

ARTICLE

## New origins to vowel tensing in Tangut: internal and comparative evidence<sup>1</sup>

Shuya Zhang<sup>1</sup>  and Yunfan Lai<sup>2</sup> 

<sup>1</sup>Shanghai Jiao Tong University, Shanghai, China and <sup>2</sup>Nanyang Technological University, Singapore  
**Corresponding author:** Lai Yunfan; [khroskyabs@gmail.com](mailto: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 parallels 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.<sup>2</sup> 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 𐞖<sup>1475</sup> *bji*<sup>1</sup> “to be thin” and 𐞖<sup>1789</sup> *bji*<sup>1</sup> “to make thin” are not only semantically related but also

<sup>1</sup> 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.

<sup>2</sup> 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 <sub>i</sub> -)	Tense vowel (< *s.C <sub>i</sub> -)
Causativization	𐽪 <sup>4614</sup> <i>nju</i> <sup>2</sup> “to suckle (vt)”	𐽪 <sup>4834</sup> <i>nju</i> <sup>2</sup> “to milk (vt)”
Causativization	𐽪 <sup>4767</sup> <i>ɣwej</i> <sup>1</sup> “to fight (vi)”	𐽪 <sup>2487</sup> <i>ɣwej</i> <sup>1</sup> “to cause to fight (vt)”
Denominalization	𐽪 <sup>0181</sup> <i>la</i> <sup>1</sup> “hole”	𐽪 <sup>2658</sup> <i>la</i> <sup>1</sup> “to bury”
Nominalization	𐽪 <sup>3469</sup> <i>sjj</i> <sup>2</sup> “to know”	𐽪 <sup>1771</sup> <i>sjj</i> <sup>2</sup> “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 𐽪<sup>1084</sup> *ɣa*<sup>2</sup> :: 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 texts<sup>3</sup> with comparative data from modern Sino-Tibetan languages, particularly from modern West Gyalrongic languages.<sup>4</sup>

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,

<sup>3</sup> 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.

<sup>4</sup> This paper adopts the recent classification of Tangut as 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; Siyewu 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<sub>i</sub>-, 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<sub>i</sub>- (without the period to differentiate them from the \*S.C<sub>i</sub>- consistently used in this paper), yield long vowels in Tangut. Second, syllabic pre-initials \*Că.C<sub>i</sub>- yield various phonation types, including tensing, rhoticizing and labial medializing, and result in initial lenition. Third, syllabic pre-initials \*Cũ.C<sub>i</sub>- yield the same phonation types but without initial lenition. Fourth, pre-initials with a neutral vowel Cə.C<sub>i</sub>- 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<sub>i</sub>- (lengthening)  
\*Ckar<sup>1</sup> > 𐰽𐰚<sup>5682</sup> *kaar*<sup>1</sup> “to weigh”
- b. Tangut \*Că.C<sub>i</sub>- (leniting)  
\*Să.kia<sup>1</sup> > \*Să.yia<sup>1</sup> > \*S.yia<sup>1</sup> > 𐰽𐰚<sup>0439</sup> *yie*<sup>1</sup> “to cook”
- c. Tangut \*Cũ.C<sub>i</sub>- (non-leniting)  
\*Sũ.kjo<sup>1</sup> > \*S.kjo<sup>1</sup> > 𐰽𐰚<sup>2278</sup> *kju*<sup>1</sup> “spring onion”
- d. Tangut \*Cə.C<sub>i</sub>- (leniting and dropping)  
\*Cə-ko<sup>1</sup> > \*Cə-yo<sup>1</sup> > \*yo<sup>1</sup> > 𐰽𐰚<sup>2750</sup> *yu*<sup>1</sup> “head”

Since the second type \*Că.C<sub>i</sub>- (in 1b) and the third type \*Cũ.C<sub>i</sub>- (in 1c) both transphonologize into tense vowels, they can be reconstructed as \*S[ũ/ă].C<sub>i</sub>-. 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<sub>i</sub>- 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<sub>i</sub>- 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 <sup>5070</sup> 𐞗 <sup>5049</sup> zji <sup>1</sup> -wja <sup>1</sup>	*S-zji <sup>1</sup> -wjaC <sup>1</sup>	𐞗 <sup>1085</sup> zji <sup>1</sup> “offspring, male”	𐞗 <sup>5049</sup> wja <sup>1</sup> “father”
“mother and son”	𐞗 <sup>5070</sup> 𐞗 <sup>0092</sup> zji <sup>1</sup> -mja <sup>1</sup>	*S-zji <sup>1</sup> -mjaC <sup>1</sup>	𐞗 <sup>1085</sup> zji <sup>1</sup> “offspring, male”	𐞗 <sup>0092</sup> mja <sup>1</sup> “mother”
“husband and wife”	𐞗 <sup>5721</sup> 𐞗 <sup>3562</sup> zji <sup>1</sup> -mjaa <sup>1</sup>	*S-zji <sup>1</sup> -mjaaC <sup>1</sup>	𐞗 <sup>1085</sup> zji <sup>1</sup> “offspring, male”	𐞗 <sup>3562</sup> mjaa <sup>1</sup> “son-in-law”
“children, baby”	𐞗 <sup>5525</sup> 𐞗 <sup>1241</sup> zji <sup>1</sup> -lji <sup>2</sup>	*S-zji <sup>1</sup> -lji <sup>2</sup>	𐞗 <sup>1085</sup> zji <sup>1</sup> “offspring, male”	𐞗 <sup>1241</sup> lji <sup>2</sup> “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, zji<sup>1</sup>, which is etymologically related to 𐞗<sup>1085</sup> zji<sup>1</sup> (Pre-Tangut \*zja) “son, offspring, male”. The form 𐞗<sup>1085</sup> zji<sup>1</sup> 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 Siyewwu Khroskyabs. The vowel alternation -ji :: -jî observed between 𐞗<sup>5070</sup> / 𐞗<sup>5721</sup> / 𐞗<sup>5525</sup> zji<sup>1</sup> and 𐞗<sup>1085</sup> zji<sup>1</sup> is explained by two morpho-phonological processes.

