8	𪛘 𪛘 𪛘 𪛘 𪛘 𪛘 𪛘 = 番漢合時掌中珠 (<i>Pearl in the Palm</i>)
4.19	𪛛	𪛛 𪛛	17.7	𪛛 𪛛 (Homophones)
4.20	𪛜	𪛜 𪛜 𪛜	106.15	𪛛 𪛛 = 雜字 (<i>Assorted Characters</i>)
4.21	𪛝	𪛝 𪛝 𪛝	167.15	𪛛 𪛛 = 雜字 (<i>Assorted Characters</i>)
4.22	𪛞	𪛞 𪛞	462.11	𪛞 𪛞 𪛞 𪛞 = 五音切韻 (<i>Joined Rimes of the Five Sounds</i>)
4.23	𪛟	𪛟 𪛟 𪛟	579.17	𪛞 𪛞 𪛞 𪛞 = 五音切韻 (<i>Joined Rimes of the Five Sounds</i>)
4.24	𪛠	𪛠 𪛠 𪛠	210.13	𪛞 𪛞 𪛞 𪛞 = 五音切韻 (<i>Joined Rimes of the Five Sounds</i>)
4.25	𪛡	𪛡 𪛡 𪛡	278.13	𪛞 𪛞 𪛞 𪛞 = 五音切韻 (<i>Joined Rimes of the Five Sounds</i>)
4.26	𪛢	𪛢 𪛢 𪛢	141.19	𪛞 𪛞 𪛞 𪛞 = 五音切韻 (<i>Joined Rimes of the Five Sounds</i>)
4.27	𪛣	𪛣 𪛣 𪛣	75.13	𪛣 𪛣 𪛣 𪛣 (Newly Assembled Precious Dual Maxims)
4.28	𪛤	𪛤 𪛤	36.7	𪛤 𪛤 𪛤 𪛤 𪛤 𪛤 (Collected Writings of the Shining Speech of Three Generations)
4.29	𪛥	𪛥 𪛥 𪛥	141.9	Stele of Sulaiman

2. Code Point Allocation

2.1 Tangut Ideographs

This document does not suggest code points for the proposed characters. There are currently eight free code points at the end of the Tangut block at 187F8..187FF, and 119 free code point in the Tangut Supplement block at 18D09..18D7F. Therefore, the 29 proposed Tangut ideographs could either be allocated at 187F8..187FF (eight characters) and 18D09..18D1D (twenty-one characters), or allocated at 18D09..18D25 (twenty-nine characters). In Table 2 these twenty-nine characters are ordered according to their sources, but they could be reordered by radical and stroke order if that is considered more appropriate.

In Unicode 13.0 the Tangut Supplement block was allocated as 144 code points at 18D00..18D8F (<https://www.unicode.org/Public/13.0.0/ucd/Blocks.txt>). In Unicode 14.0 the size of the Tangut Supplement block was revised to 128 code points at 18D00..18D7F (<https://www.unicode.org/Public/14.0.0/ucd/Blocks.txt>). After spending several years researching the characters proposed in this document, I have reached the conclusion that it was a mistake to allocate such a large size for the Tangut Supplement block, as it is very unlikely that a large number of new Tangut characters for encoding will be forthcoming unless variant characters which I personally consider to be unifiable (and best represented using IVS) are accepted for encoding (and if that is the case then the Tangut Supplement block will not be sufficient). Therefore, I suggest that WG2 and UTC consider further reducing the allocation for the Tangut Supplement block to 18D00..18D3F (64 code points), which would leave 34 free code points in the Tangut and Tangut Supplement blocks after encoding the 29 ideographs proposed in this document.

2.2 Tangut Components

The Tangut Components block is completely full. Therefore the 114 Tangut components proposed in WG2 N5218 and the two Tangut components proposed in this document will need to be allocated in a new Tangut Components Supplement block. I suggest a block size of 128 code points if the characters in both proposals are accepted for encoding.

3. Evidence and Notes for Tangut Components

3.1 Component 1 𐞁

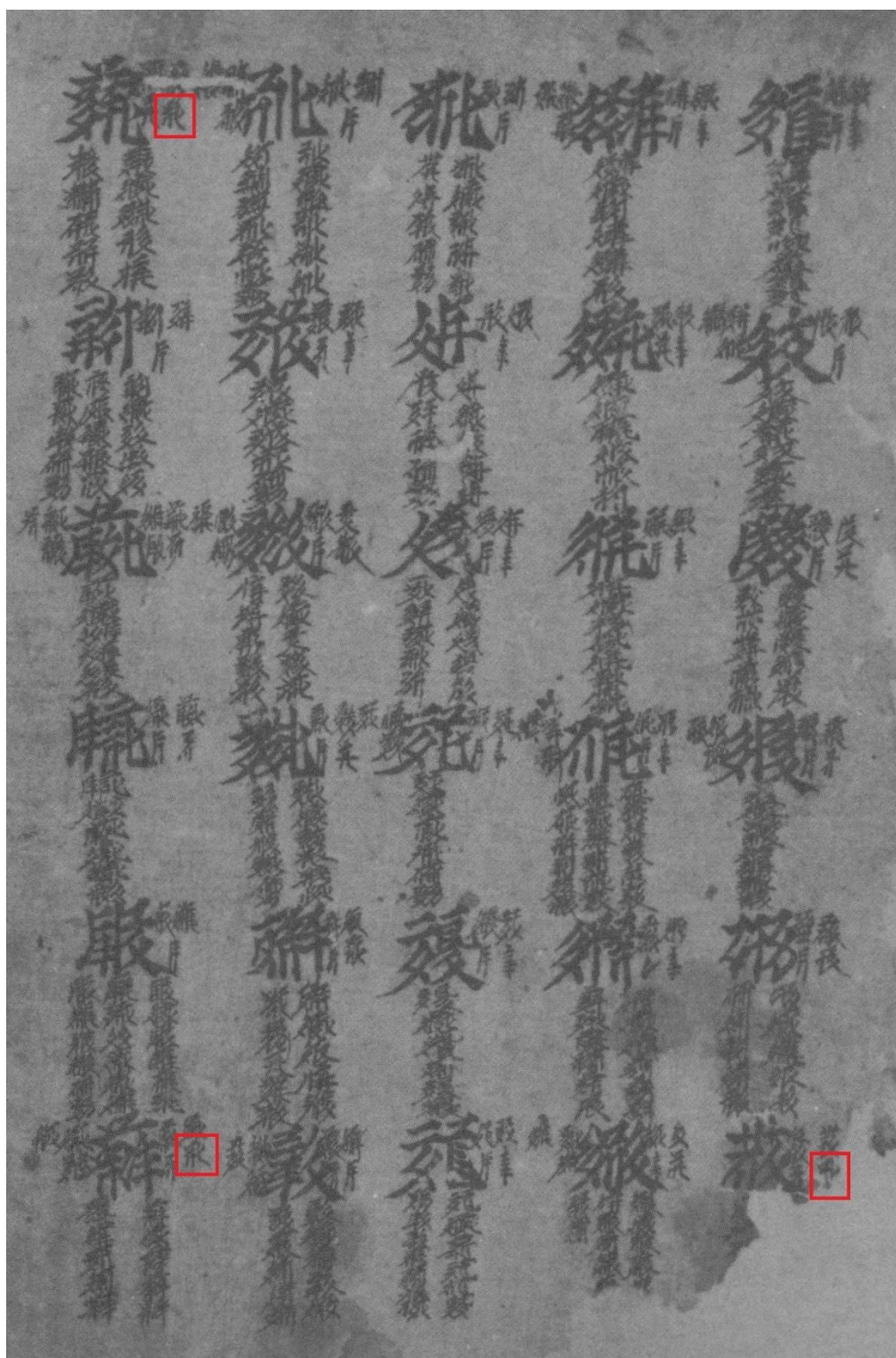


Fig. 1: IOM Inv. No. 4152 folio A ¹

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
𐰇	<i>pha</i> ¹	left side	𐰇	394
𐰈	<i>bjir</i> ²	right side	𐰈	118
𐰉	<i>yu</i> ¹	head (top)	𐰉	
𐰊	<i>tshjij</i> ¹	bottom	𐰊	89
𐰋	<i>io</i> ¹	surrounding	𐰋	428
𐰌	<i>njjj</i> ¹	middle	𐰌	9
𐰍	<i>ɲowr</i> ²	whole	𐰍	408

Six of the seven abbreviations are representable using existing encoded Tangut components, but the abbreviation for 𐰉 ‘head’ (𐰉) 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	𐰇	𐰇 𐰉 𐰇 𐰇	Top of 𐰇 and bottom of 𐰇
A51	𐰇	𐰇 𐰉 𐰇 𐰇	Top of 𐰇 and whole of 𐰇
A56	𐰇	𐰇 𐰉 𐰇 𐰇	Top [right] of 𐰇 and bottom of 𐰇
B14	𐰇	𐰇 𐰉 𐰇 𐰇	Top of 𐰇 and whole of 𐰇
B55	𐰇	𐰇 𐰉 𐰇 𐰇	Top of 𐰇 and whole of 𐰇

3.2 Component 2 𐰢



Fig. 2: 《𐰢𐰣𐰤𐰥𐰦𐰧𐰨𐰩𐰪𐰫𐰬𐰭𐰮𐰯𐰰𐰱𐰲𐰳𐰴𐰵𐰶𐰷𐰸𐰹𐰺𐰻𐰼𐰽𐰾𐰿》 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 𐰢, and assumed the damaged component above it was Component 172 𐰢 (see N4957 Table 1 Row 27). However, a more careful inspection of the sample character shows that it is actually U+180F8 𐰢, and the damaged component is actually 𐰢.

4. Evidence and Notes for Tangut Ideographs

4.1 Classic of Filial Piety with Commentary

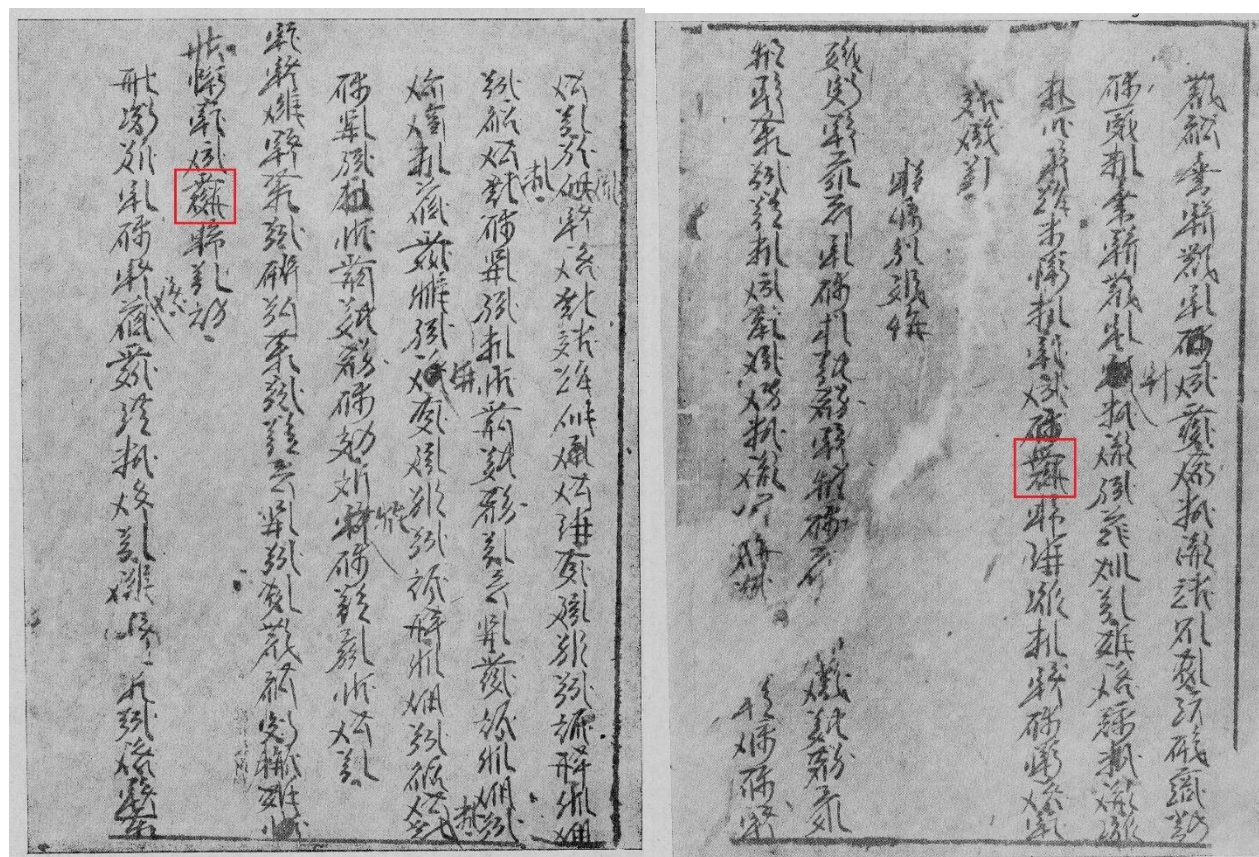


Fig. 3: IOM Tang. 1 Inv. No. 2627: 《終觀軒轅駘》 folios 6A and 7A ²

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 *Xīxiàwén Zìdiǎn* (2019) and in Hán Xiǎománg's *Xīxiàwén Cídiǎn* (2021), where it is regularized as 𐰇𐰺𐰍 by both authors. Jiǎ Chángyè gives the character construction as the top and bottom left-middle of 𐰇𐰺𐰍 *tsja*¹ 'road' and the right side of 𐰇𐰺𐰍 *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 𐰇 (Component 141) rather than 𐰇 (Component 767) or 𐰇 (Component 278), which would suggest a regularised form of 𐰇𐰺𐰍. 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 Classic of Filial Piety with Commentary

