

New origins to vowel tensing in Tangut: internal and comparative evidence¹

Shuya Zhang¹ and Yunfan Lai²

¹Shanghai Jiao Tong University, Shanghai, China and ²Nanyang Technological University, Singapore

Corresponding author: Lai Yunfan; khroskyabs@gmail.com

Abstract

This paper proposes new origins for tense vowels in Tangut by integrating textual analysis of Tangut texts with comparative data from both Gyalrongic and other Sino-Tibetan languages. It uncovers two previously unreported sources of vowel tensing in compounding: the collective prefix (*S-) and the compound linker (*-S-). Both morphemes left only a few traces, indicating their antiquity and productivity in earlier stages. The collective *S- could be an inherited morpheme which finds parallel in Tibetan, whereas the compound linker *-S- emerged as a stage of morphological merging in West Gyalrongic with (an) obscure origin(s). These findings not only advance our understanding of the origins of Tangut tense vowels but also offer insights into Sino-Tibetan nominal morphology.

Keywords: Tangut; Gyalrongic; Tense vowel; Collective; Compounding; Historical linguistics

1. The reconstruction of tense vowels in Tangut

Nishida (1964) first proposed that the Tangut rhymes in what is traditionally called the “first minor cycle” had a specific contrast in sound quality with other rhymes.² Based on Chinese and Tibetan transcriptions of Tangut, he posited that these rhymes could be reconstructed with “tense vowels (はり母音)”, as opposed to “lax vowels (ゆるみ母音)”. This view is supported by Wang (1982: 3–4), who suggested that the absence of *fanqie* spelling connections between the rhymes in the “first minor cycle” and other minor cycles could be attributed to their “laryngeal constriction (緊喉元音)”, equivalent to Nishida’s “tense vowel”. Wang (1982) further deemed that his proposal aligned well with relevant phenomena found in Lolo-Burmese languages. The reconstruction of tense vowels has received wide acceptance ever since, reused by some of the most influential Tangutologists (Arakawa 2012, 2014; Gong 1999; Li 1997).

Gong (1999) observed that a considerable number of word pairs in Tangut exhibit alternation between tense and lax vowels. For instance, the two members in the pair 𐰇¹⁴⁷⁵ *bji*¹ “to be thin” and 𐰇¹⁷⁸⁹ *bji*¹ “to make thin” are not only semantically related but also

¹ We would like to thank Nathan Hill for his constructive comments and Zhang Yongfu for generously sharing the annotated version of *The Twelve Kingdoms*. We also extend our gratitude to the two anonymous reviewers for their valuable suggestions.

² Kychanov and Sofronov (1963) initially proposed that Tangut rhymes are classified into four cycles, each containing a consecutive set of vowels with varying phonations or sound qualities. The first cycle, which comprises 58 rhymes and significantly outnumbers the others, is referred to as the “major cycle”. The remaining cycles are known as “minor cycles”.

distinguished solely by vowel tensing. This observation leads to the hypothesis that at least some tense vowels in Tangut result from morphological operations, which Gong (1999) identified as having four functions, summarized in Table 1. By analysing internal and external evidence, Gong (1999: 550) proposes that the phonological origin of vowel tensing in Tangut is an old sibilant pre-initial *s.Ci- (where Ci represents the initial consonant of a syllable).

Table 1. Functions of Tangut vowel tensing according to Gong (1999)

Function	Plain vowel (< *C _i -)	Tense vowel (< *s.C _i -)
Causativization	𐰇 ⁴⁶¹⁴ <i>nju</i> ² “to suckle (vt)”	𐰇 ⁴⁸³⁴ <i>nju</i> ² “to milk (vt)”
Causativization	𐰇 ⁴⁷⁶⁷ <i>ɣwej</i> ¹ “to fight (vi)”	𐰇 ²⁴⁸⁷ <i>ɣwej</i> ¹ “to cause to fight (vt)”
Denominalization	𐰇 ⁰¹⁸¹ <i>lə</i> ¹ “hole”	𐰇 ²⁶⁵⁸ <i>lə</i> ¹ “to bury”
Nominalization	𐰇 ³⁴⁶⁹ <i>sjj</i> ² “to know”	𐰇 ¹⁷⁷¹ <i>sjj</i> ² “wisdom”

Gong’s seminal work, although widely accepted, still leaves many instances of tense vowels in Tangut unexplained. For instance, Jacques (2014) demonstrates, through the examination of cognates shared between Tangut and modern Gyalrongic languages, that tense vowels in Tangut may also have originated from pre-initials that are intrinsic components of lexical roots (e.g. the numeral “ten”: Tangut 𐰇¹⁰⁸⁴ *ɣɑ*² :: Geshiza Horpa *zya* :: Japhug *sqi*). On the other hand, tense vowels involving morphological functions have different origins, as evidenced by the causative/denominalizing *S- and nominalizing *S-, which clearly come from distinct sources. This point highlights the eroded features of Tangut, wherein tense vowels represent a merger of different morphological functions and different pre-initial consonants.

Building upon the previous hypothesis that vowel tensing results from the transphonologization of pre-initial elements, this paper proposes new origins of Tangut tense vowels in terms of nominal morphology, specifically a collective prefix *S- and a compound linker *S-. It integrates the internal evidence from the study of Tangut texts³ with comparative data from modern Sino-Tibetan languages, particularly from modern West Gyalrongic languages.⁴

The paper is structured as follows. Section 2 outlines the reconstruction conventions employed in this paper. Sections 3 and 4 present Tangut internal evidence supporting the existence of a collective prefix *S- and a compound linker *S-, respectively. Both morphemes left only a few traces in Tangut, pointing to once-productive morphological processes which are crucial for reconstructing regular morphological processes in historical linguistics (Meillet 1925). Drawing on comparative data from Sino-Tibetan languages,

³ The Tangut examples used in this paper are mainly extracted from (i) *The Grove of Classification* (Shi et al. (eds) 1993), (ii) *Newly Collected Biographies of Affection and Filial Piety* (henceforth *Filial Piety*) (Jacques 2007, ed), (iii) *The Ode on Monthly Pleasures* (Nishida 1986, ed.), (iv) *The Twelve Kingdoms* (Solonin 1995, ed, with annotation provided by Zhang Yongfu), (v) *Tiansheng Code*, taken from Jacques (2012), (vi) *Mengzi* (Peng 2012, ed). Glosses follow the Leipzig glossing rules (Comrie et al. 2008), to which the following are added: [A] = stem A, [B] = stem B, [I] = stem I, [II] = stem II, COL = collective, DIR = directional prefixes, LNK = linker, ms = male speaker, PART = particle, POST = postposition, PN = person name, ws = female speaker.

⁴ This paper adopts the recent classification of Tangut as a West Gyalrongic, a sub-group within Qiangic (Lai et al. 2020). The comparative data used in this paper include both modern West Gyalrongic languages (Geshiza Horpa, Honkasalo 2019; Bawang Horpa, Yang 2021; Njorogs Khroskyabs, Yin 2007; Siyuewu Khroskyabs, Lai Yunfan’s field data; Wobzi Khroskyabs, Lai 2017) and East Gyalrongic languages (Japhug, Jacques 2015; 2021; Tshobdun, Sun 1997, 2006; Cogtse Situ, Lin 2016, Lin You-Jing’s field data; Bragbar Situ, Zhang 2020).

Section 5 demonstrates that the collective *S- may represent an ancient inherited morphological process, while the compound linker *-S- emerged as a stage of morphological merging in West Gyalrongic. These findings not only contribute to the reconstruction of Tangut morphology but also shed light on the origins of compounding morphology in Sino-Tibetan, an area that remains understudied.

2. Conventions of Tangut and Pre-Tangut reconstructions

Since both Tangut and Pre-Tangut forms are used in the present paper, it is necessary to elucidate the conventions adopted for these reconstructions. Tangut forms are provided with IPA transcriptions based on Gong's (2003) reconstruction and are referenced by their corresponding numbers in the *Tangut-Chinese Dictionary* (Li 1997).

Pre-Tangut forms generally follow the reconstruction by Jacques (2014) and are preceded by an asterisk *. Uncertain phonetic values are indicated with square brackets [], following Baxter and Sagart (2014). Periods, as in *Cə.C_i-, indicate a non-morphological separation between a pre-initial element and an initial. Hyphens after a pre-initial element indicate that it is considered a prefix.

The treatment of initial lenition in Pre-Tangut follows Lai (2023; 2024), who proposes that the occurrence of lenition depends on the syllabicity of the pre-initial element (including pre-initial consonants and pre-syllables) and distinguishes four types of Pre-Tangut pre-initial elements.

First, non-syllabic pre-initial consonants, noted as *CC_i- (without the period to differentiate them from the *S.C_i- consistently used in this paper), yield long vowels in Tangut. Second, syllabic pre-initials *Că.C_i- yield various phonation types, including tensing, rhoticizing and labial medializing, and result in initial lenition. Third, syllabic pre-initials *Cũ.C_i- yield the same phonation types but without initial lenition. Fourth, pre-initials with a neutral vowel Cə.C_i- cause lenition before dropping entirely. Note that the vowel distinction in the reconstructed presyllables represent pure phonological contrasts rather than actual phonetic values. See (1) below for distinctive examples of these four pre-initial types in Tangut.

- (1) a. Tangut *CC_i- (lengthening)
*Ckar¹ > 𑖇⁵⁶⁸² *kaar*¹ “to weigh”
- b. Tangut *Că.C_i- (leniting)
*Să.kia¹ > *Să.yia¹ > *S.yia¹ > 𑖇⁰⁴³⁹ *yie*¹ “to cook”
- c. Tangut *Cũ.C_i- (non-leniting)
*Sũ.kjo¹ > *S.kjo¹ > 𑖇²²⁷⁸ *kju*¹ “spring onion”
- d. Tangut *Cə.C_i- (leniting and dropping)
*Cə-ko¹ > *Cə-yo¹ > *yo¹ > 𑖇²⁷⁵⁰ *yu*¹ “head”

Since the second type *Că.C_i- (in 1b) and the third type *Cũ.C_i- (in 1c) both transphonologize into tense vowels, they can be reconstructed as *S[ũ/ă].C_i-. For the sake of brevity, the distinction between *-ă and *-ũ in the vowel tensing pre-initial elements will be mentioned only when necessary. In most cases, we unify our notation with a simple *S.C_i- when referring to the vowel tensing pre-initials in Pre-Tangut, corresponding to the third stage in (1b) and the second stage in (1c), while keeping *Că.C_i- as the pre-syllable disappeared in Tangut.

3. Collective prefix *S-

In Tangut, some compounds have initial syllables with a tense vowel, which may originate from a compound initial element *S-. This element undergoes transphonologization, resulting in a tense vowel in the subsequent syllable, as illustrated in (2).

(2) *S-CV-CV > CV-CV

In some cases, it is expected that the compound initial *S- is not an intrinsic part of the root but rather a prefix used to derive collective nouns. The four compounds listed in Table 2 are among the few remaining traces of this collective morphology in Tangut. At the synchronic level, the *S- collective prefix has become lexicalized as an inseparable component.