First, the rhyme -ji in the base form 𐞗<sup>1085</sup> zji<sup>1</sup> 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 𐞗<sup>4669</sup> 𐞗<sup>2541</sup> bji<sup>1</sup>-dzjwo<sup>2</sup> (below-people) “servant” (with the first component 𐞗<sup>4669</sup> bji<sup>1</sup> based on 𐞗<sup>3791</sup> bji<sup>2</sup> “below”), as well as in reduplication, e.g. 𐞗<sup>4669</sup> 𐞗<sup>3791</sup> bji<sup>1</sup>-bji<sup>2</sup> “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 𐞗<sup>5070</sup> 𐞗<sup>5049</sup> zji<sup>1</sup> -wja<sup>1</sup> “father and son”, 𐞗<sup>5070</sup> 𐞗<sup>0092</sup> zji<sup>1</sup> -mja<sup>1</sup> “mother and son”, and 𐞗<sup>5721</sup> 𐞗<sup>3562</sup> zji<sup>1</sup> -mjaa<sup>1</sup> “husband and wife”, which are collectives representing social relations. The second category includes 𐞗<sup>5525</sup> 𐞗<sup>1241</sup> zji<sup>1</sup> -lji<sup>2</sup> “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. 𐞗<sup>5070</sup> 𐞗<sup>5049</sup> zji<sup>1</sup> -wja<sup>1</sup> “father and son”, 𐞗<sup>5070</sup> 𐞗<sup>0092</sup> zji<sup>1</sup> -mja<sup>1</sup> “mother and son”

The two compounds 𐞗<sup>5070</sup> 𐞗<sup>5049</sup> zji<sup>1</sup> -wja<sup>1</sup> “father and son” and 𐞗<sup>5070</sup> 𐞗<sup>0092</sup> zji<sup>1</sup> -mja<sup>1</sup> “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. 𐰽<sup>5070</sup> *zji<sup>1</sup>* “son” and 𐰽<sup>5049</sup> *wja<sup>1</sup>* “father”/𐰽<sup>0092</sup> *mja<sup>1</sup>* “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 𐰽<sup>5306</sup>𐰽<sup>3508</sup> *dzjw<sup>1</sup>-bj<sup>2</sup>* “ruler and minister” and 𐰽<sup>4457</sup>𐰽<sup>1910</sup> *ljij<sup>2</sup>tji<sup>2</sup>* “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 𐰽<sup>5070</sup> 𐰽<sup>5049</sup> “son and father” retains the native word order of Tangut.<sup>5</sup>

- (3) 𐰽<sup>2219</sup>𐰽<sup>5925</sup> 𐰽<sup>1822</sup> 𐰽<sup>2983</sup> 𐰽<sup>3583</sup> 𐰽<sup>5070</sup> 𐰽<sup>5049</sup> 𐰽<sup>2533</sup> 𐰽<sup>3583</sup>  
*kji<sup>1</sup>.tsə<sup>1</sup>* *ɲwu<sup>1</sup>* *u<sup>2</sup>* *tja<sup>1</sup>* *zji<sup>1</sup>.wja<sup>1</sup>* *dji<sup>2</sup>* *tja<sup>1</sup>*  
 PN:Jingzi to.say inside TOP son.and.father outside TOP  
 𐰽<sup>5306</sup>𐰽<sup>3508</sup> 𐰽<sup>2541</sup>=𐰽<sup>1139</sup> 𐰽<sup>4457</sup> 𐰽<sup>1910</sup> 𐰽<sup>0508</sup>  
*dzjw<sup>1</sup>.bj<sup>2</sup>* *dzjwo<sup>2</sup>=ji<sup>1</sup>* *ljij<sup>2</sup>* *tji<sup>2</sup>* *ɲwu<sup>2</sup>*  
 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) 𐰽<sup>0010</sup> 𐰽<sup>5259</sup> 𐰽<sup>3423</sup> 𐰽<sup>5070</sup> 𐰽<sup>5049</sup>=𐰽<sup>0724</sup> 𐰽<sup>1045</sup>=𐰽<sup>4950</sup> 𐰽<sup>1608</sup>  
*zji<sup>2</sup>* *ljiw<sup>1</sup>.xjow<sup>1</sup>* *zji<sup>1</sup>.wja<sup>1</sup>=nj<sup>2</sup>* *da<sup>2</sup>=rji<sup>2</sup>* *law<sup>2</sup>*  
 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) 𐰽<sup>2862</sup> 𐰽<sup>0214</sup> 𐰽<sup>2484</sup> 𐰽<sup>1567</sup> 𐰽<sup>0448</sup> 𐰽<sup>4342</sup>-𐰽<sup>5875</sup> 𐰽<sup>5070</sup> 𐰽<sup>0092</sup>  
*nji<sup>1</sup>* *lu<sup>2</sup>* *niow<sup>1</sup>* *gji<sup>2</sup>* *gji<sup>2</sup>* *dja<sup>2</sup>-zji<sup>2</sup>* *zji<sup>1</sup>.mja<sup>1</sup>*  
 family be.poor LNK child one DIR1:PFV-to.sell son.and.mother

𐰽<sup>5163</sup> 𐰽<sup>1160</sup>  
*dzjow<sup>1</sup>* *ka<sup>2</sup>*  
 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. 𐰽<sup>5721</sup>𐰽<sup>3562</sup> *zji<sup>1</sup>-mjaa<sup>1</sup>* “husband and wife, couple”

The collective 𐰽<sup>5721</sup>𐰽<sup>3562</sup> *zji<sup>1</sup>-mjaa<sup>1</sup>* “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).<sup>6</sup>

<sup>5</sup> 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.

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

- (6) 藏<sup>4686</sup> 藏<sup>5262</sup> 藏<sup>0009</sup> 藏<sup>5993</sup> 藏<sup>0476</sup> 藏<sup>2484</sup> 藏<sup>5721</sup> 藏<sup>3562</sup> 藏<sup>4024</sup> 藏<sup>4027</sup>  
 khjwā<sup>1</sup> ηewr<sup>1</sup> ejwo<sup>1</sup> kha<sup>1</sup> wji<sup>1</sup> niow<sup>1</sup> zji<sup>1</sup>.mjaa<sup>1</sup> zji<sup>2</sup> njii<sup>1</sup>  
 county.chaos arise POST exist LNK husband.and.wife two two  
 藏<sup>0804</sup> 藏<sup>5754</sup> 藏<sup>2590</sup> 藏<sup>5759</sup>  
 dji<sup>2</sup> -lju<sup>2</sup> wji<sup>2</sup> -Cja<sup>2</sup>

DIR1:PFV -to.arrest<sub>[A]</sub> DIR1:PFV-to.tie.up

“There were troubles in the county, and the two spouses were caught and tied up.” (*Filial Piety*, Jacques 2007: 91)

The two components of 藏<sup>5721</sup> 藏<sup>3562</sup> zji<sup>1</sup>-mjaa<sup>1</sup> “husband and wife, couple” both refer to the husband, with the wife unexpressed in the compound. The first component 藏<sup>5721</sup> zji<sup>1</sup> represents the bound state of 藏<sup>1085</sup> zji<sup>1</sup>, based on its semantics of “male”. The second component 藏<sup>3562</sup> mjaa<sup>1</sup> is etymologically related to 藏<sup>4820</sup> mā<sup>1</sup> “son-in-law”, reflecting the etymon for “son-in-law, bridegroom”, cognate with γmáγ “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).

An alternative etymology proposed by Shi (2020: 461) suggests that 藏<sup>3562</sup> mjaa<sup>1</sup>, occurring in the compound 藏<sup>5721</sup> 藏<sup>3562</sup> zji<sup>1</sup>-mjaa<sup>1</sup>, might be related to 藏<sup>2436</sup> mjaa<sup>1</sup> “fruit”. However, our proposal that 藏<sup>3562</sup> mjaa<sup>1</sup> is etymologically related to 藏<sup>4820</sup> mā<sup>1</sup> “son-in-law” better aligns with the semantics of this collective. The term 藏<sup>4820</sup> mā<sup>1</sup> “son-in-law” is often used independently, as illustrated in (7).