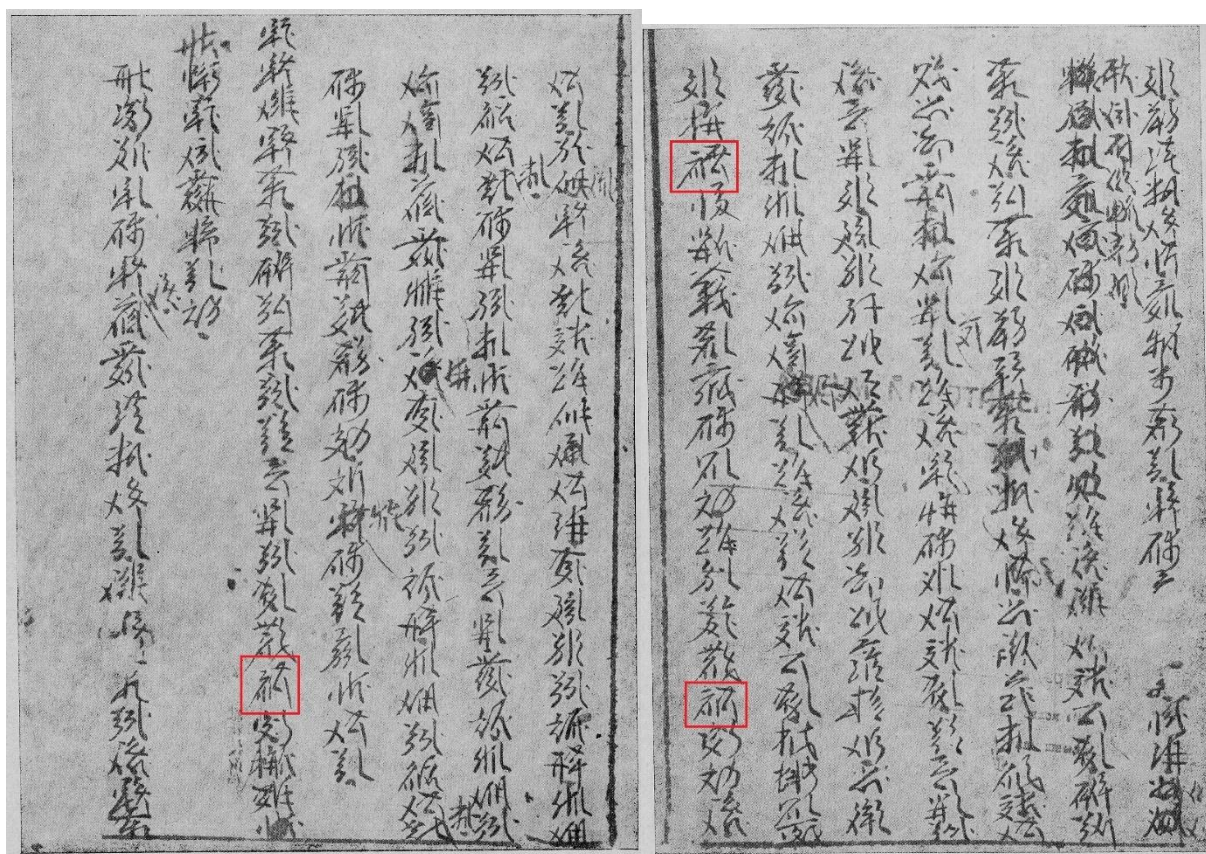


Fig. 6: IOM Tang. 1 Inv. No. 2627: 《孝經》 folios 6A and 6B³

This character occurs three times in the cursive manuscript of the Tangut translation of the *Classic of Filial Piety*. It is included in Hán Xiǎománg's *Xìxiàwén Cídiǎn* (2021) where it is regularized as 𪛗 (Fig. 7).

³ Kitajskaja klassika v tangutskom perevode pp. 145 and 146.

K. ? L. 6075 S. ?

[illegible]

Fig. 7: *Xīxiàwén Cídiǎn* (2021) vol. 5 p. 246

4.3 *Analects of Confucius*

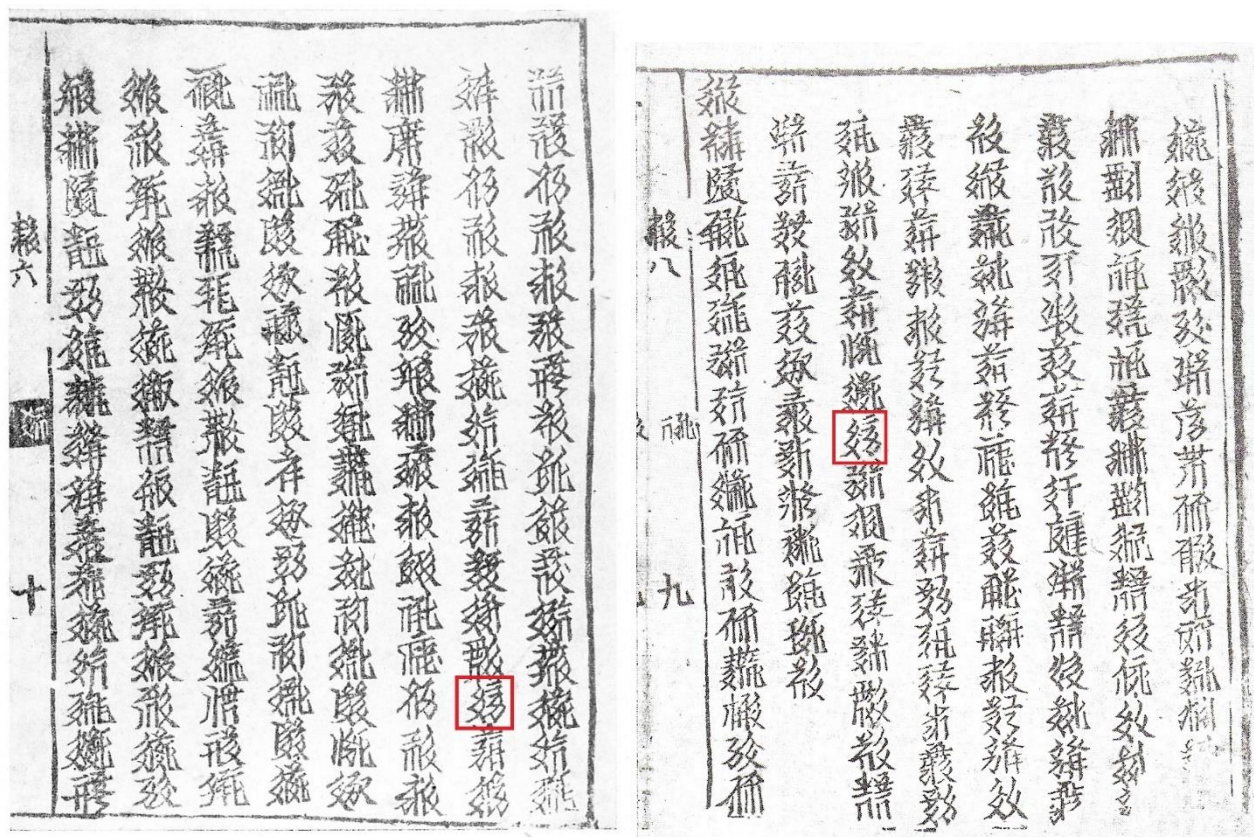


Fig. 8: IOM Tang. 1 Inv. No. 2627: 《𪛗𪛗》 6:10A and 8:9A ⁴

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 𪛗 (Fig. 9). However, the original woodblock printed edition shows that the left side is Component 766 (𪛗) not Component 267 (𪛗), so the correct form should be 𪛗.

⁴ Kitajskaja klassika v tangutskom perevode pp. 23 and 36; also one further example on p. 41.

左彡/右彡

K. ? L. ? S. ?

〔音〕?：声韵不详。

[illegible]

Fig. 9: *Xīxiàwén Cídiǎn* (2021) vol. 6 p. 7

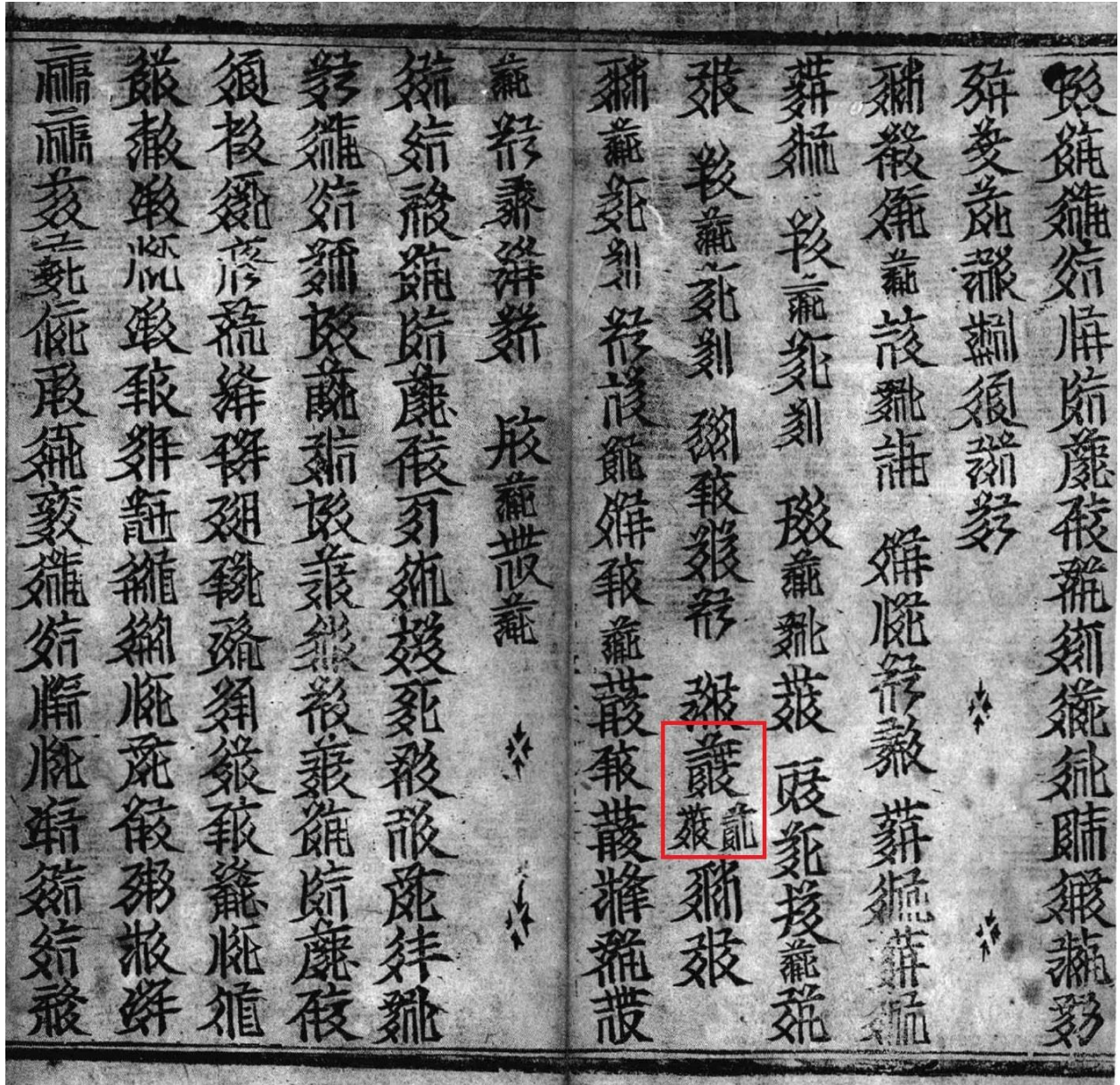


Fig. 10: National Library of China B11-028: 《繆敗傷癒禪席瓊茲》 𪛗𪛗𪛗⁵

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

This character also occurs in *Homonyms* under the homophone group 𪛗 *dwu*² ‘similar’, together with the characters 𪛗 *dwu*² and 𪛗 *twu*¹ (Fig. 11), and so should have the reading *dwu* or *twu*, which is similar to the reading *djo*² given in the *dhāraṇī* incantation.