Table 2. Traces of compound initial *S- in Tangut

Gloss	Attested forms	Pre-Tangut	Component 1	Component 2
“father and son”	𐞗 ⁵⁰⁷⁰ 𐞗 ⁵⁰⁴⁹ zji ¹ -wja ¹	*S-zji ¹ -wjaC ¹	𐞗 ¹⁰⁸⁵ zji ¹ “offspring, male”	𐞗 ⁵⁰⁴⁹ wja ¹ “father”
“mother and son”	𐞗 ⁵⁰⁷⁰ 𐞗 ⁰⁰⁹² zji ¹ -mja ¹	*S-zji ¹ -mjaC ¹	𐞗 ¹⁰⁸⁵ zji ¹ “offspring, male”	𐞗 ⁰⁰⁹² mja ¹ “mother”
“husband and wife”	𐞗 ⁵⁷²¹ 𐞗 ³⁵⁶² zji ¹ -mjaa ¹	*S-zji ¹ -mjaaC ¹	𐞗 ¹⁰⁸⁵ zji ¹ “offspring, male”	𐞗 ³⁵⁶² mjaa ¹ “son-in-law”
“children, baby”	𐞗 ⁵⁵²⁵ 𐞗 ¹²⁴¹ zji ¹ -lji ²	*S-zji ¹ -lji ²	𐞗 ¹⁰⁸⁵ zji ¹ “offspring, male”	𐞗 ¹²⁴¹ lji ² “DIM?”

Note: the symbol * represents reconstructed forms, † forms without unbound attestation, [] uncertain phonetic value. The lenited consonants z- and w- are retained in Pre-Tangut forms to aid in readability.

The first component of the four compounds listed in Table 2 shares the same phonological form, zji¹, which is etymologically related to 𐞗¹⁰⁸⁵ zji¹ (Pre-Tangut *zja) “son, offspring, male”. The form 𐞗¹⁰⁸⁵ zji¹ reflects the Proto-Gyalrongic etymon for “offspring, male” and corresponds regularly to *tə-tsa* in Cogtse Situ, *tə-zîê* in Bragbar Situ, and *zî* in Siyewu Khroskyabs. The vowel alternation *-ji* :: *-jî* observed between 𐞗⁵⁰⁷⁰ / 𐞗⁵⁷²¹ / 𐞗⁵⁵²⁵ zji¹ and 𐞗¹⁰⁸⁵ zji¹ is explained by two morpho-phonological processes.

First, the rhyme *-ji* in the base form 𐞗¹⁰⁸⁵ zji¹ shifts to a weakened sound *-jî* when occurring in the bound state, i.e. as the non-final component of a compound. This vowel alternation pattern is regular in Tangut and is also found with other compounds, such as 𐞗⁴⁶⁶⁹ 𐞗²⁵⁴¹ bji¹-dzjwo² (below-people) “servant” (with the first component 𐞗⁴⁶⁶⁹ bji¹ based on 𐞗³⁷⁹¹ bji² “below”), as well as in reduplication, e.g. 𐞗⁴⁶⁶⁹ 𐞗³⁷⁹¹ bji¹-bji² “below” (Gong 1993; Jacques 2014: 262; Wei 2022).

Second, the alternation between the lax vowel *-ji* and the tense vowel *-jî* is explained by the transphonologization of the compound initial *S- “collective prefix”, as explained in (2).

The four collective compounds in Table 2 can be classified into two categories based on their semantics. The first category includes 𐞗⁵⁰⁷⁰ 𐞗⁵⁰⁴⁹ zji¹ -wja¹ “father and son”, 𐞗⁵⁰⁷⁰ 𐞗⁰⁰⁹² zji¹ -mja¹ “mother and son”, and 𐞗⁵⁷²¹ 𐞗³⁵⁶² zji¹ -mjaa¹ “husband and wife”, which are collectives representing social relations. The second category includes 𐞗⁵⁵²⁵ 𐞗¹²⁴¹ zji¹ -lji² “children”, which is a general collective. However, the collective meanings are not always explicit in Tangut texts due to lexicalization accompanied by multiple semantic changes, which will be elaborated on in the following sections.

3.1. 𐞗⁵⁰⁷⁰ 𐞗⁵⁰⁴⁹ zji¹ -wja¹ “father and son”, 𐞗⁵⁰⁷⁰ 𐞗⁰⁰⁹² zji¹ -mja¹ “mother and son”

The two compounds 𐞗⁵⁰⁷⁰ 𐞗⁵⁰⁴⁹ zji¹ -wja¹ “father and son” and 𐞗⁵⁰⁷⁰ 𐞗⁰⁰⁹² zji¹ -mja¹ “mother and son” exhibit transparent semantics, representing the two most prominent social relationships: father-son (examples 3 and 4) and mother-son (example 5). Both

compounds follow the same word formation pattern, in which the collective prefix *S- precedes the two components overtly referring to the two participants in the denoted social relations, i.e. 𪛗⁵⁰⁷⁰ *zjɿ¹* “son” and 𪛗⁵⁰⁴⁹ *wja¹* “father”/𪛗⁰⁰⁹² *mja¹* “mother”.

In the two compounds, the younger generation participant consistently precedes the elder generation participant. This order is reversed compared to the Chinese terms 父子 *fù-zǐ* (father-son) and 母子 *mǔ-zǐ* (mother-son) in the source text, indicating that these collective forms have become fossilized in Tangut. In particular, in (3), the Tangut translation mostly adheres to a literal rendering of the original Chinese text. Terms like 𪛗⁵³⁰⁶𪛗³⁵⁰⁸ *dzjwɿ¹-bjɿ²* “ruler and minister” and 𪛗⁴⁴⁵⁷𪛗¹⁹¹⁰ *ljɿ²tjɿ²* “important relations” are adapted to match the word order of the Chinese original 君臣 *jūn-chén* and 大倫 *dà-lún*. In contrast, only 𪛗⁵⁰⁷⁰ 𪛗⁵⁰⁴⁹ “son and father” retains the native word order of Tangut.⁵

- (3) 𪛗²²¹⁹𪛗⁵⁹²⁵ 𪛗¹⁸²² 𪛗²⁹⁸³ 𪛗³⁵⁸³ 𪛗⁵⁰⁷⁰·𪛗⁵⁰⁴⁹ 𪛗²⁵³³ 𪛗³⁵⁸³
kjɿ¹.tsə¹ *ɲwuu¹* *u²* *tja¹* *zjɿ¹.wja¹* *djɿ²* *tja¹*
 PN:Jingzi to.say inside TOP son.and.father outside TOP
 𪛗⁵³⁰⁶𪛗³⁵⁰⁸ 𪛗²⁵⁴¹=𪛗¹¹³⁹ 𪛗⁴⁴⁵⁷ 𪛗¹⁹¹⁰ 𪛗⁰⁵⁰⁸
dzjwɿ¹.bjɿ² *dzjwo²=jɿ¹* *ljɿ²* *tjɿ²* *ɲwu²*
 ruler.and.ministers man=GEN important relations COP

景子曰：“內則父子，外則君臣，人之大倫也”

“Master Jing said, ‘In the family, there is the relation of father and son; outside, there is the relation of monarch and minister. These are the most important relations among men.’” (*Mengzi, Gongsunchou* 0404, Peng 2012: 122)

- (4) 𪛗⁰⁰¹⁰ 𪛗⁵²⁵⁹·𪛗³⁴²³ 𪛗⁵⁰⁷⁰·𪛗⁵⁰⁴⁹=𪛗⁰⁷²⁴ 𪛗¹⁰⁴⁵=𪛗⁴⁹⁵⁰ 𪛗¹⁶⁰⁸
zjɿ² *ljɿw¹.xjow¹* *zjɿ¹.wja¹=ɲjɿ²* *da²=ɲjɿ²* *law²*
 all PN:Liu.Xiang son.and.father= PL say= POST be.identical

皆如劉向父子所說 (Shi et al. 1993: 146)

“It was all like what Liu Xiang and his son said.” (*The Grove of Classification* 06.28B.7)

- (5) 𪛗²⁸⁶² 𪛗⁰²¹⁴ 𪛗²⁴⁸⁴ 𪛗¹⁵⁶⁷ 𪛗⁰⁴⁴⁸ 𪛗⁴³⁴²·𪛗⁵⁸⁷⁵ 𪛗⁵⁰⁷⁰·𪛗⁰⁰⁹²
ɲjɿ¹ *lu²* *niow¹* *gji²* *gji²* *dja²-zjɿ²* *zjɿ¹.mja¹*
 family be.poor LNK child one DIR1:PFV-to.sell son.and.mother

𪛗⁵¹⁶³ 𪛗¹¹⁶⁰

dzjow¹ *ka²*
 to.separate to.separate

家貧鬻一子， 母子離異 (Shi et al. 1993: 144)

“The family was impoverished and had to sell one of their children, thus the child and the mother were separated from each other.” (*The Grove of Classification* 06.27B.2)

3.2. 𪛗⁵⁷²¹𪛗³⁵⁶² *zjɿ¹-mjaa¹* “husband and wife, couple”

The collective 𪛗⁵⁷²¹𪛗³⁵⁶² *zjɿ¹-mjaa¹* “husband and wife, couple” follows a different word formation pattern. Nonetheless, the social relation between husband and wife, as represented in its meaning, is prominent in Tangut texts (e.g. 6).⁶

⁵ Peng (2012: 32–3) pointed out that there are instances of reversed translation of compounds in the Tangut version of *Mengzi*, which deserves further investigation.

⁶ There is no clear evidence for the phonological reconstruction of the initial of the character 𪛗⁵⁷⁵⁹ in example (6), which is represented by a capital C in the phonetic transcription.

- (6) 𪛗⁴⁶⁸⁶ 𪛘⁵²⁶² 𪛙⁰⁰⁰⁹ 𪛚⁵⁹⁹³ 𪛛⁰⁴⁷⁶ 𪛜²⁴⁸⁴ 𪛝⁵⁷²¹ 𪛞³⁵⁶² 𪛟⁴⁰²⁴ 𪛠⁴⁰²⁷ 256
 khjwā¹ ɲewr¹ ɕjwo¹ kha¹ wji¹ niow¹ zji¹.mjaa¹ zji² ɲji¹ 257
 county chaos arise POST exist LNK husband.and.wife two two 258
 𪛗⁰⁸⁰⁴ 𪛘⁵⁷⁵⁴ 𪛚²⁵⁹⁰ 𪛛⁵⁷⁵⁹ 259
 djɿ² -lju² wji² -Cja² 260
 DIR1:PFV -to.arrest_[A] DIR1:PFV-to.tie.up 261
 “There were troubles in the county, and the two spouses were caught and tied up.” (*Filial Piety*, Jacques 2007: 91) 262
 263
 264

The two components of 𪛝⁵⁷²¹ 𪛞³⁵⁶² zji¹-mjaa¹ “husband and wife, couple” both refer to the husband, with the wife unexpressed in the compound. The first component 𪛝⁵⁷²¹ zji¹ represents the bound state of 𪛛¹⁰⁸⁵ zji¹, based on its semantics of “male”. The second component 𪛞³⁵⁶² mjaa¹ is etymologically related to 𪛟⁴⁸²⁰ ma¹ “son-in-law”, reflecting the etymon for “son-in-law, bridegroom”, cognate with ɣmáɣ “son-in-law, husband” in Siyuewu Khroskyabs, *a-me-nmak* “son-in-law” in Japhug, *tə-nmak* “son-in-law” in Cogtse (Lin You-jing’s field data), and མག་པ། *mag.pa* “son-in-law, bridegroom” in Tibetan (Zhang et al. 2010: 2053). 265
 266
 267
 268
 269
 270
 271
 272