- (7) 藏<sup>0745</sup> = 藏<sup>1139</sup> 藏<sup>5049</sup> 藏<sup>4820</sup> = 藏<sup>1139</sup> 藏<sup>0945</sup>  
 wji<sup>2</sup> =jji<sup>1</sup> wja<sup>1</sup> mā<sup>1</sup> =jji<sup>1</sup> tshja<sup>1</sup>  
 PN:Rong=GEN father son.in.law=OBL be.angry

“Rong’s father was angry at his son-in-law.” (*Filial Piety*, Jacques 2007: 88)

The rhyme alternation observed between 藏<sup>3562</sup> mjaa<sup>1</sup> and 藏<sup>4820</sup> mā<sup>1</sup> 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 藏<sup>4820</sup> mā<sup>1</sup> (<\*S[ũ/ə].mak<sup>1</sup>) and 藏<sup>3562</sup> mjaa<sup>1</sup> (<\*Smjak<sup>1</sup>) 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 藏<sup>3562</sup> mjaa<sup>1</sup> (<\*Smjak<sup>1</sup>) is exclusively found in the compound 藏<sup>5721</sup> 藏<sup>3562</sup> zji<sup>1</sup>-mjaa<sup>1</sup>.

For an illustration of the compressing processes, see the sound changes presented in (8), with the non-compressed 藏<sup>4820</sup> mā<sup>1</sup> “son-in-law” in (8a) and the compressed 藏<sup>5721</sup> 藏<sup>3562</sup> zji<sup>1</sup>-mjaa<sup>1</sup> “husband and wife, couple” in (8b) (note that the tensing process of 藏<sup>5721</sup> zji<sup>1</sup> is omitted for clarity). The alternation between a Grade I rhyme -a in 藏<sup>4820</sup> mā<sup>1</sup> and a Grade III rhyme -jaa, reconstructed with a medial -j- in 藏<sup>3562</sup> mjaa<sup>1</sup>, could be explained by a harmonizing process mirroring the Grade III rhyme of 藏<sup>5721</sup> zji<sup>1</sup>.

- (8) a. Non-compressed: tensing  
 \*S[ũ/ə].mak<sup>1</sup> > 藏<sup>4820</sup> mā<sup>1</sup> “son-in-law”  
 b. Compressed by compounding with 藏<sup>5721</sup> zji<sup>1</sup>: lengthening  
 \*zji<sup>1</sup>-S[ũ/ə].mak<sup>1</sup> > \*zji<sup>1</sup>-Smjak<sup>1</sup> > 藏<sup>5721</sup> 藏<sup>3562</sup> zji<sup>1</sup>-mjaa<sup>1</sup> “husband and wife, couple”

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



parties. For instance, in Bragbar Situ, the term *koə-təa-jâ* “brothers and sisters, siblings” is derived from *təetəe* “younger siblings” and *a-jâ* “elder siblings” (Zhang 2020: 218, see also Section 5.1.1 for examples in Siyuewu 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, *kendze-nmā-nma*, is derived solely from *tē-nma* “husband”, whereas the other, *ke’dze-rjā-rjev*, is based only on *tē-rjev* “wife” (Sun 1997), which corresponds exactly to the Tangut case.

### 3.3. 𐞧𐞦<sup>5525</sup> 𐞧𐞦<sup>1241</sup> *zji<sup>1</sup>-lji<sup>2</sup>* “children, baby”

The compound, 𐞧𐞦<sup>5525</sup> 𐞧𐞦<sup>1241</sup> *zji<sup>1</sup>-lji<sup>2</sup>* “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, 𐞧𐞦<sup>5525</sup> 𐞧𐞦<sup>1241</sup> *zji<sup>1</sup>-lji<sup>2</sup>* can refer both to the collective concept of “children” and to an individual entity, such as a “small child”. For instance, in (10), 𐞧𐞦<sup>5525</sup> 𐞧𐞦<sup>1241</sup> *zji<sup>1</sup>-lji<sup>2</sup>* is followed by the singular indefinite marker 𐞧𐞦<sup>0448</sup> *gji<sup>2</sup>*.

- (9) 𐞧𐞦<sup>4971</sup> 𐞧𐞦<sup>3798</sup> 𐞧𐞦<sup>0089</sup> 𐞧𐞦<sup>5525</sup> 𐞧𐞦<sup>1241</sup> 𐞧𐞦<sup>0497</sup> 𐞧𐞦<sup>0010</sup> 𐞧𐞦<sup>5354</sup> = 𐞧𐞦<sup>1139</sup> 𐞧𐞦<sup>3320</sup> 𐞧𐞦<sup>3202</sup>  
*ejwi<sup>1</sup> tsəj<sup>1</sup> təhja<sup>1</sup> zji<sup>1</sup>.lji<sup>2</sup> newr<sup>2</sup> zji<sup>2</sup> thji<sup>2</sup> = jji<sup>1</sup> yiew<sup>1</sup>.rji<sup>2</sup>*  
 age small LNK children many all DEM=GEN knowledge  
 𐞧𐞦<sup>2699</sup>  
*nwə<sup>1</sup>*  
 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) 𐞧𐞦<sup>2019</sup> 𐞧𐞦<sup>4861</sup> 𐞧𐞦<sup>2862</sup> 𐞧𐞦<sup>4889</sup> 𐞧𐞦<sup>1794</sup> = 𐞧𐞦<sup>4884</sup> = 𐞧𐞦<sup>1139</sup> 𐞧𐞦<sup>5525</sup> 𐞧𐞦<sup>1241</sup> 𐞧𐞦<sup>0448</sup>  
*thja<sup>1</sup> zjo<sup>2</sup> nji<sup>1</sup> dzjwi<sup>1</sup>.o<sup>1</sup> = nji<sup>2</sup> = jji<sup>1</sup> zji<sup>1</sup>.lji<sup>2</sup> gji<sup>2</sup>*  
 DEM time home neighbour=PL=GEN child one  
 𐞧𐞦<sup>3155</sup> 𐞧𐞦<sup>2921</sup> - 𐞧𐞦<sup>4469</sup>  
*tji<sup>2</sup> -wjji<sup>2</sup> -əji<sup>2</sup>*  
 to.hold-to.hold-to.go<sub>[m]</sub>

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 𐞧𐞦<sup>5525</sup> 𐞧𐞦<sup>1241</sup> *zji<sup>1</sup>-lji<sup>2</sup>*, 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 *yī-gè guān.zhòng* (one-CLF audience) “a spectator”.