⁵ Zhōngguó cáng Xìxià Wénxiàn vol. 3 p. 155.

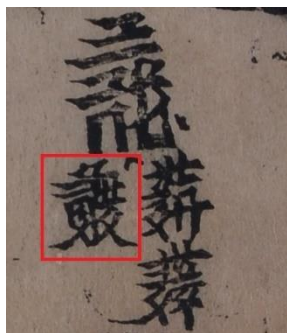


Fig. 11: 《𦵿𦵿𦵿𦵿𦵿𦵿𦵿𦵿》 folio 12B

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

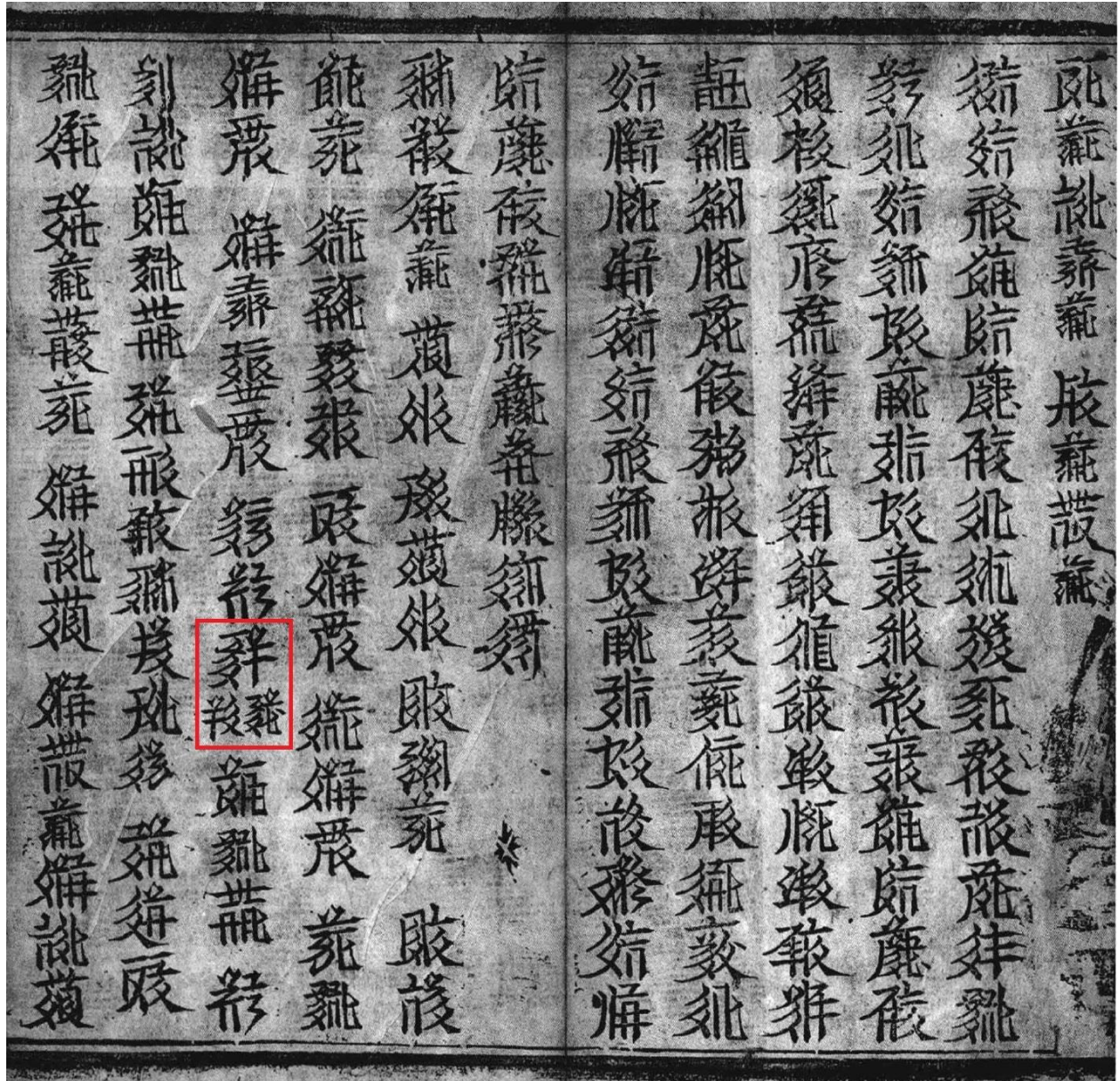


Fig. 12: National Library of China B11-028: 《釋迦傳經疏》 翻經 6

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 *nji*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 𑖦𑖦𑖦 in the Chinese version).

⁶ Zhōngguó cáng Xìxià Wénxiàn vol. 3 p. 162.

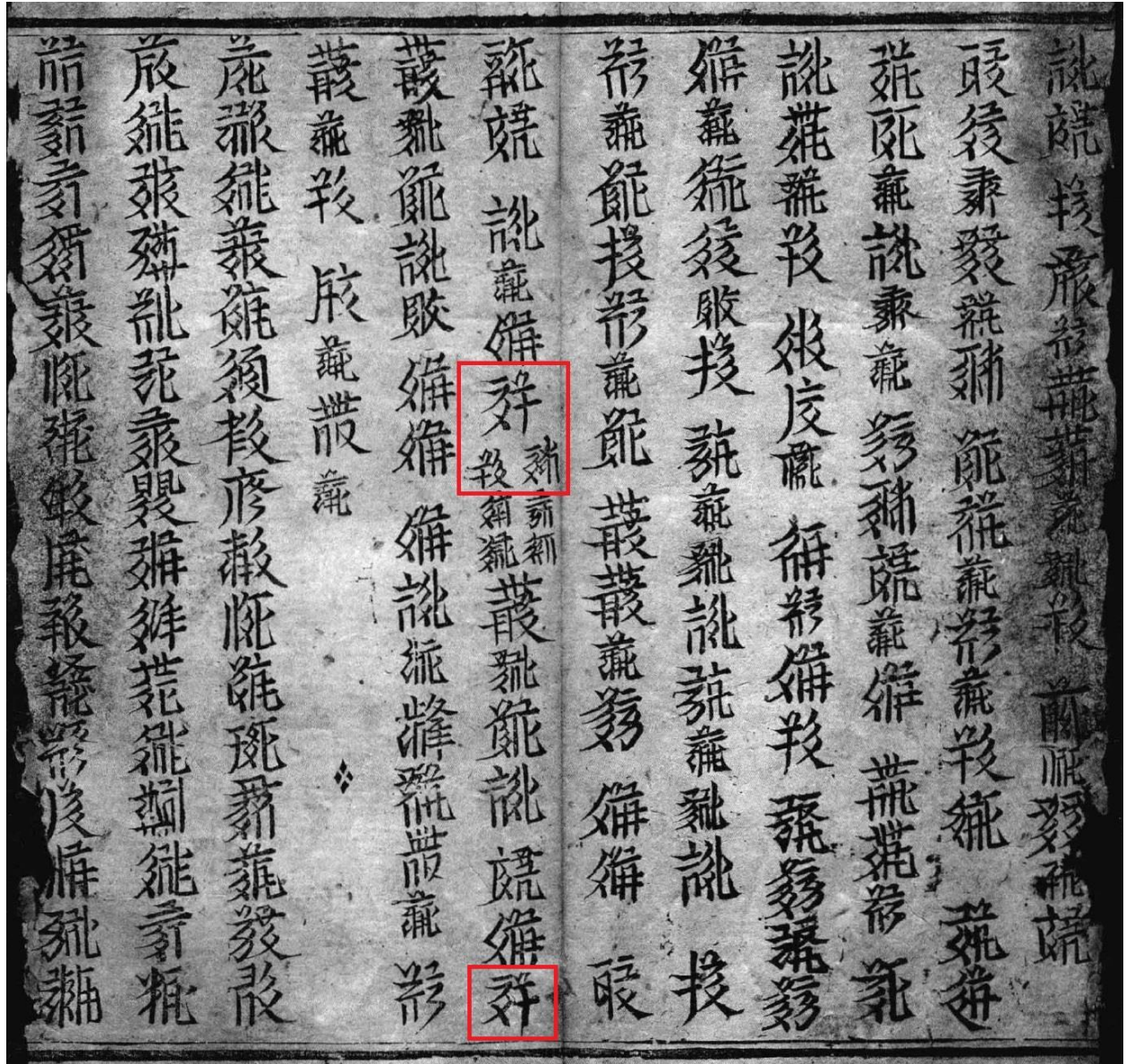


Fig. 13: National Library of China B11-028: 《繆敗傷癰痲痲痲》 𐰇𐰈𐰇𐰈⁷

The character 𐰇 is a transcription character for Sanskrit used in a *dhāraṇī* incantation, and is glossed graphically and phonetically with the two characters 𐰇 *sjij*¹ and 𐰈 *ja*² which give the *fanqie* reading of *sjja*². In this *dhāraṇī* the Tangut words 𐰇𐰇𐰇𐰇 𐰇𐰇𐰇𐰇 𐰇𐰇𐰇𐰇 𐰇𐰇𐰇𐰇 *ja mu kja nja: mja sja* and 𐰇𐰇𐰇𐰇 𐰇𐰇𐰇𐰇 𐰇𐰇𐰇𐰇 𐰇𐰇𐰇𐰇 *dja r(jir) 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).

⁷ Zhōngguó cáng Xìxià Wénxiàn vol. 3 p. 324; there is also another copy of the same edition in vol. 3 p. 285.

[illegible]

The Tangut text is a literal translation of the Chinese version of the *Mahāvaiṣṭava Sūtra* (T0279 《大方廣佛華嚴經》卷第四十一), but the character 𑖦 does not correspond to any character in the Chinese version, so its meaning is uncertain.

菩薩摩訶薩亦復如是，不捨於行，隨世所作，而於此二無所執著。

⁸ Cf. *Fǎcáng Dūnhuáng Xīxiàwén Wénxiàn* p. 215. The corresponding section of the Beijing copy (*Zhōngguó cáng Xīxià Wénxiàn* vol. 8 p. 325) is missing this part of the text due to damage.

The character 𪗇 is perhaps a mistake for U+18284 𪗇 *niow*² ‘evil’ (seen seven characters down on the same column), with the left and right sides swapped.

JTC1/SC2/WG2 N5217



Fig. 16: National Library of China B11-052: 《維摩經疏》 colophon ¹⁰

¹⁰ Zhōngguó cáng Xīxià 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. 18).