An alternative etymology proposed by Shi (2020: 461) suggests that 𪛞³⁵⁶² mjaa¹, occurring in the compound 𪛝⁵⁷²¹ 𪛞³⁵⁶² zji¹-mjaa¹, might be related to 𪛟²⁴³⁶ mjaa¹ “fruit”. However, our proposal that 𪛞³⁵⁶² mjaa¹ is etymologically related to 𪛟⁴⁸²⁰ ma¹ “son-in-law” better aligns with the semantics of this collective. The term 𪛟⁴⁸²⁰ ma¹ “son-in-law” is often used independently, as illustrated in (7). 273
 274
 275
 276
 277
 278

- (7) 𪛟⁰⁷⁴⁵ = 𪛟¹¹³⁹ 𪛟⁵⁰⁴⁹ 𪛟⁴⁸²⁰ = 𪛟¹¹³⁹ 𪛟⁰⁹⁴⁵ 279
 wji² = jji¹ wja¹ ma¹ = jji¹ tshja¹ 280
 PN:Rong=GEN father son.in.law=OBL be.angry 281
 “Rong’s father was angry at his son-in-law.” (*Filial Piety*, Jacques 2007: 88) 282
 283
 284

The rhyme alternation observed between 𪛞³⁵⁶² mjaa¹ and 𪛟⁴⁸²⁰ ma¹ can be explained by the compounding morphology that involves syllable compression. As previously explained in Section 2 (see also Lai 2023; 2024), the alternation between 𪛟⁴⁸²⁰ ma¹ (<*S[ü/ə].mak¹) and 𪛞³⁵⁶² mjaa¹ (<*Smjak¹) can be attributed to the syllabicity in the pre-initial element (see the first stages in 1a, 1b and 1c). It can be posited that the syllabic pre-initial would have been compressed in compounding, as 𪛞³⁵⁶² mjaa¹ (<*Smjak¹) is exclusively found in the compound 𪛝⁵⁷²¹ 𪛞³⁵⁶² zji¹-mjaa¹. 285
 286
 287
 288
 289
 290
 291

For an illustration of the compressing processes, see the sound changes presented in (8), with the non-compressed 𪛟⁴⁸²⁰ ma¹ “son-in-law” in (8a) and the compressed 𪛝⁵⁷²¹ 𪛞³⁵⁶² zji¹-mjaa¹ “husband and wife, couple” in (8b) (note that the tensing process of 𪛝⁵⁷²¹ zji¹ is omitted for clarity). The alternation between a Grade I rhyme -a in 𪛟⁴⁸²⁰ ma¹ and a Grade III rhyme -jaa, reconstructed with a medial -j- in 𪛞³⁵⁶² mjaa¹, could be explained by a harmonizing process mirroring the Grade III rhyme of 𪛝⁵⁷²¹ zji¹-. 292
 293
 294
 295
 296
 297
 298

- (8) a. Non-compressed: tensing 299
 *S[ü/ə].mak¹ > 𪛟⁴⁸²⁰ ma¹ “son-in-law” 300
 b. Compressed by compounding with 𪛝⁵⁷²¹ zji¹: lengthening 301
 *zji¹-S[ü/ə].mak¹ > *zji¹-Smjak¹ > 𪛝⁵⁷²¹ 𪛞³⁵⁶² zji¹-mjaa¹ “husband and wife, couple” 302
 303
 304

In modern Gyalrongic languages, social relation collectives can be formed by including both parties involved in the relation, or more commonly, by including only one of the 305
 306

parties. For instance, in Bragbar Situ, the term *koə-təa-jâ* “brothers and sisters, siblings” is derived from *tətətə* “younger siblings” and *a-jâ* “elder siblings” (Zhang 2020: 218, see also Section 5.1.1 for examples in Siyewu Khroskyabs). Alternatively, in Japhug, *kvndzi-ye* “grandparents and grandchildren” is based on *tr-ye* “grandchild” (Jacques 2021: 177). Notably, Tshobdun Gyalrong has two collective terms for “married couple”, one of them, *kəndze-nmá-nma*, is derived solely from *té-nma* “husband”, whereas the other, *keⁿdze-rjǎ-rjev*, is based only on *té-rjev* “wife” (Sun 1997), which corresponds exactly to the Tangut case.

3.3. $\text{𐵓}^{5525} \text{𐵓}^{1241} \text{zj}^1 \text{-lji}^2$ “children, baby”

The compound, $\text{𐵓}^{5525} \text{𐵓}^{1241} \text{zj}^1 \text{-lji}^2$ “children, baby” does not denote a specific social relation but rather serves as a general collective term. However, it is noteworthy that in Tangut texts, $\text{𐵓}^{5525} \text{𐵓}^{1241} \text{zj}^1 \text{-lji}^2$ can refer both to the collective concept of “children” and to an individual entity, such as a “small child”. For instance, in (10), $\text{𐵓}^{5525} \text{𐵓}^{1241} \text{zj}^1 \text{-lji}^2$ is followed by the singular indefinite marker $\text{𐵓}^{0448} \text{gji}^2$.

- (9) $\text{𐵓}^{4971} \text{𐵓}^{3798} \text{𐵓}^{0089} \text{𐵓}^{5525} \text{𐵓}^{1241} \text{𐵓}^{0497} \text{𐵓}^{0010} \text{𐵓}^{5354} = \text{𐵓}^{1139} \text{𐵓}^{3320} \text{𐵓}^{3202}$
 $\text{ejwi}^1 \text{ tsəj}^1 \text{ tɕhja}^1 \text{ zj}^1 \text{-lji}^2 \text{ newr}^2 \text{ zji}^2 \text{ thji}^2 = \text{jjj}^1 \text{ yiew}^1 \text{-rjir}^2$
age small LNK children many all DEM=GEN knowledge
 𐵓^{2699}
 nwə^1

to.know

年幼小兒皆學此才智 (Shi et al. 1993: 107)

“Even young children know this intelligence.” (*The Grove of Classification* 05.13B.6–7)

- (10) $\text{𐵓}^{2019} \text{𐵓}^{4861} \text{𐵓}^{2862} \text{𐵓}^{4889} \text{𐵓}^{1794} = \text{𐵓}^{4884} = \text{𐵓}^{1139} \text{𐵓}^{5525} \text{𐵓}^{1241} \text{𐵓}^{0448}$
 $\text{thja}^1 \text{ zj}^0 \text{ nji}^1 \text{ dzjwi}^1 \text{-o}^1 = \text{nji}^2 = \text{jjj}^1 \text{ zj}^1 \text{-lji}^2 \text{ gji}^2$
DEM time home neighbour=PL=GEN child one
 $\text{𐵓}^{3155} \text{𐵓}^{2921} \text{𐵓}^{4469}$
 $\text{tjij}^2 \text{-wjijr}^2 \text{-əji}^2$

to.hold-to.hold-to.go_[I]

時鄰家主人嬰兒抱出外 (Shi et al. 1993: 115)

“(Peng Zixun) carried the neighbour’s baby outside.” (*The Grove of Classification* 05.22A.6–7)

It is likely that $\text{𐵓}^{5525} \text{𐵓}^{1241} \text{zj}^1 \text{-lji}^2$, as the result of collective derivation in an earlier stage, has undergone semantic evolution, transitioning from a term for a group to a more general term. Semantic shifts of this nature are common; for instance, in Mandarin Chinese, the term 觀眾 *guān.zhòng* “audience” originally denoted a collective concept exclusively, but has gradually evolved into a general noun that can refer to both a group and an individual, as evidenced by the modern usage of 一個觀眾 *yí-gè guān.zhòng* (one-CLF audience) “a spectator”.

Another point worth noting is the etymology of the two components of $\text{𐵓}^{5525} \text{𐵓}^{1241} \text{zj}^1 \text{-lji}^2$. The first component is the bound state of $\text{𐵓}^{1085} \text{zji}^1$ “male, offspring, son”, while the second component $\text{𐵓}^{1241} \text{lji}^2$ is likely a diminutive suffix, as found in Siyewu Khroskyabs *z̄-lo* “son (hypocoristic)” and *m̄-lo* “darling (addressing younger generations)”. The meaning “small” in $\text{𐵓}^{5525} \text{𐵓}^{1241} \text{zj}^1 \text{-lji}^2$ may originate from this diminutive suffix. However, due to the obscurity of the collective morphology and the decreased productivity of the diminutive suffix, the lexical meaning of $\text{𐵓}^{5525} \text{𐵓}^{1241} \text{zj}^1 \text{-lji}^2$ indicating “small child, infant” has been transferred to the character $\text{𐵓}^{5525} \text{zji}^1$. This character then contrasts semantically with its base $\text{𐵓}^{1085} \text{zji}^1$, which is used as a kinship term meaning

“son, offspring”.⁷ Additionally, $\begin{smallmatrix} \text{𐺗} \\ \text{𐺗} \end{smallmatrix}^{5525} \text{zj}^1$ continues to be used in later compounding mechanisms, such as in $\begin{smallmatrix} \text{𐺗} \\ \text{𐺗} \end{smallmatrix}^{5619} \begin{smallmatrix} \text{𐺗} \\ \text{𐺗} \end{smallmatrix}^{5525} \text{mj}^2\text{-zj}^1$ meaning “mischievous child” (see Li 2012: 670).

4. Compound medial linker *-S-

A handful of compounds in Tangut have a tense vowel occurring in the second component. In such cases, the tense vowel likely originates from an *-S- element, which serves as a morphological linker connecting the two components. This compound medial *-S- later underwent transphonologization, resulting in a tense vowel in the second syllable. This process is represented in (11).

- (11) Transphonologization of the compound linker *-S-
*CV-S-CV > CV-CV

Table 3 provides a list of compounds in Tangut that potentially contain a compound linker *-S-. As the etymology of each component is not entirely transparent, we will offer detailed etymological analyses in the subsequent sections.

Table 3. Traces of compound linker *-S- in Tangut

	Attested forms	Pre-Tangut
“brothers”	$\begin{smallmatrix} \text{𐺗} \\ \text{𐺗} \end{smallmatrix}^{2447} \begin{smallmatrix} \text{𐺗} \\ \text{𐺗} \end{smallmatrix}^{0605} \text{ljo}^2\text{-tj}\phi^2$	*Cə-tjok ² -S-tjok ²
“worm”	$\begin{smallmatrix} \text{𐺗} \\ \text{𐺗} \end{smallmatrix}^{1888} \begin{smallmatrix} \text{𐺗} \\ \text{𐺗} \end{smallmatrix}^{1304} \text{b}\phi^2\text{-lu}^1$	* ⁿ bə ² -S-lu ¹
“brothers”	$\begin{smallmatrix} \text{𐺗} \\ \text{𐺗} \end{smallmatrix}^{0012} \begin{smallmatrix} \text{𐺗} \\ \text{𐺗} \end{smallmatrix}^{5873} \text{bj}\phi^1\text{-ku}^2$	* ⁿ bjϕ ¹ -S-ku ²
“snake”	$\begin{smallmatrix} \text{𐺗} \\ \text{𐺗} \end{smallmatrix}^{0111} \begin{smallmatrix} \text{𐺗} \\ \text{𐺗} \end{smallmatrix}^{0047} \text{lji}^1\text{-ljw}\phi^1$	* ⁿ lji ¹ -S-ljw[ɔ] ¹

4.1. $\begin{smallmatrix} \text{𐺗} \\ \text{𐺗} \end{smallmatrix}^{2447} \begin{smallmatrix} \text{𐺗} \\ \text{𐺗} \end{smallmatrix}^{0605} \text{ljo}^2\text{-tj}\phi^2$ “brothers”

There is general agreement that the two characters $\begin{smallmatrix} \text{𐺗} \\ \text{𐺗} \end{smallmatrix}^{2447} \text{ljo}^2$ and $\begin{smallmatrix} \text{𐺗} \\ \text{𐺗} \end{smallmatrix}^{0605} \text{tj}\phi^2$ both refer to “brother” in the context of male speakers (ms). However, there is some disagreement regarding their specific semantic representation. Kepping (1991: 5) interprets $\begin{smallmatrix} \text{𐺗} \\ \text{𐺗} \end{smallmatrix}^{2447} \text{ljo}^2$ as a term for the brothers of a male speaker, while $\begin{smallmatrix} \text{𐺗} \\ \text{𐺗} \end{smallmatrix}^{0605} \text{tj}\phi^2$ is considered a collective term, meaning “brothers”. Jacques (2012), among others, suggests that $\begin{smallmatrix} \text{𐺗} \\ \text{𐺗} \end{smallmatrix}^{2447} \text{ljo}^2$ and $\begin{smallmatrix} \text{𐺗} \\ \text{𐺗} \end{smallmatrix}^{0605} \text{tj}\phi^2$ encode relative age distinction, with the former denoting an elder brother and the latter a younger brother of a male speaker.

