Nasal Preinitials in Tangut Phonology

Xun Gong

Abstract
Gong Hwang-cherng proposed that the Tangut language has a distinction between short and long vowels. To date, however, no reliable correlates have been found regarding the actual phonological nature of the distinction. A careful examination of Chinese loanwords in Tangut and Sino-Tangut pronunciation reveals that the “vowel length” distinction should be revised to that of the presence vs. absence of a nasal preinitial. The pair 萌₃₈₀₆ “weed” vs. 墓₂₁₃₈ “tomb,” borrowed respectively from Chinese 蒲 bu and 墓 muH (the latter from a Northwest-type reflex with *mb-), hitherto reconstructed as bu¹ {bu¹} vs. buu² {buu²}, should be revised to bu°¹ vs. mbu°². The reconstructed nasal preinitial not only has a close typological parallel in Modern West Rgyalrongic, but is equally reflected in other sources of evidence, most strikingly Sanskrit transcription and fǎnqiè. The revision solves a large number of problems in the historical phonology of Tangut, though not without raising some new ones, especially in connection with the treatment of Proto-Rgyalrongic preinitials before nasals.

Keywords
Linguistics | Tangut | Northwestern Medieval Chinese | Qiangic languages

* Xun Gong (minus273cn@gmail.com) is currently an Assistant at Institut für Sprachwissenschaft, Universität Wien.

** I wish to thank the anonymous reviewers for their useful comments. This essay is dedicated to the memory of Pān “the sea slug” Léi 潘雷 (1988–2020), a fellow lover of Jerzy Petersburski and Kajio Shinji.
1. CONTEXT

1.1 Paired Columns Of Rhymes And The “Vowel Length” Distinction

Gong Hwang-cherng 龔煌城, in his 1994 article\(^1\), proposed that Tangut\(^2\) has a distinction between short and long vowels. Given that Tangut is written in a non-phonetic writing system, in which, much as in Chinese, a character denotes a syllable-morpheme whose phonetic nature must be deduced through a conjunction of more or less tangential evidence, a hypothesis about Tangut phonology such as the one at hand should be properly understood as consisting of two sub-hypotheses:

- **Categorization and phonemicity**: A partition of the Tangut characters into disjoint categories, such that syllables denoted by characters in one category share some common phonological feature distinguishing them from syllables in other categories.
- **Phonological substance**: An identification of the concrete phonological distinction that underlies the difference between the categories.

In order to properly understand the vowel length hypothesis, we start by examining Gong Hwang-cherng’s reasoning and other potential evidence from the perspectives of both categorization and phonological substance.

In terms of categorization, Gong Hwang-cherng’s hypothesis is rooted in

---

\(^2\) Tangut characters are annotated with their Lǐ number, referring to the numbering system of the second edition of the *Tangut-Chinese Dictionary* (2008). Transcriptions given inside curly braces conform to Gong Hwang-cherng’s reconstruction. Outside curly braces, transcriptions always take into account the uvularization hypothesis—see Xun Gong, “Uvulars and uvularization in Tangut phonology,” *Language and Linguistics* 21, no. 2 (2020): 175–212—and, unless otherwise clear from the context, also the nasal preinitial hypothesis proposed in this essay. A fully annotated example is \[\text{𗎭₁₈₉₂} \text{ourney} \{\text{mjii¹}\} \text{“house.”} \]
his discovery of the “paired sequences” of rhymes. Consider, for example, the group of rhymes from R.8 to R.14. Foreign-language transcription evidence shows that they share, more or less, the same main vowel. In Tangutological parlance, they belong to the same 韻 shè. Moreover, as the same 1994 article demonstrates, there is another dimension within the same shè—namely that of grades (等 děng), represented in Gong Hwang-cherng’s system as {-e}: {-ie}: {-ji}. The “paired sequences” refer to the observation that exactly the same series of rhymes, ordered by grade, seems to exist twice, juxtaposed to each other. Immediately after R.8 {-e}, R.9 {-ie}, and R.10/11 {-ji} come R.12 {-ee}, R.13 {-iee}, and R.14 {-jii}, which are, in almost every respect of foreign-language transcription, virtually equivalent to their respective counterparts in R.8–R.11.