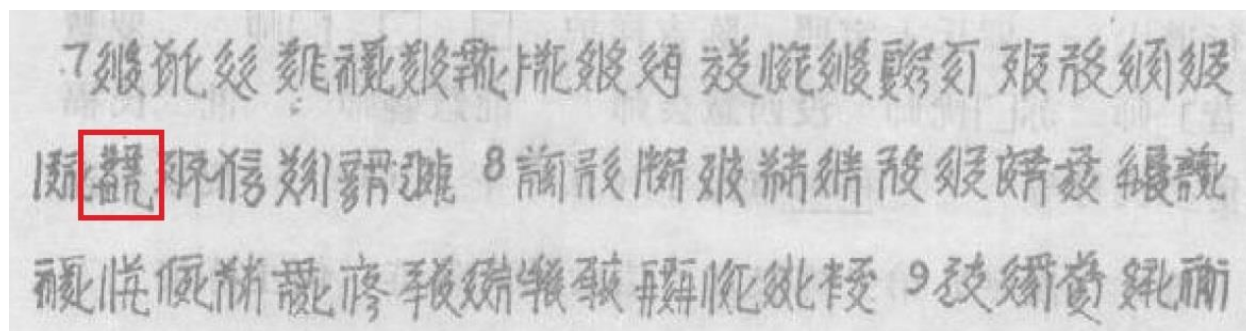


Fig. 17: *Xixia Fojiao Shilüe* p. 316

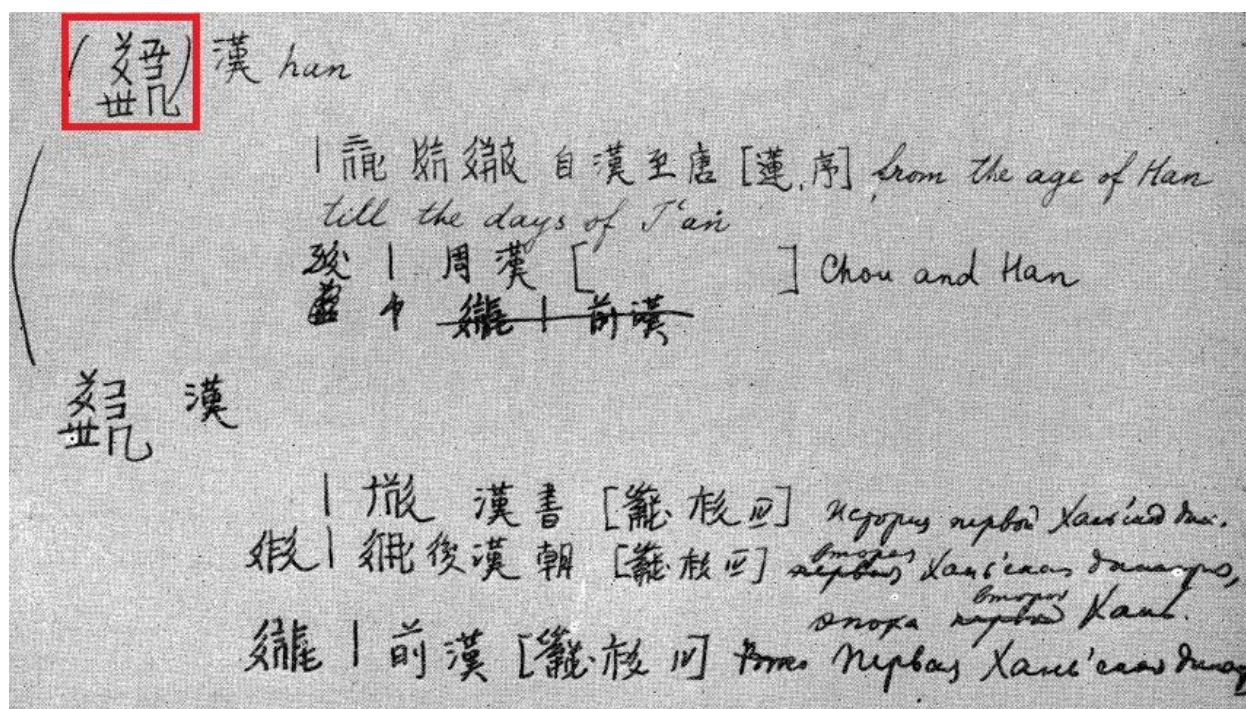


Fig. 18: *Tangutskaya filologiya* vol. II p. 655



Fig. 19: National Library of China B11-052: 《維摩經疏》 colophon ¹¹

¹¹ Zhōngguó cáng Xīxià 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 𐰽 in the phrase 𐰽𐰺𐰾𐰏 which Shǐ Jīnbō translates as 令依蕃譯 ‘caused to be translated according to Tangut [language]’. In this case, 𐰽 should be a variant form or mistake for U+1869C 𐰽 *lhjwjj*¹ ‘Tangut’.

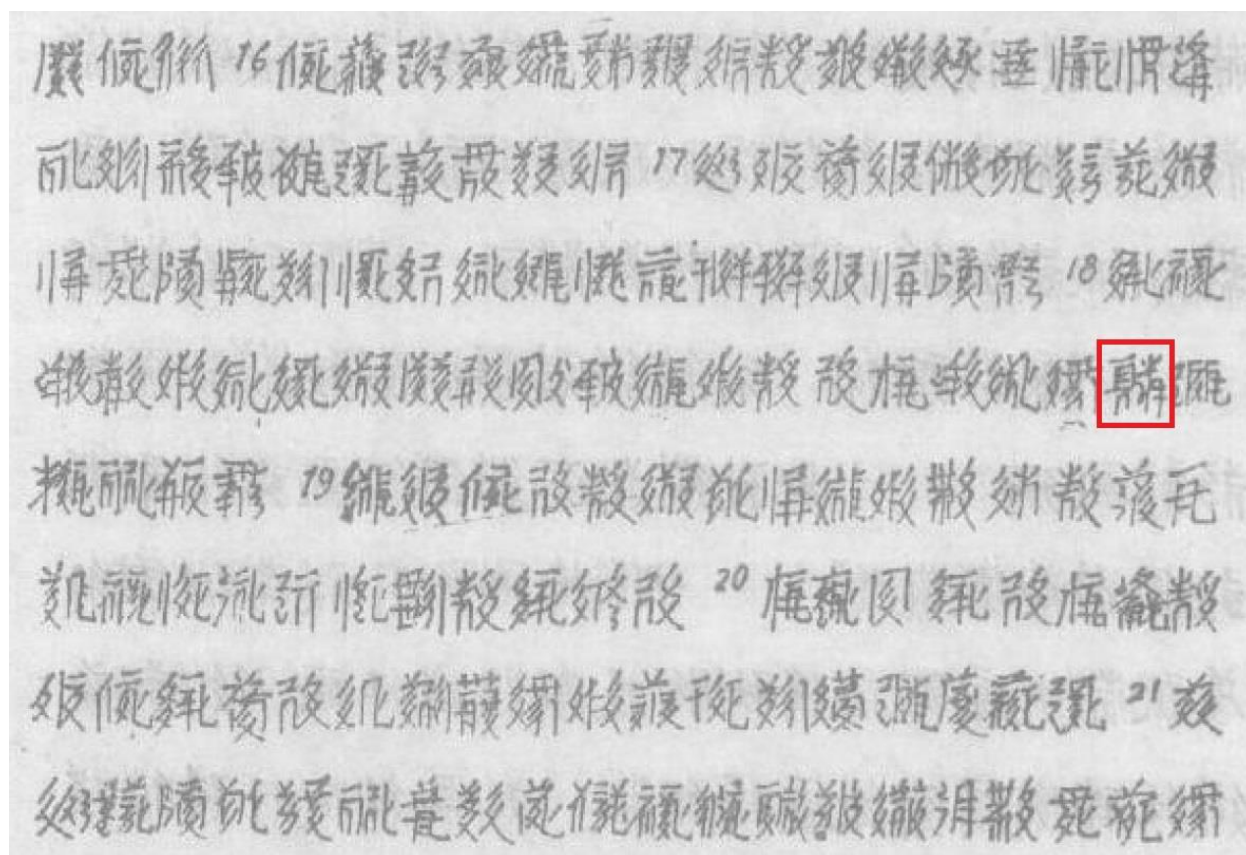


Fig. 20: *Xixia Fojiao Shilüe* p. 317

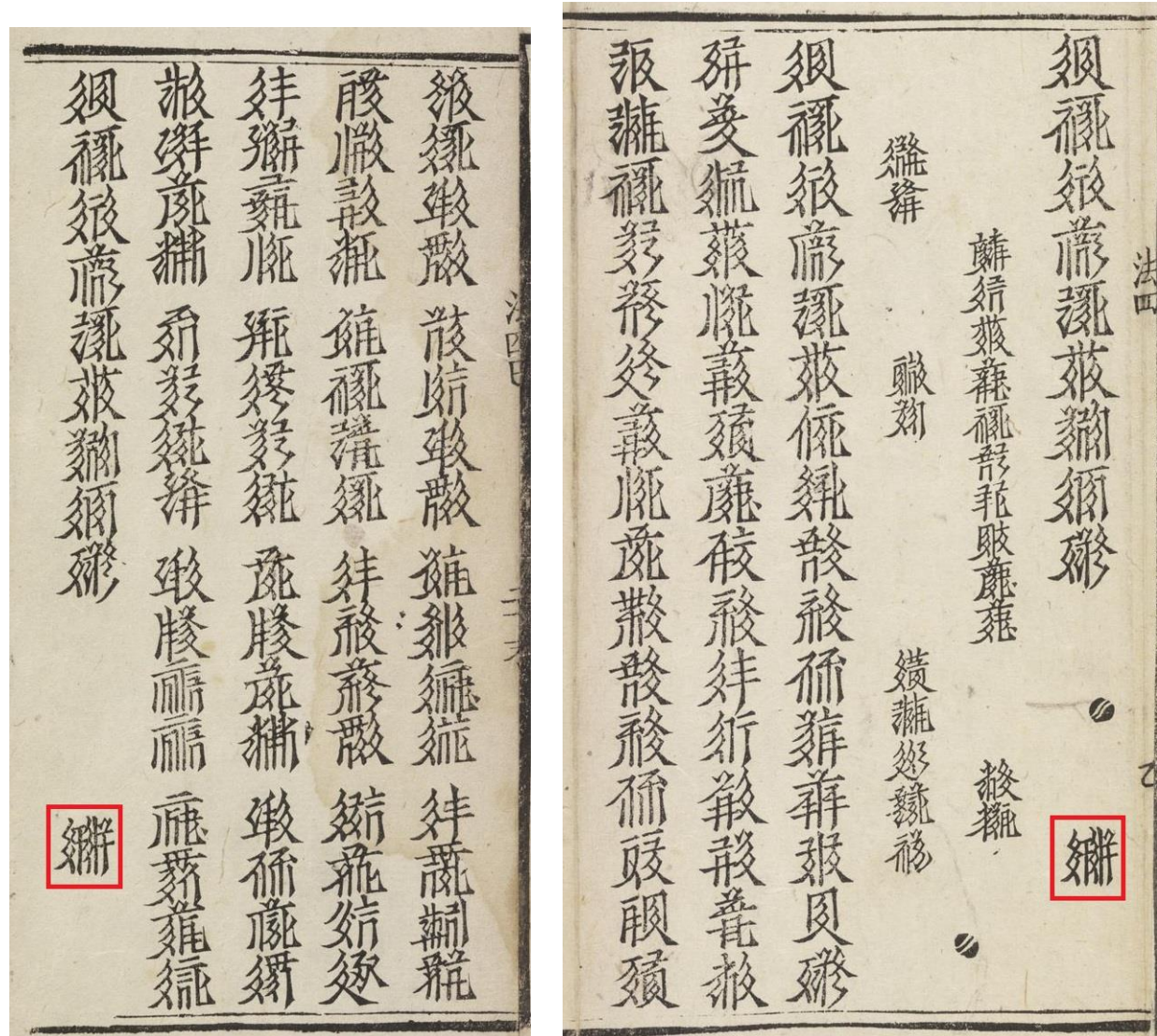


Fig. 21: Princeton BQ2053.T3 T38: 《娑祇祇祇祇祇》¹²

The character 娑 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 娑 *lhjwij*¹ 'title' with the components rearranged (Arakawa Shintarō transcribes it as 娑), 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.