A closer examination of the usage of these two characters in Tangut texts shows that the semantic distinction between $\begin{smallmatrix} \text{𐺗} \\ \text{𐺗} \end{smallmatrix}^{2447} \text{ljo}^2$ and $\begin{smallmatrix} \text{𐺗} \\ \text{𐺗} \end{smallmatrix}^{0605} \text{tj}\phi^2$ does not pertain to relative age. As illustrated in (12) and (13), $\begin{smallmatrix} \text{𐺗} \\ \text{𐺗} \end{smallmatrix}^{2447} \text{ljo}^2$ can refer to both the younger or elder brothers of a man.

- (12) $\begin{smallmatrix} \text{𐺗} \\ \text{𐺗} \end{smallmatrix}^{4962} \begin{smallmatrix} \text{𐺗} \\ \text{𐺗} \end{smallmatrix}^{3262} \begin{smallmatrix} \text{𐺗} \\ \text{𐺗} \end{smallmatrix}^{5297} \begin{smallmatrix} \text{𐺗} \\ \text{𐺗} \end{smallmatrix}^{3583} \begin{smallmatrix} \text{𐺗} \\ \text{𐺗} \end{smallmatrix}^{0707} \begin{smallmatrix} \text{𐺗} \\ \text{𐺗} \end{smallmatrix}^{3738} = \begin{smallmatrix} \text{𐺗} \\ \text{𐺗} \end{smallmatrix}^{1139} \begin{smallmatrix} \text{𐺗} \\ \text{𐺗} \end{smallmatrix}^{2447} \begin{smallmatrix} \text{𐺗} \\ \text{𐺗} \end{smallmatrix}^{3798} \begin{smallmatrix} \text{𐺗} \\ \text{𐺗} \end{smallmatrix}^{0508}$
 $\text{we}^2.\text{khow}^1.\text{gioow}^1 \quad \text{tja}^1 \quad \text{tejiw}^1.\text{kow}^1=\text{ji}^1 \quad \text{ljo}^2 \quad \text{tsaj}^1 \quad \text{jwu}^2$
 PN:Weikangshu TOP PN:Zhougong=GEN brother^{ms} little be

“Shu Feng of Kang is the younger brother of the King Wu of Zhou.” (*The Twelve Kingdoms*, Juanzhong 17–13)

⁷ Note that cognates of $\begin{smallmatrix} \text{𐺗} \\ \text{𐺗} \end{smallmatrix}^{1085} \text{zji}^1$ “son” (Pre-Tangut *zja) in modern Gyalrongic, such as *tə-tsa* in Cogtse, *tə-zie* in Bragbar, can express multiple meanings of “son, offspring” (as a kinship term), “male” and “little child”.

- (13) 𦉳⁵⁶⁴⁶.席³⁸³⁰ 𦉳²⁴⁴⁷ 𦉳²⁸⁹³ 𦉳⁵⁵⁷⁴.𦉳⁵⁹²⁵ 𦉳⁵⁶⁴⁶.席³⁸³⁰=𦉳¹¹³⁹ 𦉳⁴⁰⁶⁹ 409
 tchjiw².njj² ljo² khwej² wji².tsa¹ tchjiw².njj²=jj¹ pjwɨr¹ 410
 PN:Zhouwang brother^{ms} elder PN:Weizi PN:Zhouwang=ACC to.advise 411
 𦉳¹⁹¹⁸-𦉳³⁵⁷⁵ 𦉳²⁴⁸⁴ 𦉳²⁵⁹⁰-𦉳²⁴⁷⁴ 412
 mji¹-nji² nioow¹ wji²-rar² 413
 NEG-to.listen_[A] LNK DIR1.PFV-to.go 414
 紂王兄微子因紂王拒諫而出奔 (Shi et al. 1993: 48) 415
 “King Zhou’s brother Weizi fled because King Zhou refused to accept his 416
 advice.” (*The Grove of Classification* 03.15A.2) 417
 418

In most cases, 𦉳⁰⁶⁰⁵ tjo² is used as a bound morpheme. The two characters 𦉳²⁴⁴⁷ 𦉳⁰⁶⁰⁵ ljo²-tjo² appear together as a compound, representing the collective concept of “brothers”, as in examples (14) and (15). 419
 420
 421

- (14) 𦉳⁵⁹⁸¹-𦉳³³⁰⁵ 𦉳⁰⁰²⁸ 𦉳⁴⁹⁰⁶ 𦉳¹²⁷⁴ 𦉳⁵⁰⁴⁹.𦉳⁰¹⁵⁴ 𦉳⁰⁰⁹².𦉳⁴⁸⁹³ 423
 a⁰-kjiw¹ lhjuu¹ gjwi² wo² wja¹.o¹ mja¹.wji¹ 424
 one-year mourning to.wear_[A] should grandfather grandmother 425
 𦉳²⁴⁴⁷.𦉳⁰⁶⁰⁵ 𦉳⁵³⁰⁵.𦉳⁰⁰⁰⁴ 𦉳¹⁵⁶¹ 𦉳¹⁵⁴³ 426
 ljo².tjo² wjj².la² nji² mjo¹ 427
 brother^{ms} paternal.uncle.and.his.wife brother’s.child^{ms} non.adopted 428
 “One should wear mourning clothes for one year to one’s grandfather, grand- 429
 mother, brothers (for a male speaker), paternal uncle or his wife, and to one’s 430
 brother’s non-adopted son (for a male).” (Tiansheng Code, Jacques 2012: 238) 431
 432
- (15) 𦉳³³¹⁷ 𦉳⁰⁰⁹² 𦉳⁰¹⁸⁷ 𦉳⁰⁴⁴⁸ 𦉳³⁰⁹⁹ 𦉳¹⁸⁰¹ 𦉳²³⁶⁵ 𦉳¹⁵²⁶=𦉳³⁸¹⁸ 433
 lew¹ mja¹ nar² gj¹² dzjij¹ dzjij² pha¹ tshji²=mjjij² 434
 only mother old one to.live remain other look.after=NMLZ:S/A 435
 𦉳²⁴⁴⁷.𦉳⁰⁶⁰⁵ 𦉳¹⁶⁰² 𦉳²¹⁹⁴ 436
 ljo².tjo² ɲowr² mjjij¹ 437
 brothers^{ms} all not.exist 438
 “My old mother lives alone, and I have no brothers to look after her.” (*Filial 439
 Piety*, Jacques 2007: 14) 440
 441

The compounding morphology also accounts for the two phonological alternations 442
 between the terms 𦉳²⁴⁴⁷ ljo² and 𦉳⁰⁶⁰⁵ tjo²: (i) initial lenition alternation between l- 443
 and t-, and (ii) the tense vowel alternation between -jo and -jō. According to the transpho- 444
 nologization rule, these alternations are due to the presence of the compound linker *-S-. 445
 446

- (16) a. Leniting and dropping (see 1d) 447
 *Cə-tjok² > *Cə-ljok² > ljo² 𦉳²⁴⁴⁷ 448
 b. Non-leniting (see 1c) 449
 *-S-tjok² > -tjō² 𦉳⁰⁶⁰⁵ 450
 451
 452

The sound changes in (16) suggest that 𦉳⁰⁶⁰⁵ tjo² and 𦉳²⁴⁴⁷ ljo² share the same stem 453
 *-tjok² “brother (male speaker)” in Pre-Tangut. The lenition observed in 𦉳²⁴⁴⁷ ljo² is due 454
 to the loss of a presyllable (the fourth type shown in (1d) in Section 2), for instance, the 455
 possessive prefix *tə- still present in East Gyalrongic. Conversely, the tense vowel in 𦉳⁰⁶⁰⁵ 456
 tjo² results from the original *-S- blocking lenition (see the third type (1c) in Section 2). 457

This argument is further supported by comparative evidence. First, sibling terminology 458
 in Tangut is characterized by a clear opposition between male and female speaking 459

subsystems (Jacques 2012; Kepping 1991; Shi 2020: 462–3), a feature inherited from Proto-Gyalrongic. As illustrated in Table 4, this terminological system is also preserved in modern Gyalrongic languages, such as Siyuewu, Japhug, and Situ (with Bragbar Situ having lost the female-speaking sub-system, see Zhang and Fan 2020), where no relative age distinction is evident.

Table 4. Sibling terms in Gyalrongic languages

	West Gyalrongic			East Gyalrongic		
	Tangut	Pre-Tangut	Geshiza (Horpa)	Siyuewu (Khroskyabs)	Japhug	Bragbar Situ
“Brother ^{ms} ”	𐄂𐄃 ²⁴⁴⁷ <i>ljo</i> ²	*Cə-tjok ²	<i>ri</i>	<i>dóy</i>	<i>tx-xɬɣ</i>	<i>tə-ktiák</i>
“Brother ^{ms} ”	𐄂𐄃 ⁰⁶⁰⁵ <i>tjɔ</i> ²	*-S-tjok ²				
“Sister ^{ms} ”	𐄂𐄃 ⁰⁵⁴⁹ <i>niɔ</i> ¹	*S.njVm ¹	<i>sno</i>	<i>snəm</i>	<i>tx-snam</i>	<i>tə-snām</i>
“Brother ^{ws} ”	𐄂𐄃 ⁰³⁵⁵ <i>mju</i> ¹	*mjo ¹		<i>mô</i>	<i>tx-wɤmu</i>	
“Sister ^{ws} ”	𐄂𐄃 ³³⁶¹ <i>kiej</i> ¹	*S.kej ¹	<i>sq^he</i>	<i>sqí</i>	<i>tx-sq^haj</i>	

Second, similar initial lenition alternations observed in Tangut 𐄂𐄃²⁴⁴⁷ *ljo*² and 𐄂𐄃⁰⁶⁰⁵ *tjɔ*² are also found in cognate “brother (male speaker)” terms in Horpa languages (West Gyalrongic). As illustrated in Table 5, the lenited form *ri* (from *Cə-to, see Lai 2023) is used as an unbound form, while the non-lenited form *sti* (from *s-to) with an *s-* pre-initial occurs as the second component of the collective compound “brothers” (See Lai 2023: 362–5 for a detailed explanation). The pre-initial *s-* is comparable to the tense vowel in Tangut 𐄂𐄃⁰⁶⁰⁵ *tjɔ*², with both reflecting a compound linker *-S- (see Section 5.2 for further comparison).