Another point worth noting is the etymology of the two components of 𐞧𐞦<sup>5525</sup> 𐞧𐞦<sup>1241</sup> *zji<sup>1</sup>-lji<sup>2</sup>*. The first component is the bound state of 𐞧𐞦<sup>1085</sup> *zji<sup>1</sup>* “male, offspring, son”, while the second component 𐞧𐞦<sup>1241</sup> *lji<sup>2</sup>* is likely a diminutive suffix, as found in Siyuewu Khroskyabs *zî-lo* “son (hypocoristic)” and *mā-lo* “darling (addressing younger generations)”. The meaning “small” in 𐞧𐞦<sup>5525</sup> 𐞧𐞦<sup>1241</sup> *zji<sup>1</sup>-lji<sup>2</sup>* 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 𐞧𐞦<sup>5525</sup> 𐞧𐞦<sup>1241</sup> *zji<sup>1</sup>-lji<sup>2</sup>* indicating “small child, infant” has been transferred to the character 𐞧𐞦<sup>5525</sup> *zji<sup>1</sup>*. This character then contrasts semantically with its base 𐞧𐞦<sup>1085</sup> *zji<sup>1</sup>*, which is used as a kinship term meaning

“son, offspring”.<sup>7</sup> Additionally, 𐞑𐞑<sup>5525</sup> *zji<sup>1</sup>* continues to be used in later compounding mechanisms, such as in 𐞑𐞑<sup>5619</sup> 𐞑𐞑<sup>5525</sup> *mji<sup>2</sup>-zji<sup>1</sup>* 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”	𐞑𐞑 <sup>2447</sup> 𐞑𐞑 <sup>0605</sup> <i>ljo<sup>2</sup>-tjɔ<sup>2</sup></i>	*Cə-tjok <sup>2</sup> -S-tjok <sup>2</sup>
“worm”	𐞑𐞑 <sup>1888</sup> 𐞑𐞑 <sup>1304</sup> <i>bə<sup>2</sup>-lu<sup>1</sup></i>	* <sup>n</sup> bə <sup>2</sup> -S-lu <sup>1</sup>
“brothers”	𐞑𐞑 <sup>0012</sup> 𐞑𐞑 <sup>5873</sup> <i>bju<sup>1</sup>-ku<sup>2</sup></i>	* <sup>n</sup> bju <sup>1</sup> -S-ku <sup>2</sup>
“snake”	𐞑𐞑 <sup>0111</sup> 𐞑𐞑 <sup>0047</sup> <i>lji<sup>1</sup>-ljwu<sup>1</sup></i>	*lji <sup>1</sup> -S-ljw[o] <sup>1</sup>

4.1. 𐞑𐞑<sup>2447</sup> 𐞑𐞑<sup>0605</sup> *ljo<sup>2</sup>-tjɔ<sup>2</sup>* “brothers”

There is general agreement that the two characters 𐞑𐞑<sup>2447</sup> *ljo<sup>2</sup>* and 𐞑𐞑<sup>0605</sup> *tjɔ<sup>2</sup>* 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 𐞑𐞑<sup>2447</sup> *ljo<sup>2</sup>* as a term for the brothers of a male speaker, while 𐞑𐞑<sup>0605</sup> *tjɔ<sup>2</sup>* is considered a collective term, meaning “brothers”. Jacques (2012), among others, suggests that 𐞑𐞑<sup>2447</sup> *ljo<sup>2</sup>* and 𐞑𐞑<sup>0605</sup> *tjɔ<sup>2</sup>* 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 𐞑𐞑<sup>2447</sup> *ljo<sup>2</sup>* and 𐞑𐞑<sup>0605</sup> *tjɔ<sup>2</sup>* does not pertain to relative age. As illustrated in (12) and (13), 𐞑𐞑<sup>2447</sup> *ljo<sup>2</sup>* can refer to both the younger or elder brothers of a man.

- (12) 𐞑𐞑<sup>4962</sup> 𐞑𐞑<sup>3262</sup> 𐞑𐞑<sup>5297</sup> 𐞑𐞑<sup>3583</sup> 𐞑𐞑<sup>0707</sup> 𐞑𐞑<sup>3738</sup> = 𐞑𐞑<sup>1139</sup> 𐞑𐞑<sup>2447</sup> 𐞑𐞑<sup>3798</sup> 𐞑𐞑<sup>0508</sup>  
*we<sup>2</sup>.khow<sup>1</sup>.ioow<sup>1</sup>* *tja<sup>1</sup>* *tejiw<sup>1</sup>.kow<sup>1</sup>=jji<sup>1</sup>* *ljo<sup>2</sup>* *tsaj<sup>1</sup>* *jwu<sup>2</sup>*  
PN:Weikangshu TOP PN:Zhougong=GEN brother<sup>ms</sup> little be  
“Shu Feng of Kang is the younger brother of the King Wu of Zhou.” (*The Twelve Kingdoms*, Juanzhong 17–13)

<sup>7</sup> Note that cognates of 𐞑𐞑<sup>1085</sup> *zji<sup>1</sup>* “son” (Pre-Tangut \*zja) in modern Gyalrongic, such as *tə-tsa* in Cogtse, *tə-ziẽ* in Bragbar, can express multiple meanings of “son, offspring” (as a kinship term), “male” and “little child”.



- (13) 𡗗<sup>5646</sup>.席<sup>3830</sup> 𡗗<sup>2447</sup> 𡗗<sup>2893</sup> 𡗗<sup>5574</sup>.𡗗<sup>5925</sup> 𡗗<sup>5646</sup>.席<sup>3830</sup>=𡗗<sup>1139</sup> 𡗗<sup>4069</sup>  
 tchjiw<sup>2</sup>.nji<sup>2</sup> ljo<sup>2</sup> khwej<sup>2</sup> wji<sup>2</sup>.tsa<sup>1</sup> tchjiw<sup>2</sup>.nji<sup>2</sup>=jji<sup>1</sup> pjwɪr<sup>1</sup>  
 PN:Zhouwang brother<sup>ms</sup> elder PN:Weizi PN:Zhouwang=ACC to.advise  
 𡗗<sup>1918</sup>-𡗗<sup>3575</sup> 𡗗<sup>2484</sup> 𡗗<sup>2590</sup>-𡗗<sup>2474</sup>  
 mji<sup>1</sup>-nji<sup>2</sup> nioow<sup>1</sup> wji<sup>2</sup>-rar<sup>2</sup>  
 NEG-to.listen[A] LNK DIR1.PFV-to.go  
 紂王兄微子因紂王拒諫而出奔 (Shi et al. 1993: 48)  
 “King Zhou’s brother Weizi fled because King Zhou refused to accept his advice.” (*The Grove of Classification* 03.15A.2)

In most cases, 𡗗<sup>0605</sup> tjo<sup>2</sup> is used as a bound morpheme. The two characters 𡗗<sup>2447</sup> 𡗗<sup>0605</sup> ljo<sup>2</sup>-tjo<sup>2</sup> appear together as a compound, representing the collective concept of “brothers”, as in examples (14) and (15).