¹² Tangut Version of the Lotus Sutra in the Collection of Princeton University Library, Facsimile, Text and Linguistic Studies pp. V and LIII.

4.12 Homonyms

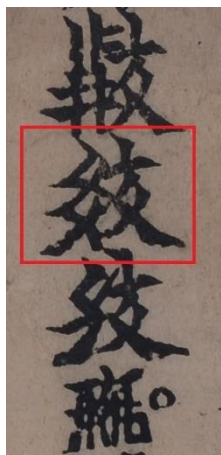


Fig. 22: 《玦玦玦玦玦玦玦玦》 folio 13A

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

4.13 Homonyms

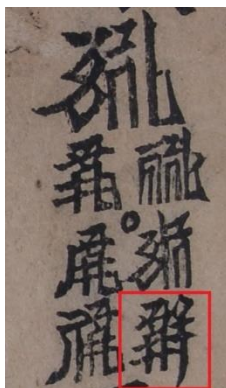


Fig. 23: 《𪛗𪛗𪛗𪛗𪛗𪛗𪛗𪛗》 folio 8A

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

4.14 Homonyms

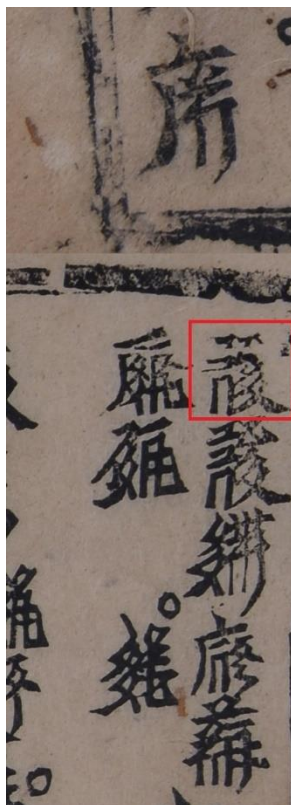


Fig. 24: 《𪛗𪛗𪛗𪛗𪛗𪛗𪛗𪛗》 folio 12B-13A

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

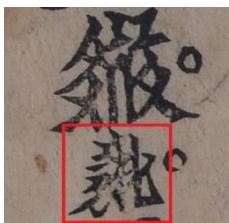


Fig. 25: 《𪚩𪚪𪚫𪚬𪚭𪚮𪚯𪚰》 folio 9B

The character 𪛗 occurs in *Homonyms* as the only character under the homophone group 𪛗 *niq*² ‘goose’. The middle component is not clear, but by comparison with other characters in this edition of *Homonyms* it appears to be 𪛗. There are no other characters with the same or similar left and right components as this character which it could be a plausible variant form of.

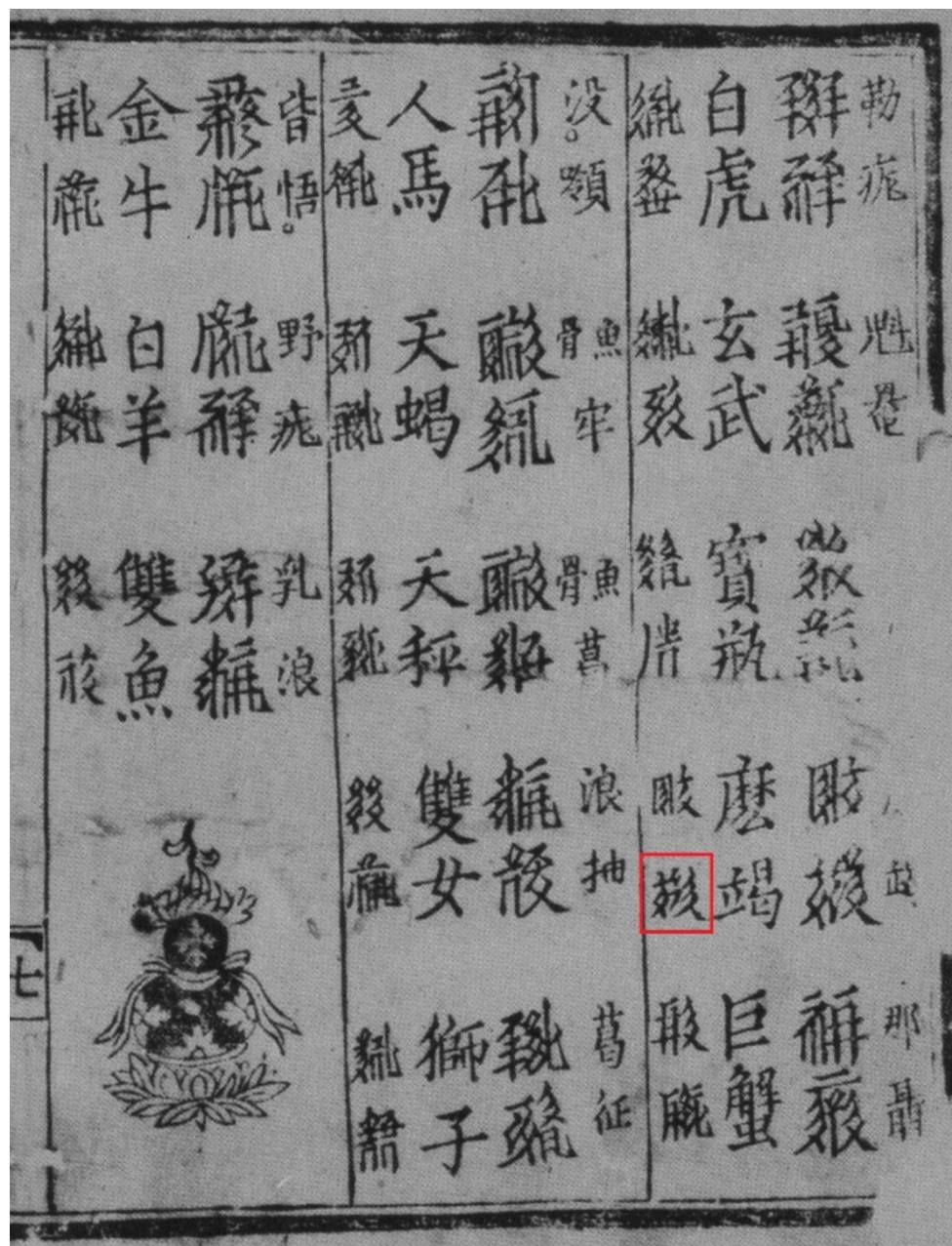


Fig. 26: IOM Tang.13: 《級殺形纖蕪玆綽絞》 Ed. A folio 7A

The character 蛟 occurs in the phonetic gloss 蛟蛟 mo^2 ($khja^2$) for Chinese 麼竭 $mó jié$ 'Capricorn'. It is likely to be a non-unifiable variant of U+18039 蛟 $khja^2$.

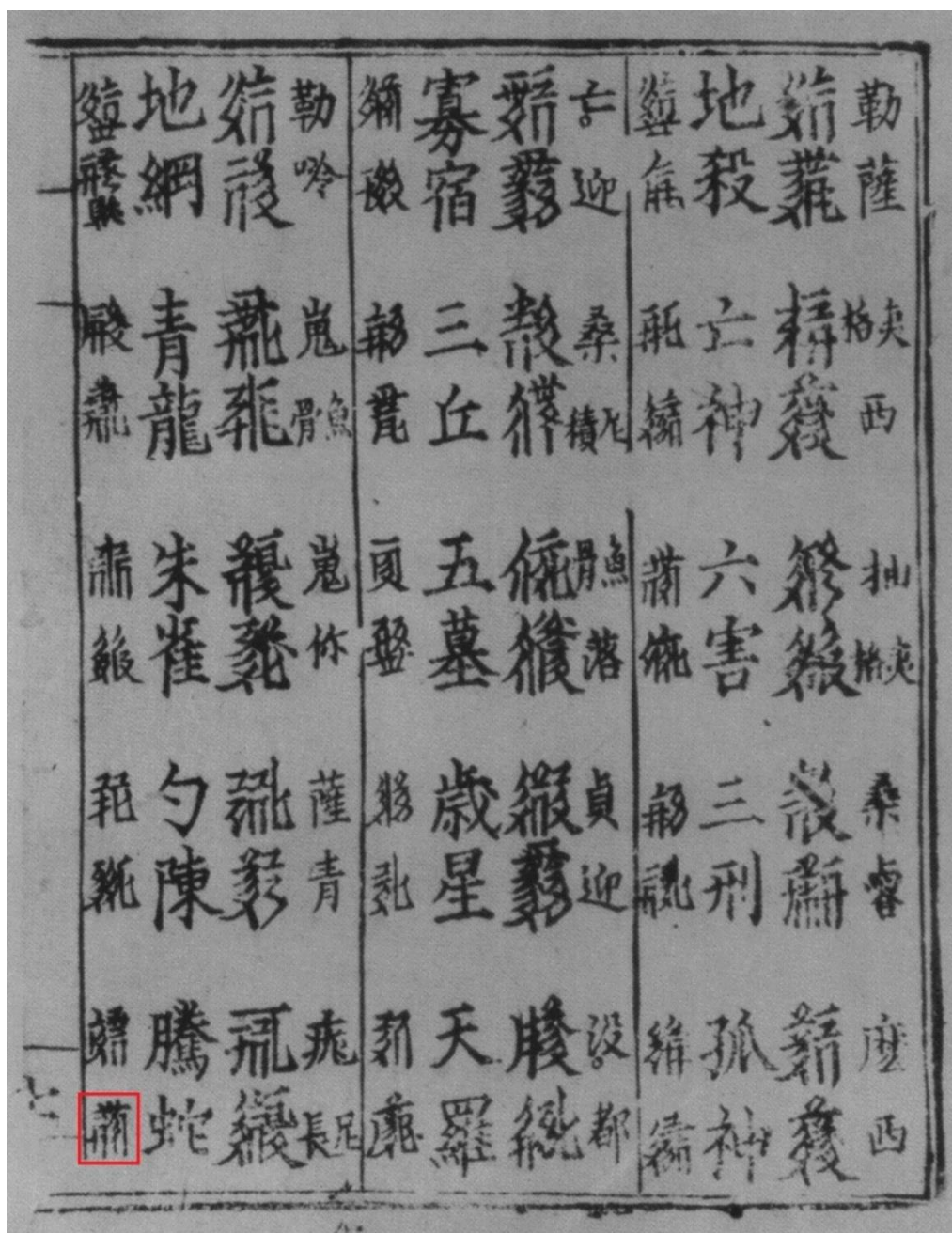


Fig. 27: IOM Tang.13: 《級殺𪚩𪚩𪚩𪚩𪚩》 Ed. B folio 7B ¹³

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

¹³ This entry also occurs in the A edition (folio 6B), but the gloss character is lost due to damage.