Table 5. Initial lenition alternation of the terms for “brothers (male speaker)” in West Gyalrongic languages

	Pre-Tangut	Geshiza Horpa	Bawang Horpa	Siyuewu Khroskyabs
“Brother ^{ms} ”	*Cə-tjok ²	<i>ri</i>	<i>ri</i>	<i>dóy</i>
“Brother ^{ms} ” (col.)	*Cə-tjok ² -S-tjok ²	<i>rmæ-s-ti</i>	<i>rmɛ-s-ti</i>	<i>rmâ-s-təy</i>

4.2. 𐄂¹⁸⁸⁸ 𐄃¹³⁰⁴ *bə*²-*lu*¹ “worms”

The second component of the compound 𐄂¹⁸⁸⁸ 𐄃¹³⁰⁴ *bə*²-*lu*¹ “worms” also carries a tense vowel,⁸ which likely originates from a compound linker *S- that transphonologized onto the second syllable, as illustrated in (17).

⁸ Jacques (2014: 72) mentions that the first component can also be written as 𐄂⁵²⁷⁰ *bə*¹, which is annotated as “maggot” in the *Tangut-Chinese Dictionary* (Li 2012: 629). However, the character 𐄂⁵²⁷⁰ *bə*¹ does not appear in our corpus, and there is insufficient evidence to determine the semantic differences between 𐄂⁵²⁷⁰ 𐄃¹³⁰⁴ *bə*¹-*lu*¹ and 𐄂¹⁸⁸⁸ 𐄃¹³⁰⁴ *bə*²-*lu*¹. Nevertheless, it is clear that 𐄂⁵²⁷⁰ *bə*¹ and 𐄂¹⁸⁸⁸ *bə*² are etymologically related, though the phonological alternations presented await further investigation.

(17) *ⁿbə²-S-lu¹ > bə²-lu¹ 𧈧¹⁸⁸⁸ 𧈩¹³⁰⁴

The compound 𧈧¹⁸⁸⁸ 𧈩¹³⁰⁴ bə²-lu¹ “worms” and 𧈪²⁴⁴⁷ 𧈫⁰⁶⁰⁵ ljo²-tjɔ² “brothers” may share the same compounding mechanism. This involves bisyllabification through stem reduplication or juxtaposition of different roots, linked by *-S-, to form a compound representing a collective concept. The semantic differences between the compound 𧈧¹⁸⁸⁸ 𧈩¹³⁰⁴ bə²-lu¹ “worms” and the unbound root 𧈧¹⁸⁸⁸ bə² “worm” can be observed in textual examples. The compound 𧈧¹⁸⁸⁸ 𧈩¹³⁰⁴ bə²-lu¹ involves a collective reading (e.g. 18), whereas 𧈧¹⁸⁸⁸ bə² denotes singular concepts, such as a particular type of insect such as locusts in (19a) and silkworm in (19b).

(18) 𧈧¹⁸⁸⁸ 𧈩¹³⁰⁴ bə²-lu¹ “worms”

a. 𧈪⁰²⁸⁹ 𧈫³²⁶⁶ 𧈬³²⁷³=𧈭¹¹³⁶ 𧈧¹⁸⁸⁸ 𧈩¹³⁰⁴ 𧈮⁴³⁴² 𧈯²⁷²⁴
 we².dzju² yar²=gu² bə².lu¹ dja² -təhju¹
 Prefect chest=LOC:inside worms DIR1.IPFV-to.have[A]

府君胸中有蟲 (Shi et al. 1993: 129)

“Prefect, there are worms in your chest.” (*The Grove of Classification* 06.11B.7)

b. 𧈰²⁰¹⁹ 𧈱⁰⁸⁶⁴ 𧈲²⁴³⁶=𧈳¹¹³⁹ 𧈴²²⁶² 𧈵⁵¹³⁴ 𧈧¹⁸⁸⁸ 𧈩¹³⁰⁴=𧈪⁰⁷²⁴
 thja¹ dew².mjaa¹=jjj¹ dzjwɔw¹.we¹ bə².lu¹=nji²
 DEM fruit=ACC birds worms=PL

𧈶³⁵²⁷ 𧈷⁴⁵¹⁷ 𧈸⁰⁷³⁴

mja¹-dzji¹-mo²

IRR-to.eat[A]-IRR

其子將為鳥雀小蟲所食 (Shi et al. 1993: 61)

“The fruit, probably the birds and insects would eat it.” (*The Grove of Classification* 03.29B.2)

(19) 𧈧¹⁸⁸⁸ bə² “insect”

a. 𧈪⁵³⁵⁴ 𧈧¹⁸⁸⁸ 𧈮⁵⁶⁸⁸ 𧈯²⁴⁸⁴ 𧈰¹⁹¹⁸ 𧈱⁵⁷⁵⁴ 𧈲⁴⁸⁸⁴
 thji² bə² wa² niow¹ mji¹-lju²-nji²
 DEM insect what LNK:cause NEG-to.catch[A]-2PL

何不捕此蟲? (Shi et al. 1993: 86)

“Why didn’t you catch this insect?” (*The Grove of Classification* 04.15B.2)

b. 𧈳⁴²⁷⁴ 𧈴³⁶⁶⁸ 𧈧¹⁸⁸⁸ 𧈵³⁰⁴² 𧈶⁵³⁶⁴
 sow¹ lji¹ bə² jur¹ lew²
 mulberry to.plant silkworm to.feed woollen.material
 𧈷³⁸⁷² 𧈸¹⁶⁰⁰ 𧈹⁰⁶³⁰ 𧈺⁵⁸⁸⁰ 𧈻⁴⁸⁶⁹ 𧈼⁰¹⁵⁴
 kjir¹ thu¹ la¹ ŋwu² kji¹.o¹
 woollen.material to.weave to.weave LNK: INSTR father.in.law

𧈽³⁹⁸⁶ 𧈾⁴⁸⁹³=𧈿¹¹³⁹ 𧉀⁰¹⁰⁵ 𧉁⁴⁸⁸⁷ 𧉂⁵¹¹³

nji¹.wji¹=jjj¹ kju¹-tshwew¹-wji¹

mother.in.law= ACC to.support-to.serve-to.do[A]

種桑育蠶，紡績織紵衣之，奉養公婆 (Shi et al. 1993: 119)

“(I) plant mulberry and raise silkworms, spin and weave clothes to support parents-in-law.” (*The Grove of Classification* 06.01B.6-7)

Comparative evidence supports the hypothesis that the tense vowel in the second component 𧈩¹³⁰⁴ lu¹ originates from a linker *-S- rather than being an inherent part of the root. As illustrated in Table 6, the bisyllabic form for “worm(s)” in West Gyalrongic contains a shared innovative root bə- as the first component. The second component corresponds to the Gyalrongic etymon for “insect, worm”, which is preserved as

unbound lexemes in East Gyalrongic with the animal prefix, such as Japhug *qa-juu* “worm” (Jacques 2014: 72) and Bragbar Situ *kə-lú* “worm”. While the correspondence of the initials is regular,⁹ the tense vowel (< *S-) in Tangut 𐞗¹³⁰⁴ *lu*¹ lacks a counterpart in modern Gyalrongic. This suggests that the tense vowel (< *S-) in Tangut comes from an extra-root element, most likely the compound linker *S- necessary for lexical bi-syllabification.¹⁰

Table 6. Comparison of the terms for “insect(s), worm(s)” in Gyalrongic languages

	West Gyalrongic		East Gyalrongic	
Pre-Tangut	Geshiza Horpa	Khroskyabs	Japhug	Bragbar Situ
* ^m bə ² -S-lu ¹	bə.zo	bə.jo	qa-juu	kə-lú

4.3. 𐞗⁰⁰¹²𐞗⁵⁸⁷³ *bju*¹-*ku*² “brothers”

The compound 𐞗⁰⁰¹²𐞗⁵⁸⁷³ *bju*¹-*ku*² “brothers” is not found in textual attestations but is recorded in dictionaries such as *Homophones* and *Sea of characters*, where it is defined as a collective term meaning “brothers”. Although there is no textual evidence that the two components can be used individually in Tangut, both components have potential cognates in other Sino-Tibetan languages.

The first component, 𐞗⁰⁰¹² *bju*¹, is likely related to the first syllable of Tibetan བུ་སྤྱལ། *bu.s-pun* “brothers” (Zhang et al. 2010: 1830).¹¹ Note that the Tibetan form contains a collective prefix *s-* in the second component *spun* (see Section 5.1.2).

The second component 𐞗⁵⁸⁷³ *ku*² is related to Burmese အကို *akui* (Proto-Burmish **kuiw*) “elder brother” and is further connected to Tibetan མུ་ *khu*, which originally meant “maternal uncle” (see Nagano 1994), and Old Chinese 舅 *[g](r)u? “maternal uncle” (Hill 2019: 77, 239; Zhang et al. 2019).¹² The semantic discrepancy is similar to the case of ལྷ་པོ་ *skud-po* “brother-in-law”, derived from མུ་ *khu* “maternal uncle” with the circumfix *s-Σ-d* (< **s-khu-d*, see Benedict 1942, Section 5.1.2).

Comparative evidence suggests that both roots lack a pre-initial element, and the tense vowel in the second component of 𐞗⁰⁰¹²𐞗⁵⁸⁷³ *bju*¹-*ku*² “brothers” likely originates from a compound linker *S-, serving to link the two co-ordinative roots. However, this proposal remains to be verified with clearer etymological evidence.

4.4. 𐞗⁰¹¹¹𐞗⁰⁰⁴⁷ *li*¹-*ljwu*¹ “snake”

The compound 𐞗⁰¹¹¹𐞗⁰⁰⁴⁷ *li*¹-*ljwu*¹ in Tangut typically signifies “snake”, as evidenced in (20), with no instances of its components being used independently. This term likely

⁹ The correspondence among the initial *l-* in Tangut, *z-* in Geshiza, *j-* in Khroskyabs, and *j-* in Japhug is regular, as observed with the etymon for “hand”: 𐞗³⁴⁸⁵ *la*¹ in Tangut, *za* in Geshiza, *jóy* in Siyuewu Khroskyabs, *tu-jax* (with an indefinite possessive prefix) in Japhug. Both etyma for “worm” and “hand” may originate from a palatalized lateral, however, the different reflexes such as *kə-lú* and *ta-ják* in Situ Gyalrong require further explanation.

¹⁰ One of the reviewers pointed out that cognate forms in Rma, Ronghong *bal*, Longxi *bù ló*, and Mianchi *bò lè*, all meaning “worm”, also lack an *-S-* element. This evidence suggests that the tense vowel in the Tangut form 𐞗¹⁸⁸⁸𐞗¹³⁰⁴ *bə²-lu¹* “worms” is a Tangut internal issue.

¹¹ The Tangut rhyme *-u* corresponds regularly to Tibetan *-u* (Jacques 2014: 64–75), as in 𐞗⁴⁶¹⁴ *nju*² “to milk” :: ལུ་ *nu.ma* “breasts” and in 𐞗³³⁸⁸ *ŋju*² “to cry, weep” :: ལུ་ *ngu* “to cry, weep”.