- (14) 𡗗<sup>5981</sup>-𡗗<sup>3305</sup> 𡗗<sup>0028</sup> 𡗗<sup>4906</sup> 𡗗<sup>1274</sup> 𡗗<sup>5049</sup>.𡗗<sup>0154</sup> 𡗗<sup>0092</sup>.𡗗<sup>4893</sup>  
 a<sup>0</sup>-kjiw<sup>1</sup> lhjuu<sup>1</sup> gjwi<sup>2</sup> wo<sup>2</sup> wja<sup>1</sup>.o<sup>1</sup> mja<sup>1</sup>.wji<sup>1</sup>  
 one-year mourning to.wear[A] should grandfather grandmother  
 𡗗<sup>2447</sup>.𡗗<sup>0605</sup> 𡗗<sup>5305</sup>.𡗗<sup>0004</sup> 𡗗<sup>1561</sup> 𡗗<sup>1543</sup>  
 ljo<sup>2</sup>.tjo<sup>2</sup> wji<sup>2</sup>.la<sup>2</sup> nji<sup>2</sup> mjo<sup>1</sup>  
 brother<sup>ms</sup> paternal.uncle.and.his.wife brother’s.child<sup>ms</sup> non.adopted  
 “One should wear mourning clothes for one year to one’s grandfather, grandmother, brothers (for a male speaker), paternal uncle or his wife, and to one’s brother’s non-adopted son (for a male).” (Tiansheng Code, Jacques 2012: 238)
- (15) 𡗗<sup>3317</sup> 𡗗<sup>0092</sup> 𡗗<sup>0187</sup> 𡗗<sup>0448</sup> 𡗗<sup>3099</sup> 𡗗<sup>1801</sup> 𡗗<sup>2365</sup> 𡗗<sup>1526</sup>=𡗗<sup>3818</sup>  
 lew<sup>1</sup> mja<sup>1</sup> nar<sup>2</sup> gjt<sup>2</sup> dzjii<sup>1</sup> dzjij<sup>2</sup> pha<sup>1</sup> tshji<sup>2</sup>=mjiir<sup>2</sup>  
 only mother old one to.live remain other look.after=NMLZ:S/A  
 𡗗<sup>2447</sup>.𡗗<sup>0605</sup> 𡗗<sup>1602</sup> 𡗗<sup>2194</sup>  
 ljo<sup>2</sup>.tjo<sup>2</sup> ɲowr<sup>2</sup> mjiij<sup>1</sup>  
 brothers<sup>ms</sup> all not.exist  
 “My old mother lives alone, and I have no brothers to look after her.” (*Filial Piety*, Jacques 2007: 14)

The compounding morphology also accounts for the two phonological alternations between the terms 𡗗<sup>2447</sup> ljo<sup>2</sup> and 𡗗<sup>0605</sup> tjo<sup>2</sup>: (i) initial lenition alternation between *l*- and *t*-, and (ii) the tense vowel alternation between *-jo* and *-jə*. According to the transphonologization rule, these alternations are due to the presence of the compound linker \*-S-.

- (16) a. Leniting and dropping (see 1d)  
 \*Cə-tjok<sup>2</sup> > \*Cə-ljok<sup>2</sup> > ljo<sup>2</sup> 𡗗<sup>2447</sup>  
 b. Non-leniting (see 1c)  
 \*-S-tjok<sup>2</sup> > -tjo<sup>2</sup> 𡗗<sup>0605</sup>

The sound changes in (16) suggest that 𡗗<sup>0605</sup> tjo<sup>2</sup> and 𡗗<sup>2447</sup> ljo<sup>2</sup> share the same stem \*-tjok<sup>2</sup> “brother (male speaker)” in Pre-Tangut. The lenition observed in 𡗗<sup>2447</sup> ljo<sup>2</sup> is due to the loss of a presyllable (the fourth type shown in (1d) in Section 2), for instance, the possessive prefix \*tə- still present in East Gyalrongic. Conversely, the tense vowel in 𡗗<sup>0605</sup> tjo<sup>2</sup> results from the original \*-S- blocking lenition (see the third type (1c) in Section 2).

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

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 <sup>ms</sup> ”	𐞧𐞮 <sup>2447</sup> ljo <sup>2</sup>	*Cə-tjok <sup>2</sup>	ri	dóy	tx-xtry	tə-ktiák
“Brother <sup>ms</sup> ”	𐞧𐞮 <sup>0605</sup> tjɔ <sup>2</sup>	*-S-tjok <sup>2</sup>				
“Sister <sup>ms</sup> ”	𐞧𐞮 <sup>0549</sup> niɔ <sup>1</sup>	*S.njVm <sup>1</sup>	sno	snám	tx-snam	tə-snám
“Brother <sup>ws</sup> ”	𐞧𐞮 <sup>0355</sup> mju <sup>1</sup>	*mjo <sup>1</sup>		mô	tx-wɹmu	
“Sister <sup>ws</sup> ”	𐞧𐞮 <sup>3361</sup> kiej <sup>1</sup>	*S.kej <sup>1</sup>	sq <sup>h</sup> e	sqí	tx-sq <sup>h</sup> aj	

Second, similar initial lenition alternations observed in Tangut 𐞧𐞮<sup>2447</sup> ljo<sup>2</sup> and 𐞧𐞮<sup>0605</sup> tjɔ<sup>2</sup> 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 𐞧𐞮<sup>0605</sup> tjɔ<sup>2</sup>, 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 <sup>ms</sup> ”	*Cə-tjok <sup>2</sup>	ri	ri	dóy
“Brother <sup>ms</sup> ” (col.)	*Cə-tjok <sup>2</sup> -S-tjok <sup>2</sup>	rmæ-s-ti	rmɛ-s-ti	rmæ-s-təy

4.2. 𐞧𐞮<sup>1888</sup> 𐞧𐞮<sup>1304</sup> bə<sup>2</sup>-lɿ<sup>1</sup> “worms”

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

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

(17) \*<sup>n</sup>bə<sup>2</sup>-S-lu<sup>1</sup> > bə<sup>2</sup>-lu<sup>1</sup> 𧈧<sup>1888</sup> 𧈧<sup>1304</sup>

The compound 𧈧<sup>1888</sup> 𧈧<sup>1304</sup> bə<sup>2</sup>-lu<sup>1</sup> “worms” and 𧈧<sup>2447</sup> 𧈧<sup>0605</sup> ljo<sup>2</sup>-tjə<sup>2</sup> “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<sup>1888</sup> 𧈧<sup>1304</sup> bə<sup>2</sup>-lu<sup>1</sup> “worms” and the unbound root 𧈧<sup>1888</sup> bə<sup>2</sup> “worm” can be observed in textual examples. The compound 𧈧<sup>1888</sup> 𧈧<sup>1304</sup> bə<sup>2</sup>-lu<sup>1</sup> involves a collective reading (e.g. 18), whereas 𧈧<sup>1888</sup> bə<sup>2</sup> denotes singular concepts, such as a particular type of insect such as locusts in (19a) and silkworm in (19b).