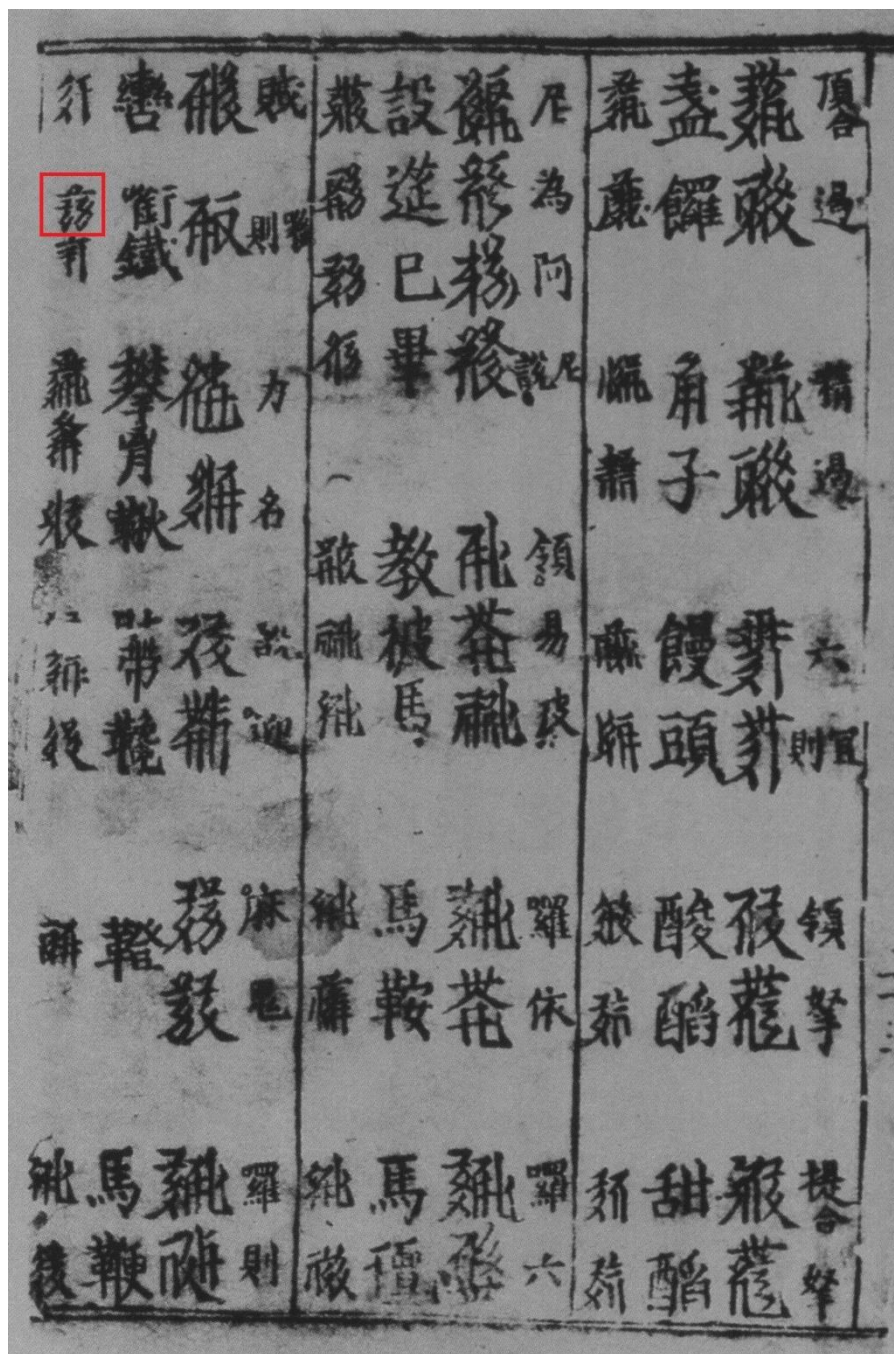


Fig. 28: IOM Tang.13: 《級殺形繼繼玆解絞》 Ed. A folio 33B

The character 𐰇 occurs in the phonetic gloss 𐰇𐰺 ($x\tilde{a}^1$) $thij^2$ for Chinese 嚼鐵 *xián tiě* 'horse's bit'. It is likely to be a non-unifiable variant of U+17C54 𐰇 $x\tilde{a}^1$. This character is also given in Nevsky's draft manuscript Tangut dictionary (Fig. 29). This is the only example I have ever encountered where a character with Component 185 𐰇 is written with Component 75 𐰇.



Fig. 30: IOM Tang.18: 《龍龍》 Ed. A folio 52B

The character 𠂔 occurs as the gloss for 𠂔 *ljwⁱ²* ‘colt’ in the ‘A’ edition of *Homophones*. In the ‘B’ edition the gloss character is U+1733B 𠂔 *jar²* ‘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.

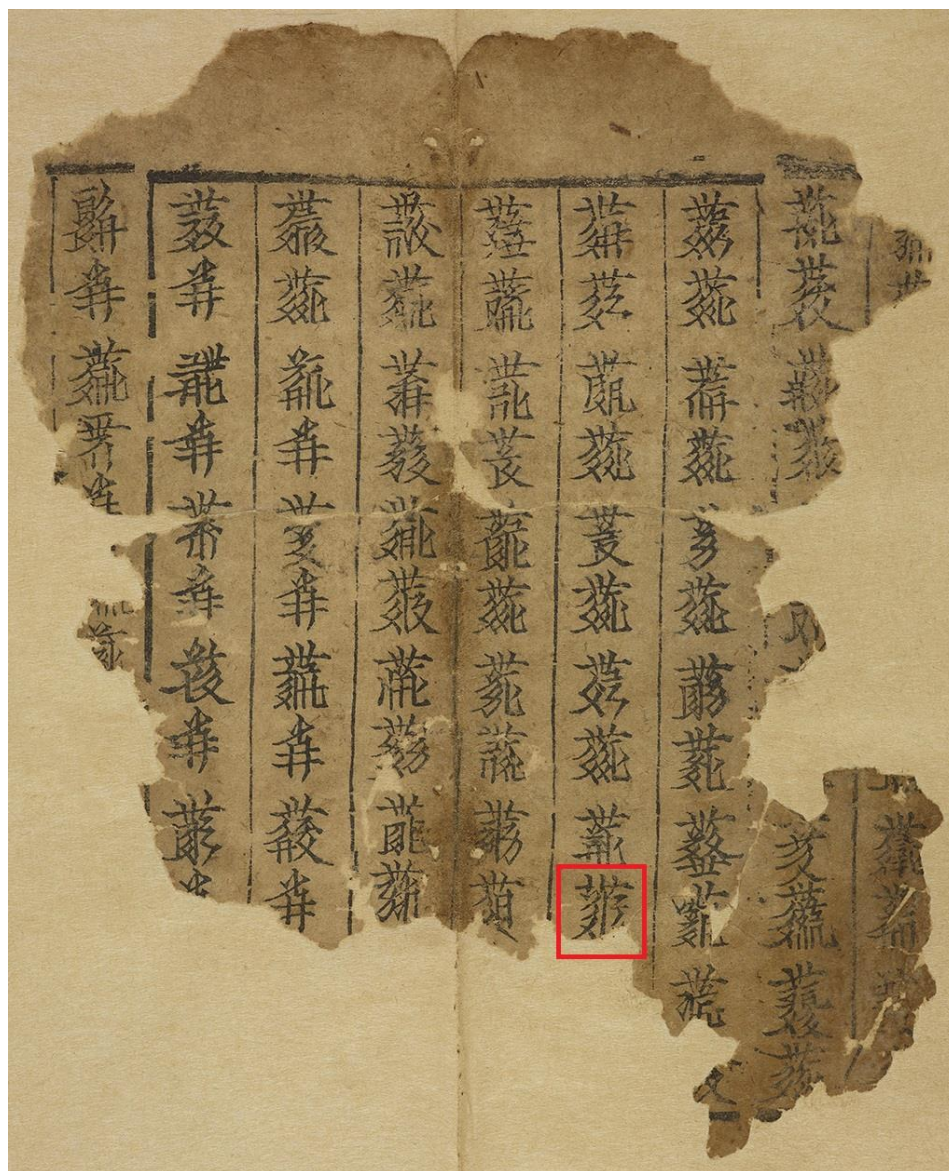


Fig. 31: British Library Or.12380/1843 (G): 《佩𦵏》

The character 薊 occurs in the word 薊薊 (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 薊薊 (Tang.19 folio 7A), so 薊 should be a variant form of U+17768 薊. However, it cannot be considered to be a unifiable variant because it uses a completely different bottom left component.



Fig. 32: British Library Or.12380/1843 (H): 《佩𪔐》

The character 𪔐 occurs in the word 𪔐𪔐 (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 𪔐𪔐 (Tang.19 folio 8A), so 𪔐 should be a variant form of U+17BC0 𪔐. However, it cannot be considered to be a unifiable variant because it uses a completely different middle component (Component 419 𪔐 instead of Component 112 𪔐).

4.22 Joined Rimes of the Five Sounds

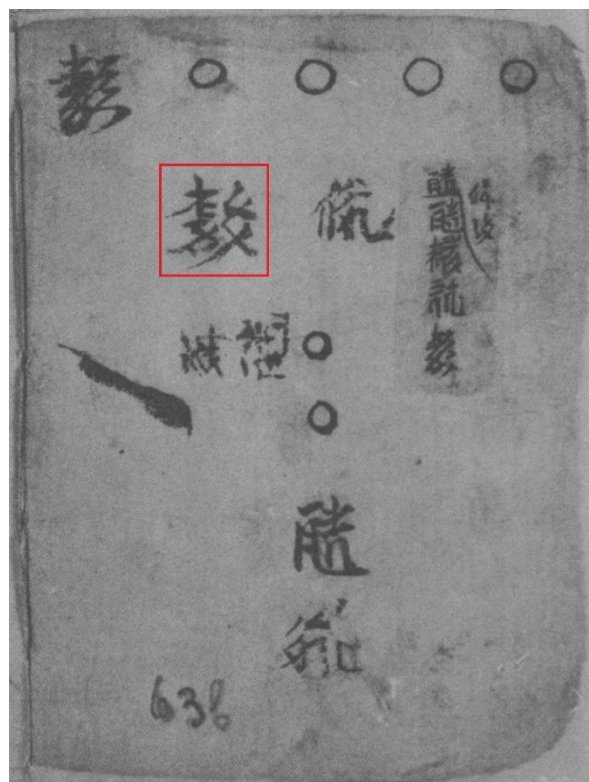


Fig. 33: IOM Tang.22: 《𪛗𪛗𪛗𪛗》 MS D folio 63B

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 𪛗 *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 𪛗 *zew*² (see Fig. 34), which should be under Rime Table 48, but none of the manuscripts give 𪛗 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. 34: 'Wǔyīn qièyùn' yǔ 'Wénhǎi bǎoyùn' Bǐjiào Yánjiū p. 183

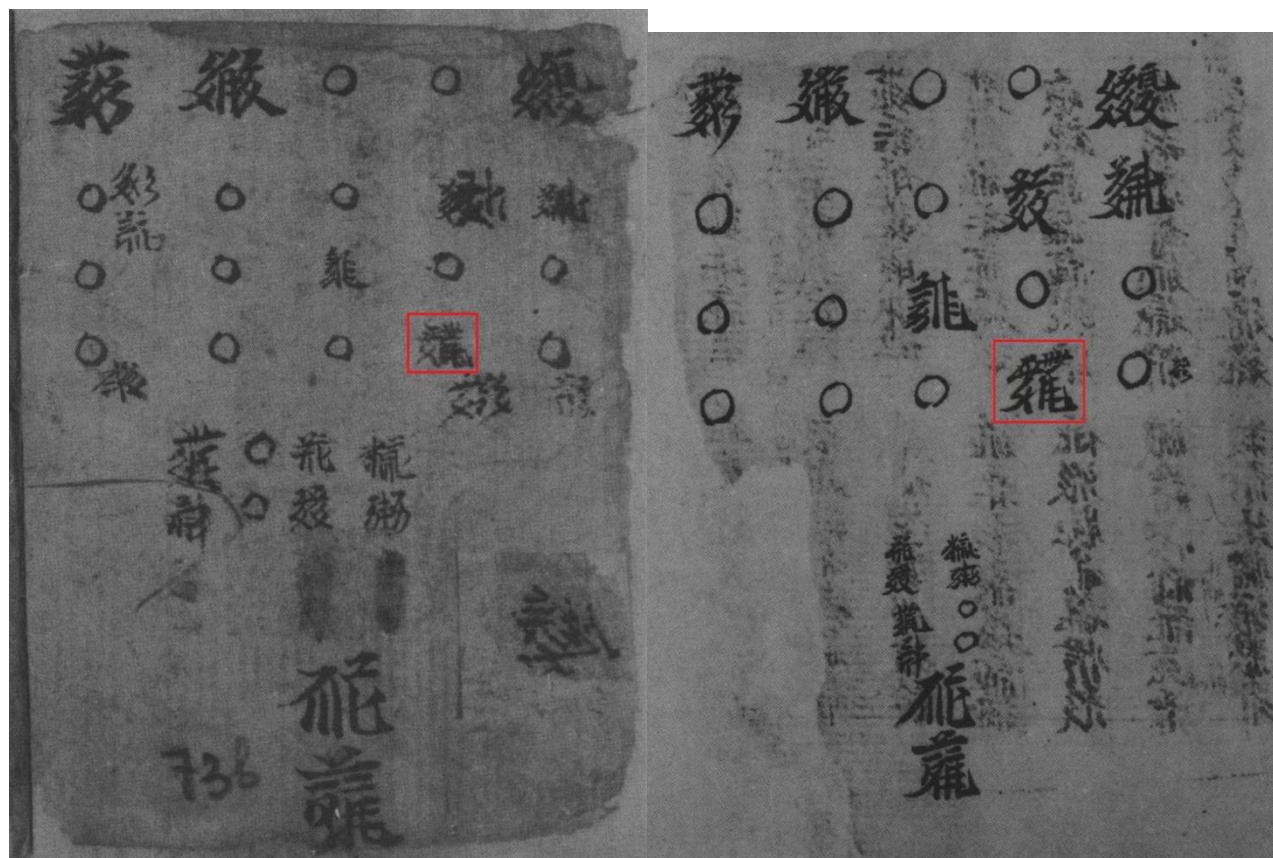


Fig. 35: IOM Tang.22: 《侏龍被龍》 MS D folio 73B and MS E page 43

The 'C', 'D', and 'E' manuscripts of the Tangut rime tables each have a somewhat different character in the fourth position of the second column (for dental initials) of Rime Table 70 侏 *bji*¹ / 龍 *bji*². MS 'C' appears to give U+18688 𐰇 *tju*²; whereas MS 'D' gives 𐰇, and MS 'E' gives 𐰇. The D and E forms are unencoded, but are obviously variants of the same character with a one-stroke difference in the top right component. As the 'D' form could plausibly be constructed from the left side of 𐰇 *tju*² and the top and right side of 𐰇 *ljw*¹ (shown under the liquids column in this rime table), giving the plausible *fanqie* reading *tjw*¹, I believe that 𐰇 is the correct glyph form, and 𐰇 can be treated as a unifiable variant.