¹² Tangut 𐞗⁰⁵⁹⁷ *yi*¹ “maternal uncle” is also a potential cognate. The lenition alternation between 𐞗⁵⁸⁷³ *ku*² “brothers” and 𐞗⁰⁵⁹⁷ *yi*¹ “maternal uncle” can be explained by the sound laws in Section 2. Specifically, the character 𐞗⁵⁸⁷³ *ku*² belongs to the non-leniting type (see 1c) and 𐞗⁰⁵⁹⁷ *yi*¹ to the leniting and dropping type (see 1d). However, the tonal alternation and the rhyme alternation between *-u* and *-ji* remain to be explained.

originates from an ideophone, capturing the serpentine movement characteristic of a snake, later extending its meaning to the animal itself. It is potentially related to Wobzi Khroskyabs *z-bæ-ljâ~lja* “to lie prone, to crawl”.

- (20) 𪛗⁰¹¹¹.𪛗⁰⁰⁴⁷ 𪛗³⁹¹² 𪛗⁴⁷⁶² 𪛗⁴²⁴⁶=𪛗²⁹⁸³ 𪛗³⁴⁹¹
ljî¹.ljwü¹ bji² tɕhji¹ lhejr²=u² rjur¹
 creep.creep step to.go woods=LOC:inside star
 大蛇は歩行し、叢の中に光る。(Nishida 1986: 53)
 “The serpent slithered, and the stars sparkled in the woods.” (*The Ode on Monthly Pleasures*, 5-61)

Should the hypothesized ideophonic origin of this compound hold true, its formation process can be elucidated by reduplication. Although Gong (1993) does not document the *-ji :: -jwu* alternation pattern,¹³ it is plausible to hypothesize that the first component 𪛗⁰¹¹¹ *ljî¹* serves as the reduplicant, while the second component 𪛗⁰⁰⁴⁷ *ljwü¹* represents the root. Thus, the tense vowel in the second component is likely not inherent to the root but instead results from the transphonologization of the compound linker **-S-*. However, this hypothesis requires validation through the establishment of phonological alternation rules.

5. Origins of the **S* elements in Tangut compounds

Internal evidence suggests that the collective prefix **S-* and the compound medial **-S-* serving as a linking element in Tangut must be distinguished synchronically. These two morphological processes are attested with only a few traces, which provide important clues for revealing the regular morphology of an earlier stage.

This section provides a comparative study of the corresponding morphemes, showing that the two morphological processes are also distinct at the West Gyalrongic level, shared among Tangut, Horpa and Khroskyabs. The collective **S-* likely represents inherited morphology with parallels in Tibetan (Section 5.1), whereas the compound medial **-S-* represents a stage of morphological merging in West Gyalrongic, with an unclear origin (Section 5.2).

5.1. Historical status of the collective prefix **S-*

In both West Gyalrongic and Tibetan, traces of a collective prefix **S-* have been retained, indicating that this morphology is likely archaic.

5.1.1. West Gyalrongic

Within West Gyalrongic, Siyuewu Khroskyabs retains a collective prefix *s-*, observed in a few collectives of social relations (see Table 7). Similar to Tangut, social relation collectives in Siyuewu also involve kinship terms, with both parts in the denoted social relation overtly expressed by the two components (see Table 5).

The first compound *s-γ̂-vzî* “maternal uncle and his sister’s children” is composed of *γ̂-* (the bound state of *γ̂ô* “maternal uncle”) and a bound root *†vzî* “sister’s children (for a male speaker)”. Both components are inherited Proto-Gyalrongic kinship terms. The Siyuewu *γ̂ô* “maternal uncle” reflects the Gyalrongic etymon for maternal uncle, as in Tangut 𪛗⁰⁵⁹⁷ *γ̂ji¹* (Pre-Tangut **CV-kji¹*) “maternal uncle” and *a-kû* “maternal uncle” in

¹³ While the vowel alternation between *i :: u* is also attested with 𪛗⁰⁶⁸⁰𪛗⁰³⁶⁶ *tɕhji²-tɕhjuu²* “reversal, be inverted” (Wei 2022: 20), the absence of the medial *-w-* in the reduplicant remains to be explained.

Bragbar Situ (Zhang and Fan 2020). Although Siyuewu $\dagger vzi$ is unattested as an unbound morpheme, it is related to Tangut 𑖇²¹³⁴ zjw_i^1 (Pre-Tangut *S-pǎ.tsa) “cross nephew, child of different-sex siblings”.¹⁴

Table 7. Social relation collective *s-* in Siyuewu Khroskyabs

Gloss	Form	Compound 1	Compound 2
“maternal uncle and his nephew”	<i>s-γə-vzi</i>	<i>əγə</i> “maternal uncle”	$\dagger vzi$ “nephew”
“maternal aunt and her nephew”	<i>s-la-vdi</i>	<i>lalā</i> “maternal aunt”	$\dagger vdi$ “nephew”
“grandfather and grandchild”	<i>s-və-vlā</i>	<i>vāvə</i> “grandfather”	<i>vlā</i> “grandson”
“Goatherd’s family” (house name)	$\dagger\text{-}ts^h\hat{e}\text{-}jəm$	$ts^h\hat{e}d$ “goat”	<i>jəm</i> “house”

Note: the † indicates forms without unbound attestation.

The second collective *s-la-vdi* “maternal aunt and her sister’s children” is built upon *la-* (the bound state of *lalā* “maternal aunt”) and $\dagger vdi$ “nephew, sister’s children (for a female speaker)”. The unbound root $\dagger vdi$ is cognate with *vdé* in Njorogs Khroskyabs (Yin 2007) and *tə-mdi* “nephew” in Cogtse Situ (Lin You-Jing’s field note), among others.

In the third collective *s-və-vlā* “grandfather and grandchild”, the first component *və-* represents the Proto-Gyalrongic term for “grandfather”, preserved in Bragbar *ta-wū* and Japhug *tx-wuu*, and also occurs as the second component of *vā-və* “grandfather” in Siyuewu.

Since the three collectives mentioned above contain bound roots that are not attested individually, it is likely that the collective prefix *s-* in Siyuewu is archaic. However, it is worth noting that this morphological process seems to have lost its productivity in Siyuewu only recently. A remnant of the *s-* collective is found in a Siyuewu house name $\dagger\text{-}ts^h\hat{e}\text{-}jəm$ (COL-goat-house), in which the initial consonant \dagger is a conditioned variant of the collective *s-* prefix (see Lai 2016 on the Siyuewu *s-* allomorphy). The $\dagger\text{-}ts^h\hat{e}\text{-}jəm$ family are goatherds, and the house name reflects the close relation between goats and their owners.¹⁵ The $\dagger\text{-}ts^h\hat{e}\text{-}jəm$ family became goatherds during the people’s commune period in China (1950–60s), and the house name was thus created during that time. This indicates that the collective prefix *s-* remains productive in Siyuewu up to that time.

5.1.2. Tibetan

The preservation of the collective prefix *S- in both Tangut and Siyuewu suggests that this morphology dates back to Proto-West-Gyalrongic. Moreover, the presence of potential cognate morphemes in Tibetan further supports the antiquity of the West Gyalrongic collective prefix *S-.

In Tibetan, there are two collective circumfixes *s-Σ-d* and *s-Σ-n* (Benedict 1942: 323–5; Hill 2014: 628), in which the *s-* element is comparable to the West Gyalrongic collective

¹⁴ This interpretation follows Kepping (1991), as 𑖇²¹³⁴ zjw_i^1 is used for both sister’s children (for a male speaker) and brother’s children (for a female speaker). Cognates of this term in East Gyalrongic include Japhug *tx-ftsa* “father’s sister’s child, sister’s child”, and Bragbar *a-tsá-pu* “father’s sister’s child” (Jacques 2012; Zhang and Fan 2020). While the semantic mismatch requires further explanation, the cognacy is supported by regular sound correspondence.

¹⁵ Similar derivations are found in Japhug (East Gyalrongic), although using a non-cognate prefix *kyndzi-*, such as *kyndzi-ts^hyt* (COL-goat) “goat and its owners”, *kyndzi-mbro* (COL-horse) “horseman and his horse” and *kyndzi-ftsox* (COL-female.hybrid.yak) “female hybrid yak and its owners” (Jacques 2021: 177).

*S-. Both circumfixes in Tibetan are unproductive and appear in only five collective terms derived from kinship terms, as listed in (21) and (22).¹⁶ It is worth noting that the loss of aspiration in the derived forms with the *s-* pre-initial is explained by Shafer’s law, i.e. *s-kh- > sk-, *s-ph- > sp- (see Hill 2011; Li 1933; Shafer 1950–51).

(21) Collective circumfix *s-Σ-d*

པ *pha* “father” → པ་སླད *pha-spad* “father and children”

མ *ma* “mother” → མ་སླད *ma-smad* “mother and children”

ཁྱ *khu* “paternal uncle” → སླད་པོ་ *skud-po* “brother-in-law”

(22) Collective circumfix *s-Σ-n*

ཕྱ *phu* “brother” → སླན་ *spun* “siblings”, also in བུ་སླན་ *bu-spun* “brothers”

Except for སླད་པོ་ *skud-po* “brother-in-law”, which bears a non-transparent semantic relationship with the base form ཁྱ *khu* “paternal uncle”,¹⁷ the other forms in (21) and (22) clearly convey collective meanings. It is plausible to assume that the collective meaning in these forms likely originates from the *s-* prefix. However, the exact mechanism by which this prefix interacts with the nominal suffixes *-n* and *-d* to form a circumfix remains unclear.

5.2. Historical status of the compound linker *-S-

The compound linker *-S-, while leaving only a few traces in Tangut, appears to be a morphological process shared among West Gyalrongic languages. Data from modern West Gyalrongic languages further indicate that the linker *-S- is used not only to derive co-ordinative compounds with collective meaning, as seen in Tangut, but also to form determinative compounds.

Table 8 shows compounds with a linker *-s-* in Siyuewu Khroskyabs, along with glosses of their components.

Table 8. Traces of compound linker *-S- in Siyuewu Khroskyabs

	Gloss	Form	Compound 1	Compound 2
Co-ordinative	brothers	<i>rmã-s-təy</i>	<i>rmã</i> “man, others”	<i>dóy</i> “brother”
Co-ordinative	adult woman	<i>və-s-mé</i>	<i>və</i> “adult woman”	<i>†mé</i> “woman”
Co-ordinative	recently	<i>fsê-s-kʰə</i>	<i>fsê</i> “to be early”	<i>kʰə</i> “to be late”
Determinative	white cedar	<i>læ-s-pʰrəm</i>	<i>læ-po</i> “cedar”	<i>pʰrəm</i> “to be white”
Determinative	black cedar	<i>læ-s-ɲéŋ</i>	<i>læ-po</i> “cedar”	<i>ɲéŋ</i> “to be black”
Determinative	pig pen	<i>pʰəy-s-jəm</i>	<i>pʰəy</i> “pig”	<i>jəm</i> “house”

Co-ordinative compounds in Siyuewu juxtapose two synonymous or antonymous components. For example, the compound *rmã-s-təy* “brothers” combines two synonymous

¹⁶ Benedict (1942: 324) also includes ཚན་ *tshan* (derived from ཚམོ་ *tsha.po*) in the *s-Σ-n* pattern, as seen in མ་ཚན་ *ma-tshan* “cousins on the mother’s side” and པ་ཚན་ *pha-tshan* “cousins on the father’s side”. He attributes the absence of the *s-* element to Tibetan phonotactic rules.