(18) 𧈧<sup>1888</sup> 𧈧<sup>1304</sup> bə<sup>2</sup>-lu<sup>1</sup> “worms”

- a. 𧈧<sup>0289</sup> 𧈧<sup>3266</sup> 𧈧<sup>3273</sup> = 𧈧<sup>1136</sup> 𧈧<sup>1888</sup> 𧈧<sup>1304</sup> 𧈧<sup>4342</sup> 𧈧<sup>2724</sup>  
 we<sup>2</sup>.dzju<sup>2</sup> yar<sup>2</sup>=gu<sup>2</sup> bə<sup>2</sup>.lu<sup>1</sup> dja<sup>2</sup> -təhju<sup>1</sup>  
 Prefect chest=LOC:inside worms DIR1.IPFV-to.have<sub>[A]</sub>  
 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. 𧈧<sup>2019</sup> 𧈧<sup>0864</sup> 𧈧<sup>2436</sup> = 𧈧<sup>1139</sup> 𧈧<sup>2262</sup> 𧈧<sup>5134</sup> 𧈧<sup>1888</sup> 𧈧<sup>1304</sup> = 𧈧<sup>0724</sup>  
 thja<sup>1</sup> dew<sup>2</sup>.mjaa<sup>1</sup>=jjj<sup>1</sup> dzjwow<sup>1</sup>.we<sup>1</sup> bə<sup>2</sup>.lu<sup>1</sup>=nji<sup>2</sup>  
 DEM fruit=ACC birds worms=PL  
 𧈧<sup>3527</sup> 𧈧<sup>4517</sup> 𧈧<sup>0734</sup>  
 mja<sup>1</sup>-dzji<sup>1</sup>-mo<sup>2</sup>  
 IRR-to.eat<sub>[A]</sub>-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<sup>1888</sup> bə<sup>2</sup> “insect”

- a. 𧈧<sup>5354</sup> 𧈧<sup>1888</sup> 𧈧<sup>5688</sup> 𧈧<sup>2484</sup> 𧈧<sup>1918</sup> 𧈧<sup>5754</sup> 𧈧<sup>4884</sup>  
 thji<sup>2</sup> bə<sup>2</sup> wa<sup>2</sup> niow<sup>1</sup> mji<sup>1</sup>-lju<sup>2</sup>-nji<sup>2</sup>  
 DEM insect what LNK:cause NEG-to.catch<sub>[A]</sub>-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. 𧈧<sup>4274</sup> 𧈧<sup>3668</sup> 𧈧<sup>1888</sup> 𧈧<sup>3042</sup> 𧈧<sup>5364</sup>  
 sow<sup>1</sup> lji<sup>1</sup> bə<sup>2</sup> jur<sup>1</sup> lew<sup>2</sup>  
 mulberry to.plant silkworm to.feed woollen.material  
 𧈧<sup>3872</sup> 𧈧<sup>1600</sup> 𧈧<sup>0630</sup> 𧈧<sup>5880</sup> 𧈧<sup>4869</sup> 𧈧<sup>0154</sup>  
 kjir<sup>1</sup> thu<sup>1</sup> la<sup>1</sup> ŋwu<sup>2</sup> kji<sup>1</sup>.o<sup>1</sup>  
 woollen.material to.weave to.weave LNK: INSTR father.in.law  
 𧈧<sup>3986</sup> 𧈧<sup>4893</sup> = 𧈧<sup>1139</sup> 𧈧<sup>0105</sup> 𧈧<sup>4887</sup> 𧈧<sup>5113</sup>  
 nji<sup>1</sup>.wji<sup>1</sup>=jjj<sup>1</sup> kju<sup>1</sup>-tshwew<sup>1</sup>-wji<sup>1</sup>  
 mother.in.law= ACC to.support-to.serve-to.do<sub>[A]</sub>  
 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 𧈧<sup>1304</sup> lu<sup>1</sup> 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,<sup>9</sup> the tense vowel (< \*S-) in Tangut 𐰚<sup>1304</sup> *lu*<sup>1</sup> 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.<sup>10</sup>

**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
* <sup>m</sup> bə <sup>2</sup> -S-lu <sup>1</sup>	bə.zo	bə.jo	qa-juu	kə-lú

**4.3. 𐰚<sup>0012</sup>𐰚<sup>5873</sup> *bju*<sup>1</sup>-*ku*<sup>2</sup> “brothers”**

The compound 𐰚<sup>0012</sup>𐰚<sup>5873</sup> *bju*<sup>1</sup>-*ku*<sup>2</sup> “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, 𐰚<sup>0012</sup> *bju*<sup>1</sup>, is likely related to the first syllable of Tibetan བུ་སྤྱོད་ *bu.s-pun* “brothers” (Zhang et al. 2010: 1830).<sup>11</sup> Note that the Tibetan form contains a collective prefix *s-* in the second component *spun* (see Section 5.1.2).

The second component 𐰚<sup>5873</sup> *ku*<sup>2</sup> 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).<sup>12</sup> The semantic discrepancy is similar to the case of སུ་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 𐰚<sup>0012</sup>𐰚<sup>5873</sup> *bju*<sup>1</sup>-*ku*<sup>2</sup> “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. 𐰚<sup>0111</sup>𐰚<sup>0047</sup> *lji*<sup>1</sup>-*ljwu*<sup>1</sup> “snake”**

The compound 𐰚<sup>0111</sup>𐰚<sup>0047</sup> *lji*<sup>1</sup>-*ljwu*<sup>1</sup> in Tangut typically signifies “snake”, as evidenced in (20), with no instances of its components being used independently. This term likely

<sup>9</sup> 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”: 𐰚<sup>3485</sup> *la*<sup>1</sup> 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.

<sup>10</sup> One of the reviewers pointed out that cognate forms in Rma, Ronghong *bal*, Longxi *bù lò*, and Mianchi *bò lò*, all meaning “worm”, also lack an -S- element. This evidence suggests that the tense vowel in the Tangut form 𐰚<sup>1888</sup>𐰚<sup>1304</sup> *bə<sup>2</sup>-lu<sup>1</sup>* “worms” is a Tangut internal issue.

<sup>11</sup> The Tangut rhyme -u corresponds regularly to Tibetan -u (Jacques 2014: 64–75), as in 𐰚<sup>4614</sup> *nju*<sup>2</sup> “to milk” :: བུ་ *nu.ma* “breasts” and in 𐰚<sup>3388</sup> *ŋwu*<sup>2</sup> “to cry, weep” :: ཉུ་ *ngu* “to cry, weep”.