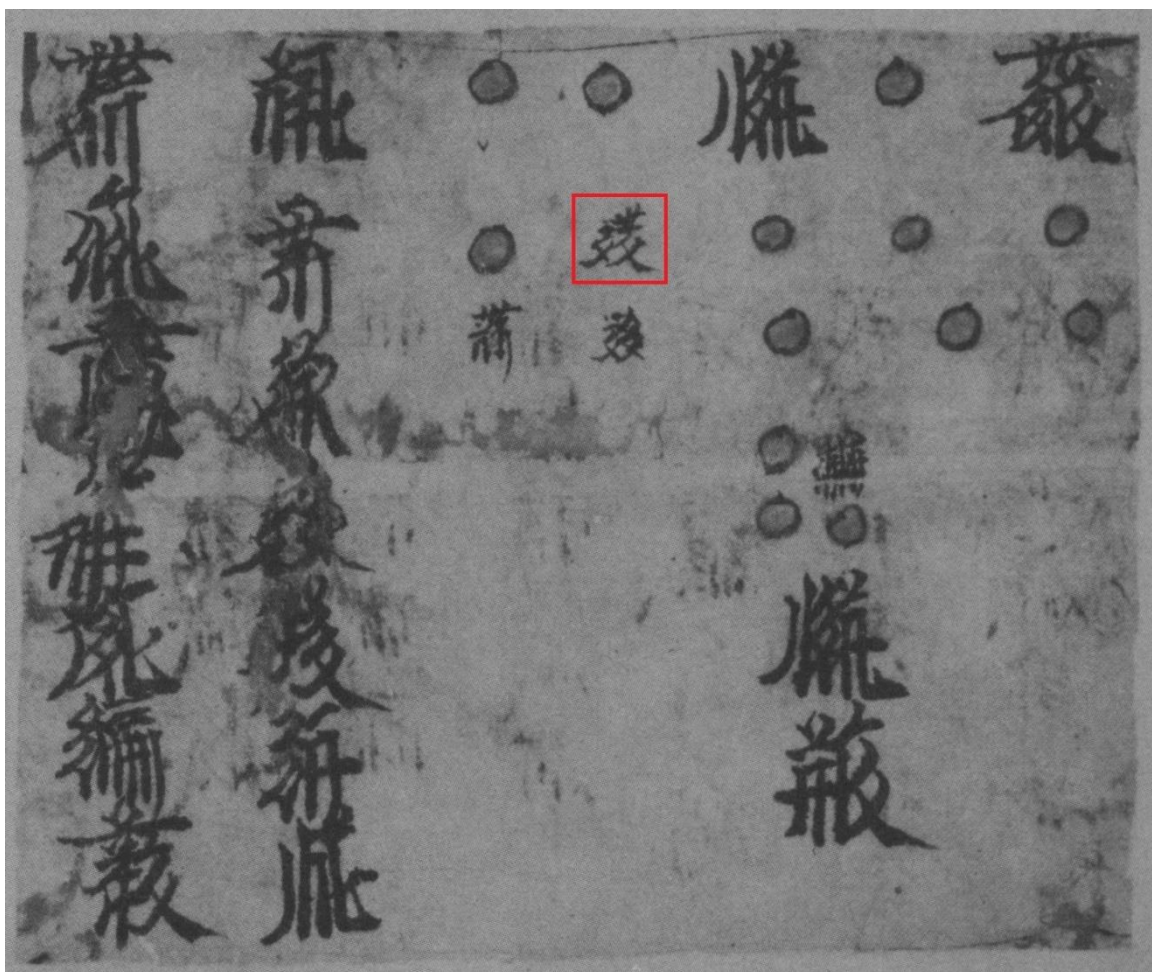


Fig. 36: IOM Tang.22: 《𐰽𐰚𐰽𐰚𐰽𐰚𐰽𐰚》 MS F page 25

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

4.25 Joined Rimes of the Five Sounds

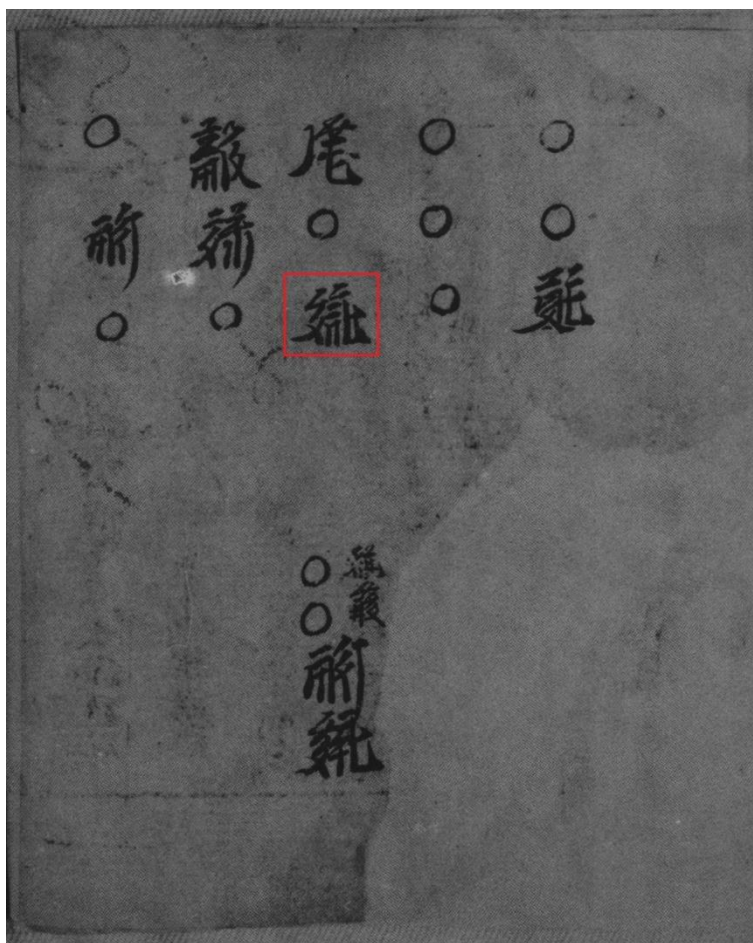


Fig. 37: IOM Tang.22: 《儀龍被龍》 MS E page 52

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

4.26 Joined Rimes of the Five Sounds

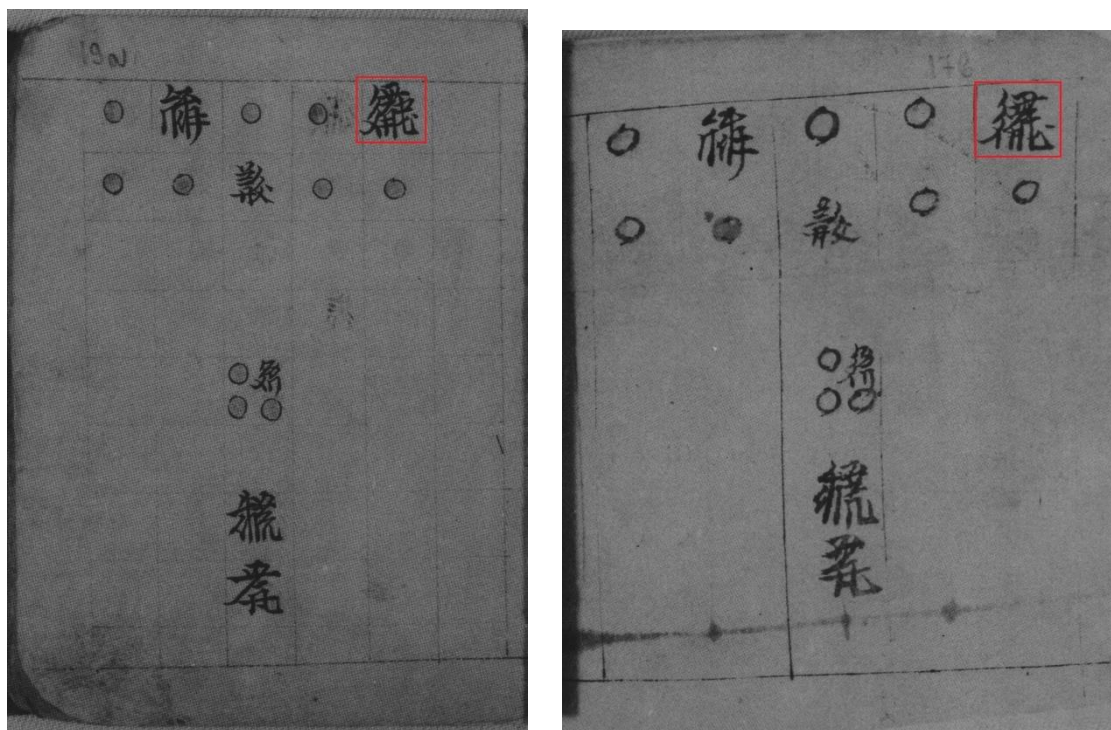


Fig. 38: IOM Tang.22: 《𐰚𐰍𐰏𐰐𐰑》 MS A folio 19A and MS B folio 17B

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

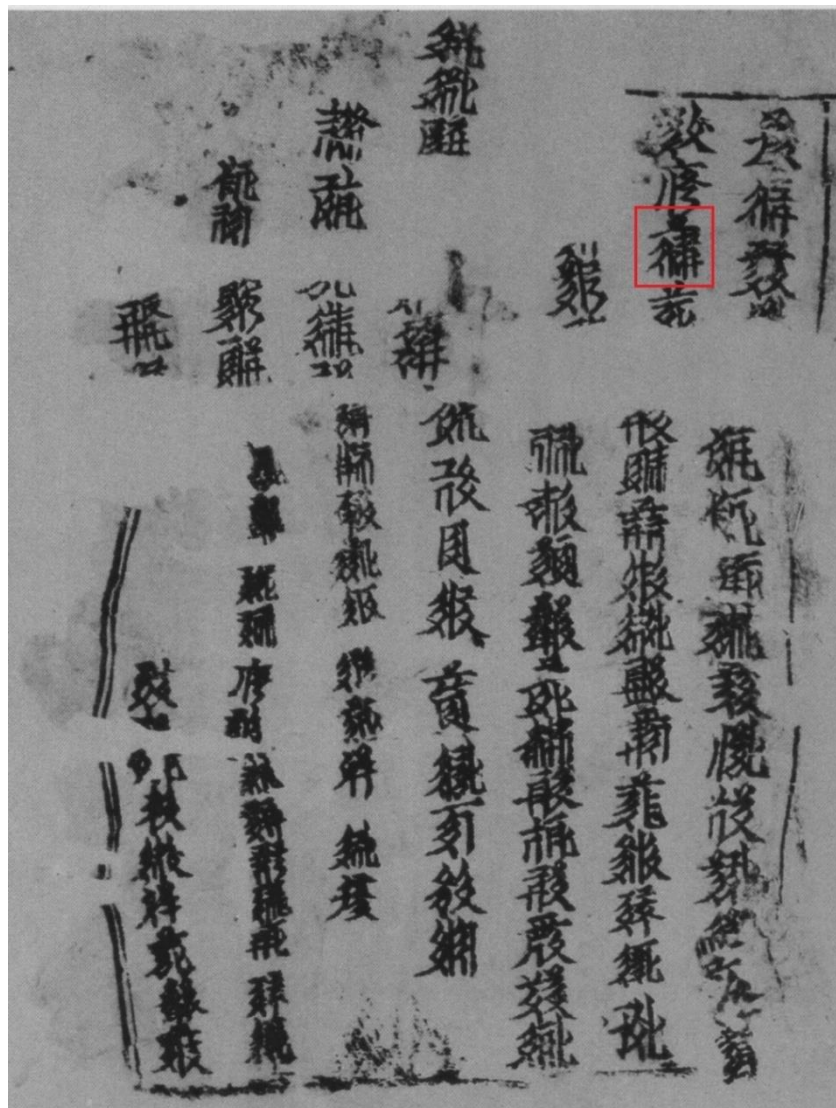


Fig. 39: IOM Tang. 35/1: 《新集寶雙》 folio 31B ¹⁴

The character 禱 occurs in the phrase 效禱禱苑 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 效禱苑, but does not assign any meaning to 禱.¹⁵ Niè Hóngyīn transcribes the character as U+174A9 禱 *nju*¹, but he omits the phrase with this character from his translation.¹⁶ 禱 is a rare character which only occurs in the family name 禱効 *nju*¹ *rjijr*², and does make any sense here, so it is not convincing that 禱 is a variant form of 禱.