¹⁷ Tibetan ཁྱ *khu* originally meant maternal uncle but later underwent semantic shift to paternal uncle (for details, see Nagano 1994). Benedict (1942: 323–4) suggests that སླད་པོ་ *skud-po* “brother-in-law” reflects an equation between consanguineous and affinal kinship terms (mother’s brother’s son = wife’s brother) under cross cousin marriage: a man marries his mother’s brother’s daughter, making his mother’s brother’s son (*skud*) his wife’s brother.

components: *rmê* “man, others” and *dóy* “brother”, connected by the linker *-s-* (for a discussion of the etymology see Section 4.1, Table 5).

The compound *və-s-mé* is formed through a similar process.¹⁸ Its first component *və-*, though unattested as a free morpheme, is related to the second component in *gə-və* “wife” (further related to Tangut 𐞗𐞧²⁴⁵⁵ 𐞗𐞧²¹²⁹ *gji²-bjij²* “wife”, see Lai et al. 2020). The second component *-s-mé* reflects the Gyalrongic etymon for “woman, girl”, as in Japhug *tu-me* and Bragbar Situ *tə-mí*.

The compound *fsê-s-k^hə* juxtaposes two antonymous components, *fsê* “to be early” and *k^hə* “to be late”, linked by *-s-*. This compound expresses a collective meaning “early or late”, hence “recently”.

Siyuewu determinative compounds can be further divided into two types based on their internal syntax – left-headed and right-headed.¹⁹ An example of a left-headed compound concerns *lə-s-p^hrəm* “white cedar” and *lə-s-nəŋ* “black cedar”, which denote two sub-species of cedar. In such compounds, the linker *-s-* connects the head *lə-* “cedar” and the modifiers, *p^hrəm* “to be white” and *nəŋ* “to be black”.

The term *p^hay-s-jəm* “pig pen” is a case of right-headed compound, in which the linker *-s-* connects the modifier *p^hay* “pig” and the head *jəm* “house”.²⁰

Traces of the compound linker **-S-* are also found in Horpa languages, as exemplified by Geshiza Horpa in Table 9.²¹

Table 9. Traces of compound linker **-S-* in Geshiza (data from Honkasalo 2019)

	Gloss	Form	Component 1	Component 2
Co-ordinative	“brothers”	<i>rmæ-s-ti</i>	<i>rmæ-</i> “sibling”	<i>ri</i> “brother”
Determinative	“tree roots”	<i>s^hə-s-q^ha</i>	<i>s^hə-p^ho</i> “tree”	<i>ʈq^ha</i> “root”
?	“clothes”	<i>ts^hæ-z-gə</i>	<i>ʈts^hæ</i> ?	<i>gə</i> “wear”

Geshiza *rmæ-s-ti* “brothers”, which is cognate with Siyuewu *rmê-s-təy*, is a co-ordinative compound, in which the two synonymous components are linked together by *-s-*.

The determinative compound *s^hə-s-q^ha* “tree roots” is right-headed, in which the linker *-s-* connects the modifier *s^hə-* “tree, wood (bound state)” and the head *ʈq^ha* “root”, a bound root. The second component is related to Guanyinqiao Khroskyabs *qé* “root”, *sni-q^hé* “tongue root”, and Japhug Gyalrong *u-qa* “root”). These Gyalrongic cognates suggest a

¹⁸ An alternative explanation is that *vəsmé* “adult woman” may be a borrowing from Tibetan བུད་མེད *bud.med* “woman”, pronounced [vətmet] in local Amdo Tibetan. However, in most cases, Siyuewu faithfully reproduces Tibetan codas in loanwords as *-d*. If it had borrowed Tibetan བུད་མེད *bud.med*, it would likely have been realized as *ʈvədméd*, instead of *vəsmé*.

¹⁹ The terms “left-headed” and “right-headed” follow Bialek (2018).

²⁰ It should be mentioned that *s-jəm* is also reanalysed as an independent noun in Siyuewu, meaning “lair, net”. This process involves reanalysing the compound linker *-s-* as the pre-initial of the root, which might explain the presence of an additional *s-* pre-initial in Tangut 𐞗𐞧⁰⁹⁶⁰ *mjjij¹* “woman, girl” (Pre-Tangut **S-mjjij¹*), Geshiza *s-me*, etc. (Lai et al. 2020: 177). These forms correspond to the Gyalrongic etymon for “woman” or “girl”, as seen in Japhug *tu-me* and Bragbar Situ *tə-mí*. Notably, no sibilant initial is ever recorded in East Gyalrongic, suggesting that this *s-* pre-initial element is an innovation exclusive to West Gyalrongic (see Lai et al. 2020: 177).

²¹ The three Geshiza compounds in Table 9 have direct cognates in Bawang, *rmə-s-ti* “brothers”, *ts^hə-z-gwə* “clothes”, *s^hə-s-q^ha* “tree roots” (data from Yang 2021).

proto-form for “root” without a sibilant pre-initial. Thereby the presence of the linker *-s-* in *s^hə-s-q^ha* “tree roots” suggests the productivity of the compound linker *-s-* after the branch-off of Horpa.

The third compound *ts^hə-z-gə* “clothes”, in which the linker *-s-* is assimilated to *-z-*, is currently only found in Geshiza and Bawang (*ts^he-z-gwə* “clothes”). Parallel compounds with cognate roots but lacking a sibilant linker morpheme exist within Gyalrongic, such as Khanggsar Stau (Horpa) *tsə-gə*, Siyewu Khroskyabs *ts^hə-gí*, Tangut 𐞗⁵⁶¹⁰𐞗⁵⁵⁹⁸ *tshji¹-gjwi²* “clothes” (Li 2012: 667, 669), and beyond, Pengbuxi Minyag *tse-ŋgə* (Gao 2016), Guiqiong *ts^hε³³-wε⁵³* (Zàngmiǎnyǔ Yǔyīn hé Cíhuì Biānxiězǔ 1992), all meaning “clothes”. In the Geshiza form *ts^hə-z-gə* “clothes”, while the second component *-gə* is related to the verb “to wear”, the first component *ts^hə-* is not attested as a free lexeme. The sporadic appearance of the compound linker *-s-* in Geshiza and Bawang forms resembles the case of “worms”, where the sibilant linker is found only in Tangut 𐞗¹⁸⁸⁸𐞗¹³⁰⁴ *bə²-lu¹* “worms” (see Table 6). It suggests that the compound linker **-S-* was still productive upon the separation of Tangut and Horpa.

While the compound linker **-S-* probably began to emerge during the stage of Proto-West-Gyalrongic, such morphology is not expected to have arisen spontaneously; it may have resulted from the merger of multiple morphemes. For example, the *-s-* linker in co-ordinative compounds might be related to a collective prefix, re-analysed from a compound medial context like Tibetan མ་སྐད་ *ma-smad* “mother and daughter”. However, re-analysing this pattern from a collective prefix to a linker in determinative compounds would require generalization.

Alternatively, the Geshiza compound *ts^hə-z-gə* “clothes” might suggest another possibility. If we consider *ts^hə* cognate with Ersu *ts^ha⁵⁵* “classifier for clothes” (Zàngmiǎnyǔ Yǔyīn hé Cíhuì Biānxiězǔ 1992), then Geshiza *ts^hə-z-gə* “clothes” can be analysed as a left-headed determinative compound, with the second part being a nominalized verb. Thus, the linker *-z-* likely originates from a sibilant nominalizer (**S-*) used to derive oblique nouns (i.e. the instrument with which to wear).²² This oblique nominalizer is no longer productive in West Gyalrongic but leaves traces in Wobzi Khroskyabs *s-p^həm* “lid” (derived from *p^həm* “to cover”) (Lai 2017: 158, 511), as well as in the nominalizing tense vowel in Tangut, e.g. 𐞗⁵²⁰⁵ *ya¹* “sword, weapon” (derived from 𐞗⁵⁶⁵³ *ya¹* “to butcher, chop”) (Jacques 2014: 256).²³ We defer a full exploration of this issue to future research.

6. Conclusion

The present research uncovers two previously unrecognized sources of vowel tensing in Tangut: the collective prefix (**S-*) and the compound linker (**-S-*). These findings not only deepen our understanding of Tangut nominal morphology but also shed light on the approximate age of these two morphemes. Comparative evidence suggests that the collective prefix **S-* can be traced back at least to the common ancestor of Burmo-Qiangic and Tibetic, while the compound linker **-S-* appears to have emerged during the West-Gyalrongic period.

This study also raises questions about the historical status of linker elements in Sino-Tibetan compounding morphology, which are often discerned through traces with obscure origins (see for instance Downer 1959: 289–90 on the non-final *qusheng* in Old Chinese compounds; Bialek 2018: 233–45 on the linker elements in Old Tibetan). Evidence from West Gyalrongic further supports the idea that compound linkers were

²² The internal syntax of *ts^hə-z-gə* “clothes” could be similar to Chinese 掛飾 *guà-shì* “hanging ornament”.

²³ The oblique nominalizer **S-* in West Gyalrongic languages corresponds to a highly productive syllabic oblique nominalizer *sV-* in East Gyalrongic (Jacques 2016; Sun 2006; Sun and Lin 2007; Zhang 2023, etc.).

historically unstable, potentially resulting from morphological merger and subject to rapid disappearance.

By investigating Tangut tense vowels, this study underscores the importance of combining careful analysis of textual attestations with comparative studies of related languages for the morphological reconstruction of highly eroded languages. We do not, however, claim to have definitely resolved the origins of Tangut tense vowels. Future research with new examples will be necessary to refine or amend our conclusions.

Funding information. This research is supported by the Horizon Europe Marie-Sklodowska-Curie Actions Postdoctoral Fellowship (101110215 Kinship Systems in Gyalrong: History and Transformation; Zhang Shuya), the Irish Research Council under the SFI-IRC Pathway Programme (Project ID: 21/PATH-A/9374, Gyalrongic unveiled: Languages, Heritage, Ancestry; Yunfan Lai) and Nanyang Technological University, Singapore under the Nanyang Assistant Professorship (NAP 2024, #024576-00001; Yunfan Lai).