Figure 1: Paired columns of rhymes in the major cycle of Tangut rhymes (R.1–R.60)

Table: Paired columns of rhymes

Note: Column-2 rhymes in boldface; “!!” marks rhymes assigned to column 2 in Gong Hwang-cherng’s system, as reflected in Gong (2003) and all versions of the Tangut-Chinese Dictionary (Li Fànwén 1997, 2008, 2012), but considered in this essay to be column-1 rhymes.

As Figure 1 shows, this pattern of paired sequences is repeated over and over in the native rhyme ordering system. In this essay, I refer to them as paired columns, enumerated as column 1 and column 2. The question of phonemicity naturally comes into play. How do we know, apart from rhyme ordering in native metalinguistic resources, if the distinction between R.1–R.4 and R.5–R.7 is the same as the distinction between R.8–R.11 and R.12–R.14? Gong Hwang-cherng typically addresses such questions through the method he labels phonological alternation⁴, which he did not apply to paired columns. Nevertheless, applying his method to this phenomenon indeed shows a tight connection. For example, 𘙣₀₇₁₆ {śjii¹} is the stem A of a verb “to butcher” whose stem B is 𘙣₄₅₇₁ {śjoo¹}. The stem A is in rhyme R.14, belonging to the second column; the stem B, in rhyme R.55, also belongs to the second column. On the other hand, an alternating verb, such as “to eat,” whose stem A 𗡅₄₅₁₇ {dzji¹} belongs to R.10, a rhyme in the first column, has a stem B 𗠈₄₅₄₇ {dzjo¹} with the rhyme R.53 in the first column too. This shows that the distinction between R.8–R.11 : R.12–R.14 is indeed analogous to the distinction between R.51–R.53 : R.54–R.55, thereby suggesting that they reflect a basic phonemic distinction of the language.

Concerning the phonological substance, Gong Hwang-cherng reconstructs the distinction between column 1 and column 2 as one of vowel length. Column-1 rhymes are reconstructed with short vowels, column-2 with the long vowels. Hence, 𗻍₃₈₀₆ “weed,” having the column-1 rhyme R.1 (1.1), is reconstructed as bu¹¹ {bu¹}, whereas 𗽰₂₁₃₈ “tomb,” which belongs to the column-2 rhyme R.5 (1.5–2.5), is reconstructed as buu²² {buu²}.

Gong Hwang-cherng’s vowel length hypothesis relies on one single observation: “rhymes representing Chinese loanwords [i.e., rhymes with nasalized vowels] have only a short vowel sequence and no corresponding long vowel sequence.”⁵ In this statement, he is referring to the rhyme sequences R.15–R.16 and R.25–R.27, reconstructed in most reconstruction systems with the nasal vowels {ĩ / ĕ} and {ã}. As can be seen in Figure 1, these rhyme sequences are not divided into paired columns. Since these rhyme sequences mostly involved loanwords from Chinese, Gong Hwang-cherng considers that the lack of paired

---


⁵ Gong Hwang-cherng, Xīxià yǔwén yánjiù lùnwénjí, 150.
columns can be imputed to the absence of vowel length distinction in Chinese. Hence, he hypothesizes the paired column distinction as one of vowel length.

This line of argument is weak. To date, in both loanword materials and etymological comparison, the literature has not revealed any external correlates of the paired columns.

Gong Hwang-cherng’s partition of Tangut rhymes into column-1 and column-2 rhymes gained wide acceptance among Tangutologists; his theory that this distinction reflects one of vowel quantity less so. Gong Hwang-cherng himself, in an interview conducted by Jackson T.-S. Sun, stated his lack of certainty concerning the actual value of “vowel length.” Later reconstruction schemes, such as those of Arakawa Shintarō and Marc Miyake, recognize both the validity of paired columns and the tenuity of the vowel quantity theory by marking column-2 rhymes with the semantically vacuous prime symbol -´.

1.2 Chronological Layers in Chinese-to-Tangut Transcription Materials

Four types of material are essential for the reconstruction of Tangut phonology: native dictionaries compiled by Tangut scholars, transcription/loanwords, comparison between Tangut and Burmo-Qiangic languages, especially the closely related modern Rgyalrongic languages, and internal reconstruction based on what Gong Hwang-cherng terms “phonological alternations.”