¹⁴ *Vnov' sobrannye dragotsennye parnye izrechenija* p. 212.

¹⁵ *Vnov' sobrannye dragotsennye parnye izrechenija* p. 125.

¹⁶ “Xīxià yíwén lù” p. 158.

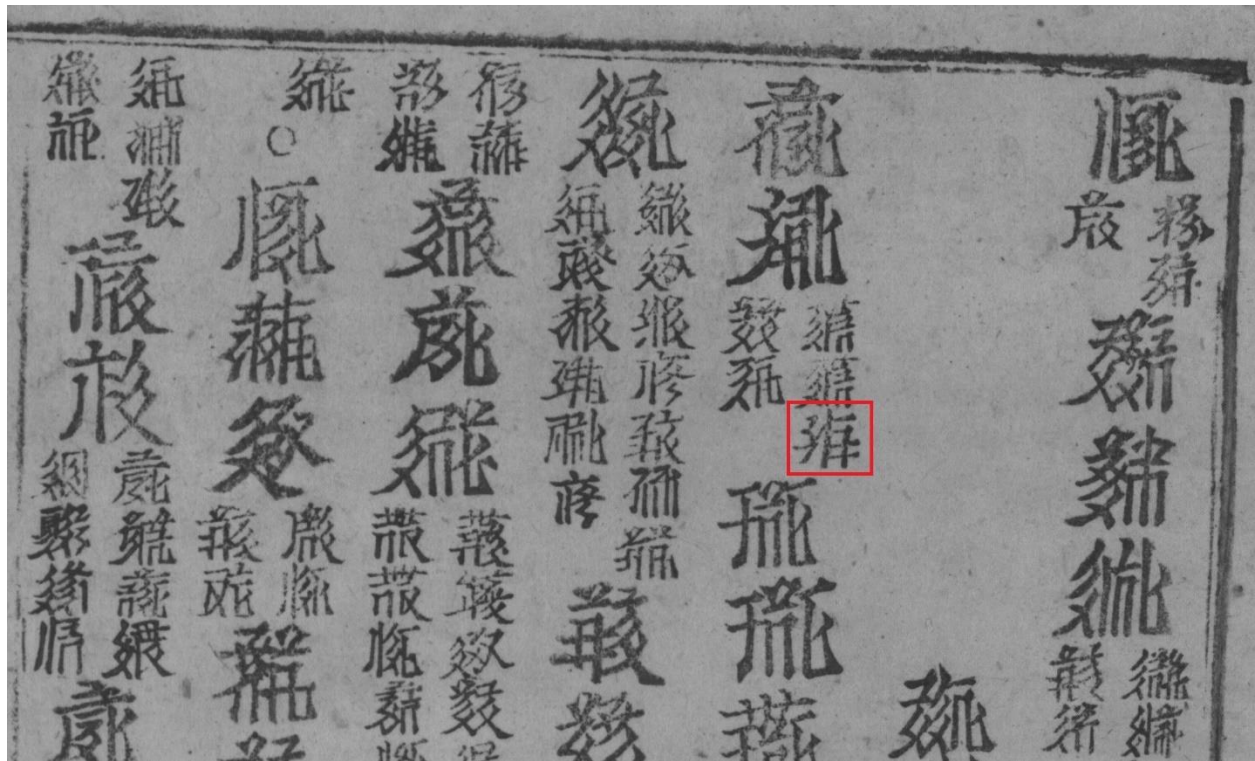


Fig. 40: IOM Tang.27: 《𠂔𠂔𠂔𠂔𠂔𠂔𠂔𠂔》 folio 9A

The character 𠂔 occurs in the phrase 𠂔𠂔𠂔𠂔𠂔, 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.

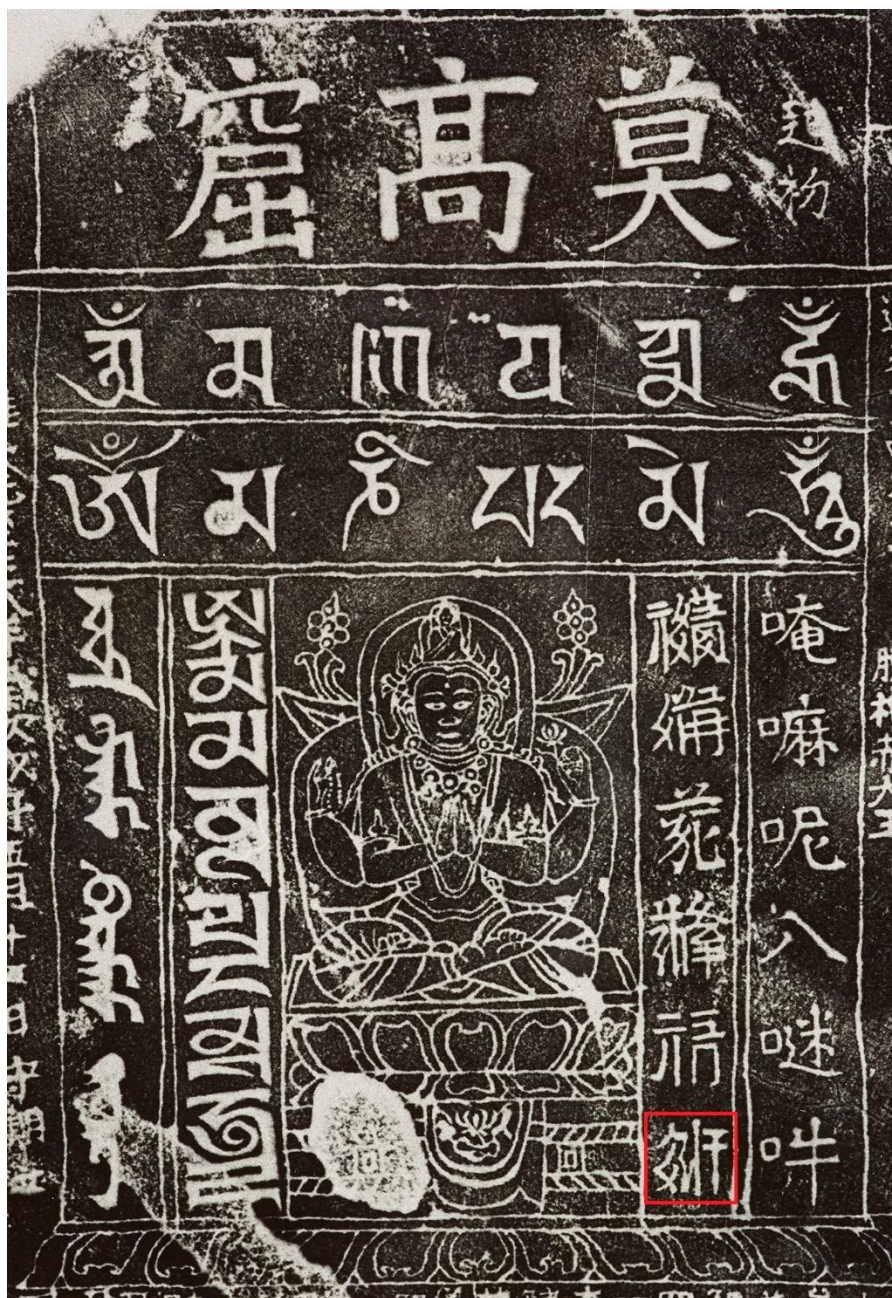


Fig. 41: Detail of the Stele of Sulaiman

The *Stele of Sulaiman* erected at the Mogao caves near Dunhuang in 1348 is engraved with the Buddhist mantra *Oṃ maṇi padme hūṃ* in six scripts, including Tangut. The Tangut version of this mantra should be written 𐰇𐰏𐰢𐰆𐰣𐰚𐰢𐰏𐰤 ·a mja¹ nji² pja¹ mjij¹ xo, but on the stele the last character is written with the non-unifiable variant form 𐰤.

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. Kumajskaja klassika v tangutskom perevode* (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. Kychanov] 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. Tangutskaya filologiya. Issledovaniya i slovarʹ* [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 韩小忙. Beijing: Zhongguo shehui kexue chubanshe, 2008.
- Vnov’ sobrannye dragotsennye parnye izrechenija. Vnov’ sobrannye dragotsennye parnye parnye izrechenija* [The Newly-Assembled Precious Dual Maxims]. E. I. Kychanov (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 韩小忙 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 Jiànlù 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.

7. Proposal Summary Form

SO/IEC JTC 1/SC 2/WG 2 PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646¹⁷

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:	Proposal to encode 2 Tangut components and 29 Tangut ideographs
2. Requester's name:	Andrew West
3. Requester type (Member body/Liaison/Individual contribution):	Individual contribution
4. Submission date:	2023-05-31
5. Requester's reference (if applicable):	
6. Choose one of the following:	
This is a complete proposal:	YES
(or) More information will be provided later:	

B. Technical – General

1. Choose one of the following:		
a. This proposal is for a new script (set of characters):	NO	
Proposed name of script:		
b. The proposal is for addition of character(s) to an existing block:	YES	
Name of the existing block:	Tangut Supplement	
2. Number of characters in proposal:	31	
3. Proposed category (select one from below - see section 2.2 of P&P document):		
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	X	G-Obscure or questionable usage symbols
4. Is a repertoire including character names provided?	YES	
a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?	YES	
b. Are the character shapes attached in a legible form suitable for review?	YES	
5. Fonts related:		
a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?	Andrew West	
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):	Andrew West	
6. References:		
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?	YES	
b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?	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)?	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.

¹⁷ 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?	NO
If YES explain	
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.)?	NO
If YES, with whom?	
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?	NO
Reference:	
4. The context of use for the proposed characters (type of use; common or rare)	rare
Reference:	
5. Are the proposed characters in current use by the user community?	YES
If YES, where? Reference:	
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?	NO
If YES, is a rationale provided?	
If YES, reference:	
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?	NO
If YES, is a rationale for its inclusion provided?	
If YES, reference:	
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?	NO
If YES, is a rationale for its inclusion provided?	
If YES, reference:	
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?	NO
If YES, is a rationale for its inclusion provided?	
If YES, reference:	
11. Does the proposal include use of combining characters and/or use of composite sequences?	NO
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:	
12. Does the proposal contain characters with any special properties such as control function or similar semantics?	NO
If YES, describe in detail (include attachment if necessary)	
13. Does the proposal contain any Ideographic compatibility characters?	NO
If YES, are the equivalent corresponding unified ideographic characters identified?	
If YES, reference:	