References

- Arakawa, Shintaro. 2012. "Re-analysis of 'Tangut-Tibetan' phonological materials", in Nathan W. Hill (ed.), *Medieval Tibeto-Burman Languages IV*, 171–89. Leiden: Brill.
- Arakawa, Shintaro. 2014. *Seikabun kongōkyō no kenkyū* 西夏文金剛經の研究 (Studies on the Tangut version of *Vajracchedikā-prajñāpāramitā*). Kyoto: Shōkadō Shoten.
- Baxter, William H. and Laurent Sagart. 2014. *Old Chinese: A New Reconstruction*. Oxford: Oxford University Press.
- Benedict, Paul K. 1942. "Tibetan and Chinese kinship terms", *Harvard Journal of Asiatic Studies* 6/3–4, 313–37.
- Bialek, Joanna. 2018. *Compounds and Compounding in Old Tibetan. A Corpus Based Approach. Vol 1*. Marburg: Indica et Tibetica Verlag.
- Comrie, Bernard, Martin Haspelmath and Balthasar Bickel. 2008. "The Leipzig glossing rules: conventions for interlinear morpheme-by-morpheme glosses", Department of Linguistics of the Max Planck Institute for Evolutionary Anthropology and the Department of Linguistics of the University of Leipzig. <https://www.eva.mpg.de/lingua/pdf/Glossing-Rules.pdf>. Retrieved 3 December 2024.
- Downer, G.B. 1959. "Derivation by tone-change in Classical Chinese", *Bulletin of the School of Oriental and African Studies* 22/2, 258–90.
- Gao, Yang. 2016. "Description de la langue menya: Phonologie et syntaxe", Doctoral dissertation, École des hautes études en sciences sociales, Paris.
- Gong, Hwang-cherng. 1993. "Xīxià yǔ de yīnyùn zhuǎnhuàn yǔ gòucí fǎ 西夏語的音韻轉換與構詞法 (Phonological alternations and word formation in Tangut)", *Bulletin of the Institute of History and Philology* 64/4, 952–4.
- Gong, Hwang-cherng. 1999. "Xīxià yǔ de jīn yuányīn jí qǐ yuán 西夏語的緊元音及其起源 (the tense vowels in Tangut and their origins)", *Bulletin of the Institute of History and Philology* 70/2, 531–58.
- Gong, Hwang-cherng. 2003. "Tangut", in Randy LaPolla and Graham Thurgood (eds), *The Sino-Tibetan Languages*, 602–20. London: Routledge.
- Hill, Nathan W. 2011. "An inventory of Tibetan sound laws", *Journal of the Royal Asiatic Society* 21/4, 441–57.
- Hill, Nathan W. 2014. "Tibetan", in Rochelle Lieber and Pavol Štekauer (eds), *The Oxford Handbook of Derivational Morphology*, 620–30. Oxford: Oxford University Press.
- Hill, Nathan W. 2019. *The Historical Phonology of Tibetan, Burmese, and Chinese*. Cambridge: Cambridge University Press.
- Honkasalo, Sami. 2019. "A grammar of Eastern Geshiza", Doctoral dissertation, University of Helsinki.
- Jacques, Guillaume. 2007. *Textes tangoutes I, Nouveau recueil sur l'amour parental et la piété filiale*. (Languages of the World/Text Collections 25.) Munich: Lincom Europa.
- Jacques, Guillaume. 2012. "The Tangut kinship system", in Nathan W. Hill (ed.), *Medieval Tibeto-Burman Languages IV*, 211–58. Leiden: Brill.
- Jacques, Guillaume. 2014. *Esquisse de phonologie et de morphologie historique du tangoute*. Leiden: Brill.
- Jacques, Guillaume (ed.). 2015. *Dictionnaire japhug-chinois-français*. Paris: Projet HimalCo.
- Jacques, Guillaume. 2016. "Subjects and objects in Japhug and relativization", *Journal of Chinese Linguistics* 44/1, 1–28.
- Jacques, Guillaume. 2021. *A Grammar of Japhug*. Berlin: Language Science Press.
- Kepping, Ksenija Borisovna. 1991. "Tangut (Xixia) degrees of mourning", *Linguistics of the Tibeto-Burman Area* 14/2, 1–63.

- Kychanov, Evgenij Ivanovich and Mikhail Viktorovich Sofronov. 1963. *Issledovanija po fonetike tangutskogo jazyka (predvaritel'nye rezul'taty) Исследования по фонетике тангутского языка (предварительные результаты)* [Studies on the Phonetics of the Tangut Language (Preliminary Results)]. Moscow: Izdatel'stvo vostochnoi literatury. 919
- Lai, Yunfan. 2016. "Causativisation in Wobzi and other Khroskyabs dialects", *Cahiers de Linguistique – Asie Orientale* 45/22, 148–75. 920
- Lai, Yunfan. 2017. "Grammaire du khroskyabs de Wobzi", Doctoral dissertation, Université Sorbonne Nouvelle (Paris 3). 921
- Lai, Yunfan. 2023. "Lenition alternation in West Gyalrongic and its implication for Southeast Asian panchronic phonology", *Diachronica* 40/3, 341–83. 922
- Lai, Yunfan. 2024. "Mutual predictiveness of sound correspondences hints at reconstruction and language sub-grouping: The case of Gyalrongic preinitials", *Diachronica* (online first). 923
- Lai, Yunfan, Xun Gong, Jesse P. Gates and Guillaume Jacques. 2020. "Tangut as a West Gyalrongic language", *Folia Linguistica Historica* 41/1, 171–203. 924
- Li, Fang-Kuei. 1933. "Certain phonetic influences of the Tibetan prefixes upon the root initials", *Bulletin of the Institute of History and Philology* 6/2, 135–57. 925
- Li, Fanwen. 1997. *Xià-Hàn zìdiǎn 夏漢字典 (Tangut–Chinese Dictionary)*. Beijing: China Social Sciences Press. 926
- Li, Fanwen. 2012. *Jiǎnmíng Xià-Hàn zìdiǎn 簡明夏漢字典 (Tangut–Chinese Dictionary)*. Second edition. Beijing: Zhongguo Shehui Kexue Chubanshe. 927
- Lin, You-Jing. 2016. *Jiāróngyǔ Zhuókèjì huà yǔfǎ biāozhù wénběn 嘉戎語卓克基話語法標注文本 (Fully analysed texts of Cogtse Rgyalrong with a sketch grammar of the language)*. Beijing: Social Sciences Academic Press. 928
- Meillet, Antoine. 1925. *La méthode comparative en linguistique historique*. Oslo: H. Aschehoug & Co. 929
- Nagano, Sadako. 1994. "A note on the Tibetan kin terms khu and zhang", *Linguistics of the Tibeto-Burman Area* 17/2, 103–15. 930
- Nishida, Tatsuo. 1964. *Seika-go no kenkyū: Seika-go no saikōsei to seika moji no kaidoku 西夏語の研究：西夏語の再構成と西夏文字の解説 (A study of the Tangut language: Reconstruction of the Tangut language and decipherment of Tangut script)*. Tokyo: Zauhō kankōkai. 931
- Nishida, Tatsuo. 1986. "Seika-go 'Getsu gestu raku shi' no kenkyū 西夏語『月々樂詩』の研究 (Research on the Tangut Ode on Monthly Pleasures)". 京都大學文學部研究紀要 (*Memoirs of the Department of Literature, Kyoto University*) 25, 1–116. 932
- Peng, Xiangqian. 2012. *Xixiàwén "Mèngzǐ" zhěnglǐ yánjiū 西夏文《孟子》整理研究 (Research and Compilation of the Book of Mencius in Tangut Script)*. Shanghai: Shanghai Guji Chubanshe. 933
- Shafer, Robert. 1950/51. "Studies in the morphology of the Bodic verbs", *Bulletin of the School of Oriental and African Studies* 13/3, 13/4. 702–24, 1017–31. 934
- Shi, Jinbo. 2020. *Tangut Language and Manuscripts: An introduction*. Leiden and Boston: Brill. 935
- Shi, Jinbo, Zhenhua Huang and Hongyin Nie. 1993. 類林研究 *Lèi lín yánjiū* (A study on "the Grove of Classification"). Yinchuan: Ningxia Renmin Chubanshe. 936
- Solonin, Kirill J. 1995. *Dvenadcat' carstv: Perevod s tangustkogo, predislovie i kommentarii Двенадцать царств: Перевод с тангутского, предисловие и комментарии (The Twelve Kingdoms: Translated from Tangut, Preface, and Comments)*. Saint Petersburg: Centr Peterburgskogo Vostokovedeni. 937
- Sun, Jackson T.-S. 1997. "Caodeng" (unpublished manuscript contributed to Sino-Tibetan Etymological Dictionary and Thesaurus as electronic text documents). Accessed via STEDT database <<http://stedt.berkeley.edu/search/>> on 15 March 2024. 938
- Sun, Jackson T.-S. 2006. "Jiāróngyǔ dòngcí de pàishēng xíngtài 嘉戎語動詞的派生形態 (Derivational morphology in the Rgyalrong verb)". 民族語文 (*Minority Languages of China*) 3/4, 3–14. 939
- Sun, Jackson T.-S. and You-Jing Lin. 2007. "Constructional variation in Rgyalrong relativization: how to make a choice?", in *Pre-Conference Proceedings of the International Workshop on Relative Clauses*, 205–26. Taipei: Institute of Linguistics, Academia Sinica. 940
- Wang, Jingru. 1982. "Xixiàyǔ yīnxì dǎoyán 西夏語音系導言 (Introduction to Tangut phonology)". 民族語文 (*Minority Languages of China*) 2, 1–13. 941
- Wei, Hao. 2022. "Xixiàyǔ dòngcí chóngdié yánjiū 西夏語動詞重疊研究 (Research on verb reduplication of the Tangut language)". 西夏研究 (*Tangut Research*) 2, 17–21. 942
- Yang, Chih-Fan. 2021. "Bāwàng Huò'ěr yǔ shí, tǐ, shìzhèng, qíngtài fānchóu de xíngtài jǔfǎ 巴旺霍爾語時、體、示證、情態範疇的形態句法" (The morpho-syntax of tense, aspect, evidentiality and modality in Bawang Horpa). Doctoral dissertation, National Taiwan Normal University. 943
- Yin, Weibin. 2007. *Yèlóng Lāwūróngyǔ Yánjiū 業隆拉姆戎語研究 (Study on the 'Jorogs Lavrung language)*. Beijing: Nationalities Press. 944

- Zàngmiǎnyǔ Yǔyīn hé Cíhuì Biānxiězǔ (ed.). 1992. *Zàngmiǎnyǔ yǔyīn hé cíhuì* 藏缅语语音和词汇 (Phonology and vocabulary of Tibeto-Burman languages). Beijing: China Social Sciences Press. 970
- Zhang, Shuya. 2020. “Le rgyalrong situ de Brag-bar et sa contribution à la typologie de l’expression des relations spatiales: l’orientation et le mouvement associé”, Doctoral dissertation, Institut National des Langues et Civilisations Orientales. 971
- Zhang, Shuya. 2023. “Sītǔ Jiāróngyǔ Báiwān huà héxīn lùn yuán guānxì huà 四土嘉绒语白湾话核心论元关系化” (Core argument relativisation in the Brag-bar dialect of Situ Rgyalrong). *语言科学 (Linguistic Sciences)* 22/1, 76–89. 972
- Zhang, Shuya and Jingming Fan. 2020. “Brag-bar kinship system in synchronic and diachronic perspectives”, *Bulletin of the School of Oriental and African Studies* 83/3, 479–503. 973
- Zhang, Shuya, Guillaume Jacques and Yunfan Lai. 2019. “A study of cognates between Gyalrong languages and Old Chinese”, *Journal of Language Relationship* 17/1, 73–92. 974
- Zhang, Yisun et al. (eds). 2010. *Zàng-Hàn dàcídiǎn* 藏汉大辞典 (Tibetan–Chinese Comprehensive Dictionary). 11th edition. Beijing: 民族出版社 (Publishing House of Minority Nationalities). 975
- 976
- 977
- 978
- 979
- 980
- 981
- 982
- 983
- 984
- 985
- 986
- 987
- 988
- 989
- 990
- 991
- 992
- 993
- 994
- 995
- 996
- 997
- 998
- 999
- 1000
- 1001
- 1002
- 1003
- 1004
- 1005
- 1006
- 1007
- 1008
- 1009
- 1010
- 1011
- 1012
- 1013
- 1014
- 1015
- 1016
- 1017
- 1018
- 1019
- 1020

Cite this article: Zhang, Shuya., Lai, Yunfan. 2025. “New origins to vowel tensing in Tangut: internal and comparative evidence”, *Bulletin of the School of Oriental and African Studies* 1–20. <https://doi.org/10.1017/S0041977X24000673>