In this essay, I am drawing in particular on Chinese-to-Tangut transcription materials; in other words, Tangut words borrowed from Chinese as

---

10 See Gong, “Phonological Alternations in Tangut”; etc.
well as Tangut transcription of Chinese. These materials form part of Sino-
Tangutica—namely, the totality of transcriptional and lexical materials
that arose in the language contact between Tangut and Chinese. Apart from
Chinese-to-Tangut materials, which we will examine in some detail, we also
have at our disposal materials in the opposite direction, from Tangut to Chi-
nese, of which the best-known example is the language textbook Pearl in the
Palm (𗼇𘂜𗟲𗿳𗖵𘃎𘇂𗊏 {mji²zar¹ ŋwuu¹dzj ɨ j¹ bju¹pjạ¹gu²njị²}, 番漢合時掌中珠
Fānhàn Héshí Zhǎngzhōngzhū), in which Tangut words and phrases are phoneti-
cally transcribed in Chinese.

Sino-Tangutica has been essential to the reconstruction of Tangut phonol-
ogy from the very beginning of the enterprise. Its familiarity, however, should
not diminish its interest. The Chinese-to-Tangut material is particularly im-
portant for an often overlooked feature: its internal divergence into different
chronological strata, which shed light on sound changes both within Tangut
itself and in the source Chinese varieties.

The majority of Chinese-to-Tangut evidence can be subsumed into one of
two categories:

- I use the term established borrowings to refer to the cases analyzed in
  Gong Hwang-cherng’s groundbreaking article, “Chinese loanwords in the
  Tangut language.” They concern words of Chinese origin that had either
  been assigned a dedicated Tangut character or had otherwise been identi-
  fied as such in native character dictionaries. As Gong has shown, this
corpus, which dates to the mid-11th century CE, already shows a degree of
  internal divergence. In particular, there is an older layer, corresponding
to an older stage of the Chinese language, closer to Early Middle Chinese
and a newer layer, basically resembling the Late Sino-Tangut pronunciation.

- I use the term Late Sino-Tangut pronunciation to refer to the system(s) of
  pronouncing Chinese characters as reflected in proper names and spe-

11 The term Sino-Tangutica designates that which in Chinese is called Xià-Hàn duiyīn cāiliào 夏
漢對音材料. The conventional translation of duiyīn 對音 as transliteration / transcription mis-
characterizes the situation, since we are not, outside of language textbooks like the Pearl in the Hand,
dealing with transcriptions per se. Instead, the sociolinguistic situation much resembles that of English words in Hindi-Urdu or Japanese, where bilingualism, at least in
terms of vocabulary, is prevalent; where almost any Chinese word can be borrowed in Tang-
ut; but where there is nevertheless a heavy adaptation to the target-language phonology.
The ambiguous term Sino-Tangutica better captures this ill-defined middle ground between
transcription, code-switching, and borrowing.

12 Gong, Hwang-cherng, “Xīxiàyǔ zhōng de Hànyǔ jièci” [Chinese Loanwords in the Tangut Lan-
cialist vocabulary of Chinese origin in a wide range of Tangut-language literature, especially legal and technical literature, as well as translations from Chinese originals, mainly from the mid-12th century CE onwards, in conjunction with the Tangut transcription of Chinese in the *Pearl in the Hand*.

In this essay, I mainly rely on Gong Hwang-cherng’s forementioned article (1981c) as my source for established borrowings, and on Gong Hwang-cherng’s article from 1991 — an analysis of proper names and bureaucratic titles of Chinese origin in the *Lèilín* (類林, 統類 {diyi¹ bo¹})— for Late Sino-Tangut pronunciation.

Two major consonant shifts demarcate Mediaeval Héxī河西 Chinese, the dialect(s) of Mediaeval Northwestern Chinese in heavy contact with Tangut, from its precursor, Early Middle Chinese (EMC):

1. Héxī *mb- < EMC *m- 明: in Mediaeval Northwestern Chinese dialects in general, EMC nasals turned, either allophonically or definitely, into pre-nasalized voiced stops;
2. Héxī *ph- < EMC *b- 並: in Mediaeval Héxī Chinese, EMC voiced stops turned into their voiceless aspirated counterparts.