<sup>12</sup> Tangut 𐰚<sup>0597</sup> *ŋji*<sup>1</sup> “maternal uncle” is also a potential cognate. The lenition alternation between 𐰚<sup>5873</sup> *ku*<sup>2</sup> “brothers” and 𐰚<sup>0597</sup> *ŋji*<sup>1</sup> “maternal uncle” can be explained by the sound laws in Section 2. Specifically, the character 𐰚<sup>5873</sup> *ku*<sup>2</sup> belongs to the non-leniting type (see 1c) and 𐰚<sup>0597</sup> *ŋji*<sup>1</sup> 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) 𐄎<sup>0111</sup>.𐄎<sup>0047</sup> 𐄎<sup>3912</sup> 𐄎<sup>4762</sup> 𐄎<sup>4246</sup>=𐄎<sup>2983</sup> 𐄎<sup>3491</sup>  
*ljî<sup>1</sup>.ljwu<sup>1</sup> bji<sup>2</sup> tɕhji<sup>1</sup> lɕejr<sup>2</sup>=u<sup>2</sup> rjur<sup>1</sup>*  
 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,<sup>13</sup> it is plausible to hypothesize that the first component 𐄎<sup>0111</sup> *ljî<sup>1</sup>* serves as the reduplicant, while the second component 𐄎<sup>0047</sup> *ljwu<sup>1</sup>* 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-ɣə-vzɪ* “maternal uncle and his sister’s children” is composed of *ɣə-* (the bound state of *ɣô* “maternal uncle”) and a bound root *†vzɪ* “sister’s children (for a male speaker)”. Both components are inherited Proto-Gyalrongic kinship terms. The Siyuewu *ɣô* “maternal uncle” reflects the Gyalrongic etymon for maternal uncle, as in Tangut 𐄎<sup>0597</sup> *ɣji<sup>1</sup>* (Pre-Tangut *\*CV-kji<sup>1</sup>*) “maternal uncle” and *a-kû* “maternal uncle” in

<sup>13</sup> While the vowel alternation between *i :: u* is also attested with 𐄎<sup>0680</sup>𐄎<sup>0366</sup> *tɕhji<sup>2</sup>-tɕhjuu<sup>2</sup>* “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  $\dagger v z i$  is unattested as an unbound morpheme, it is related to Tangut 𐰇𐰺<sup>2134</sup>  $z j w i^1$  (Pre-Tangut \*S-pǎ.tsa) “cross nephew, child of different-sex siblings”.<sup>14</sup>

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

Gloss	Form	Compound 1	Compound 2
“maternal uncle and his nephew”	<i>s-γə-vzɪ</i>	<i>əyô</i> “maternal uncle”	$\dagger v z i$ “nephew”
“maternal aunt and her nephew”	<i>s-la-vdɪ</i>	<i>lalâ</i> “maternal aunt”	$\dagger v d i$ “nephew”
“grandfather and grandchild”	<i>s-və-vlâ</i>	<i>vâvə</i> “grandfather”	<i>vlâ</i> “grandson”
“Goatherd’s family” (house name)	$\dagger t s^h â-jəm$	$t s^h â d$ “goat”	<i>jəm</i> “house”

Note: the † indicates forms without unbound attestation.

The second collective *s-la-vdɪ* “maternal aunt and her sister’s children” is built upon *la-* (the bound state of *lalâ* “maternal aunt”) and  $\dagger v d i$  “nephew, sister’s children (for a female speaker)”. The unbound root  $\dagger v d i$  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-vlâ* “grandfather and grandchild”, the first component *və-* represents the Proto-Gyalrongic term for “grandfather”, preserved in Bragbar *ta-wû* and Japhug *tx-wuu*, and also occurs as the second component of *vâ-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  $\dagger t s^h â-jəm$  (COL-goat-house), in which the initial consonant  $\dagger t$ - is a conditioned variant of the collective *s-* prefix (see Lai 2016 on the Siyuewu *s-* allomorphy). The  $\dagger t s^h â-jəm$  family are goatherds, and the house name reflects the close relation between goats and their owners.<sup>15</sup> The  $\dagger t s^h â-jəm$  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

<sup>14</sup> This interpretation follows Kepping (1991), as 𐰇𐰺<sup>2134</sup>  $z j w i^1$  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.

<sup>15</sup> Similar derivations are found in Japhug (East Gyalrongic), although using a non-cognate prefix *kɿndzi-*, such as *kɿndzi-ts<sup>h</sup>rt* (COL-goat) “goat and its owners”, *kɿndzi-mbro* (COL-horse) “horseman and his horse” and *kɿndzi-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).<sup>16</sup> 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་སྒྲ *pha-spad* “father and children”

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

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

(22) Collective circumfix *s-Σ-n*

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

Except for སྒྲ་པོ *skud-po* “brother-in-law”, which bears a non-transparent semantic relationship with the base form ཁྱ *khu* “paternal uncle”,<sup>17</sup> 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>rm̩-s-təŋ</i>	<i>rm̩</i> “man, others”	<i>dóŋ</i> “brother”
Co-ordinative	adult woman	<i>və-s-mé</i>	<i>və</i> “adult woman”	<i>ʔmé</i> “woman”
Co-ordinative	recently	<i>fs̩-s-kʰə</i>	<i>fs̩</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>pʰəŋ-s-jəm</i>	<i>pʰəŋ</i> “pig”	<i>jəm</i> “house”

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

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

<sup>17</sup> 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 སྒྲ་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.<sup>18</sup> Its first component *və-*, though unattested as a free morpheme, is related to the second component in *gə-və* “wife” (further related to Tangut 𐽄𐽀<sup>2455</sup> 𐽄𐽀<sup>2129</sup> *gji<sup>2</sup>-bjij<sup>2</sup>* “wife”, see Lai et al. 2020). The second component -s-mé reflects the Gyalrongic etymon for “woman, girl”, as in Japhug *tui-me* and Bragbar Situ *tə-mí*.

The compound *fsê-s-k<sup>h</sup>ə* juxtaposes two antonymous components, *fsê* “to be early” and *k<sup>h</sup>ə* “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.<sup>19</sup> An example of a left-headed compound concerns *lə-s-p<sup>h</sup>rəm* “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<sup>h</sup>rəm* “to be white” and *nəŋ* “to be black”.

The term *p<sup>h</sup>ay-s-jəm* “pig pen” is a case of right-headed compound, in which the linker -s- connects the modifier *p<sup>h</sup>ay* “pig” and the head *jəm* “house”.<sup>20</sup>

Traces of the compound linker \*-S- are also found in Horpa languages, as exemplified by Geshiza Horpa in Table 9.<sup>21</sup>

**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<sup>h</sup>ə-s-q<sup>h</sup>a</i>	<i>s<sup>h</sup>ə-p<sup>h</sup>o</i> “tree”	<i>†q<sup>h</sup>a</i> “root”
?	“clothes”	<i>ts<sup>h</sup>ə-z-gə</i>	<i>†ts<sup>h</sup>ə</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<sup>h</sup>ə-s-q<sup>h</sup>a* “tree roots” is right-headed, in which the linker -s- connects the modifier *s<sup>h</sup>ə-* “tree, wood (bound state)” and the head *†q<sup>h</sup>a* “root”, a bound root. The second component is related to Guanyinqiao Khroskyabs *qé* “root”, *ʃni-q<sup>h</sup>é* “tongue root”, and Japhug Gyalrong *u-qa* “root”). These Gyalrongic cognates suggest a

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

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

<sup>20</sup> 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 𐽄𐽀<sup>0960</sup> *mji<sup>1</sup>* “woman, girl” (Pre-Tangut \*S-mji<sup>1</sup>), Geshiza *s-me*, etc. (Lai et al. 2020: 177). These forms correspond to the Gyalrongic etymon for “woman” or “girl”, as seen in Japhug *tui-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).

<sup>21</sup> The three Geshiza compounds in Table 9 have direct cognates in Bawang, *rme-s-ti* “brothers”, *ts<sup>h</sup>ə-z-gwə* “clothes”, *s<sup>h</sup>ə-s-q<sup>h</sup>a* “tree roots” (data from Yang 2021).

proto-form for “root” without a sibilant pre-initial. Thereby the presence of the linker *-s-* in *s<sup>h</sup>ə-s-q<sup>h</sup>a* “tree roots” suggests the productivity of the compound linker *-s-* after the branch-off of Horpa.

The third compound *ts<sup>h</sup>ə-z-gə* “clothes”, in which the linker *-s-* is assimilated to *-z-*, is currently only found in Geshiza and Bawang (*ts<sup>h</sup>ə-z-gwə* “clothes”). Parallel compounds with cognate roots but lacking a sibilant linker morpheme exist within Gyalrongic, such as Khang,gsar Stau (Horpa) *tsə-gə*, Siyewu Khroskyabs *ts<sup>h</sup>ə-gí*, Tangut 𐽳𐽲<sup>5610</sup>𐽳𐽲<sup>5598</sup> *tshji<sup>1</sup>-gjiw<sup>2</sup>* “clothes” (Li 2012: 667, 669), and beyond, Pengbuxi Minyag *tse-ŋgə* (Gao 2016), Guiqiong *ts<sup>h</sup>ə<sup>33</sup>-wə<sup>53</sup>* (Zàngmiǎnyǔ Yǔyīn hé Cíhuì Biānxiězǔ 1992), all meaning “clothes”. In the Geshiza form *ts<sup>h</sup>ə-z-gə* “clothes”, while the second component *-gə* is related to the verb “to wear”, the first component *ts<sup>h</sup>ə-* 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 𐽳𐽲<sup>1888</sup>𐽳𐽲<sup>1304</sup> *bə<sup>2</sup>-lu<sup>1</sup>* “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་སྐྱ མ་སྐྱ *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<sup>h</sup>ə-z-gə* “clothes” might suggest another possibility. If we consider *ts<sup>h</sup>ə* cognate with Ersu *ts<sup>h</sup>a<sup>55</sup>* “classifier for clothes” (Zàngmiǎnyǔ Yǔyīn hé Cíhuì Biānxiězǔ 1992), then Geshiza *ts<sup>h</sup>ə-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).<sup>22</sup> This oblique nominalizer is no longer productive in West Gyalrongic but leaves traces in Wobzi Khroskyabs *s-p<sup>h</sup>əm* “lid” (derived from *p<sup>h</sup>əm* “to cover”) (Lai 2017: 158, 511), as well as in the nominalizing tense vowel in Tangut, e.g. 𐽳𐽲<sup>5205</sup> *ya<sup>1</sup>* “sword, weapon” (derived from 𐽳𐽲<sup>5653</sup> *ya<sup>1</sup>* “to butcher, chop”) (Jacques 2014: 256).<sup>23</sup> 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

<sup>22</sup> The internal syntax of *ts<sup>h</sup>ə-z-gə* “clothes” could be similar to Chinese 掛飾 *guà-shì* “hanging ornament”.

<sup>23</sup> 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-Sklódowska-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. "Xīxiàyǔ 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 qí qǐyuán 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)* [Studies on the Phonetics of the Tangut Language (Preliminary Results)]. Moscow: Izdatel'stvo vostochnoi literatury.
- Lai, Yunfan. 2016. "Causativisation in Wobzi and other Khroskyabs dialects", *Cahiers de Linguistique – Asie Orientale* 45/22, 148–75.
- Lai, Yunfan. 2017. "Grammaire du khroskyabs de Wobzi", Doctoral dissertation, Université Sorbonne Nouvelle (Paris 3).
- Lai, Yunfan. 2023. "Lenition alternation in West Gyalrongic and its implication for Southeast Asian panchronic phonology", *Diachronica* 40/3, 341–83.
- Lai, Yunfan. 2024. "Mutual predictiveness of sound correspondences hints at reconstruction and language sub-grouping: The case of Gyalrongic preinitials", *Diachronica* (online first).
- Lai, Yunfan, Xun Gong, Jesse P. Gates and Guillaume Jacques. 2020. "Tangut as a West Gyalrongic language", *Folia Linguistica Historica* 41/1, 171–203.
- 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.
- Li, Fanwen. 1997. *Xià-Hàn zìdiǎn* 夏漢字典 (Tangut–Chinese Dictionary). Beijing: China Social Sciences Press.
- Li, Fanwen. 2012. *Jiǎnmíng Xià-Hàn zìdiǎn* 簡明夏漢字典 (Tangut–Chinese Dictionary). Second edition. Beijing: Zhongguo Shehui Kexue Chubanshe.
- Lin, You-Jing. 2016. *Jiāróngyǔ Zhuókèjī 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.
- Meillet, Antoine. 1925. *La méthode comparative en linguistique historique*. Oslo: H. Aschehoug & Co.
- Nagano, Sadako. 1994. "A note on the Tibetan kin terms khu and zhang", *Linguistics of the Tibeto-Burman Area* 17/2, 103–15.
- 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読 (A study of the Tangut language: Reconstruction of the Tangut language and decipherment of Tangut script). Tokyo: Zauhō kankōkai.
- 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.
- Peng, Xiangqian. 2012. *Xīxiàwén "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.
- 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.
- Shi, Jinbo. 2020. *Tangut Language and Manuscripts: An introduction*. Leiden and Boston: Brill.
- 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.
- Solonin, Kirill J. 1995. *Dvenadcat' carstv: Perevod s tangustskogo, 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.
- 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.
- 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.
- 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.
- Wang, Jingru. 1982. "Xīxiàyǔ yīnxì dǎoyán 西夏語音系導言 (Introduction to Tangut phonology)". 民族語文 (*Minority Languages of China*) 2, 1–13.
- Wei, Hao. 2022. "Xīxiàyǔ 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.
- Yang, Chih-Fan. 2021. "Bāwàng Huò'ěr yǔ shí, tǐ, shìzhèng, qíngtài fànchóu de xíngtài jùfǎ 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.
- Yin, Weibin. 2007. *Yèlóng Lāwūróngyǔ Yánjiū* 業隆拉瑪戎語研究 (Study on the 'Jorogs Lavrung language). Beijing: Nationalities Press.

- 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.
- 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.
- 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.
- 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.
- 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.
- 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).

---

**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* 88/1, 103–122. <https://doi.org/10.1017/S0041977X24000673>