As Gong Hwang-cherng demonstrates in his article “Chinese loanwords in the Tangut language,” two chronological layers can be distinguished within the corpus of established loanwords. In an earlier layer, Tangut initials reflect the original EMC forms; in a later layer, Tangut initials reflect later Héxī reflexes. The Late Sino-Tangut pronunciation, unsurprisingly, also reflects the chrono-
logically later stage. The stages of sound change reflected in the chronological layers of Chinese-to-Tangut evidence are shown in Table 1.

Table 1: Chronological layers of Chinese-to-Tangut evidence.

<table>
<thead>
<tr>
<th>EMC Initial</th>
<th>Established borrowings</th>
<th>Later layer</th>
<th>Late Sino-Tangut</th>
</tr>
</thead>
<tbody>
<tr>
<td>p- 幫</td>
<td>Tangut {p-}</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>跬3802 {pjji²} &lt; 邊pen</td>
<td>跬4196 {pio¹} &lt; 包æw</td>
<td></td>
</tr>
<tr>
<td>ph- 潋</td>
<td>Tangut {ph-}</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>跬4007 {pha¹} &lt; 破phaH</td>
<td>跬2489 {phej²} &lt; 沛phajH</td>
<td></td>
</tr>
<tr>
<td>b- 並</td>
<td>Tangut {b-}</td>
<td></td>
<td>跬0876 {bã¹} &lt; 盤ban</td>
</tr>
<tr>
<td></td>
<td>跬3648 {phiej³} &lt; 跬bcH</td>
<td>跬2667 {phu¹} &lt; 部buwX</td>
<td></td>
</tr>
<tr>
<td>m- 明</td>
<td>Tangut {m-}</td>
<td></td>
<td>跬1130 {mjii²} &lt; 麦mie</td>
</tr>
<tr>
<td></td>
<td>跬4964 {bioo¹} &lt; 跬maew</td>
<td>跬2736 {biaaa²} &lt; 麦mæX</td>
<td></td>
</tr>
</tbody>
</table>

1.3 Scope and Structure of this Essay

Gong Hwang-cherng reconstructed “long vowels,” or, in the noncommittal terms of the present essay, column-2 rhymes, systematically in the major cycle (R.01–R.60). He also reconstructed “long vowels” for several rhymes in the second minor cycle (R.80–R.98) and the totality of the third minor cycle (R.99–R.103). Only the rhymes of the first minor cycle (R.61–R.79) do not show any phenomenon of paired columns. The scope of this essay, however, is restricted to Tangut rhymes in the major cycle.

This choice follows, most of all, from the fact that the Sino-Tangut materials, which lie at the fulcrum of the argument, are found almost exclusively in the major cycle. This disproportionate concentration also applies, to a lesser degree, to the other sources examined in this essay. Preliminary research, moreover, shows that the rhymes assigned to column 2 by Gong Hwang-cherng in the minor cycles do not show the same behavior with respect to transcriptional and etymological data as column 2 rhymes in the major cycle, suggesting that those rhymes can be considered an entirely different phenomenon, if indeed they can be regarded as one single class at all. I consider what Gong Hwang-cherng assigned to column 2 in the second and third minor cycles to be unrelated to the subject of this essay and relegate discussion of the nature of these rhymes to future papers.
After this introductory section, Section 2 discusses the behavior of Tangut paired columns in Sino-Tangut transcription and loanword materials and proposes the hypothesis that in syllables with voiced stop initials, the concrete nature of column-2 rhymes is akin to prenasalized voiced stops. Section 3 shows how this hypothesis is to be generalized across initial types: column-2 rhymes are proposed to indicate a nasal preinitial in Tangut phonology. Before moving on to the conclusion, Section 4 discusses how the nasal preinitial hypothesis interacts with, is supported by, or otherwise improves the treatment of other sources for the reconstruction of Tangut phonology, notably fānqìè evidence in native dictionaries as well as comparative evidence in modern Rgyalrongic languages.

2. PAIRED COLUMNS OF RHYMES IN SINO-TANGUT MATERIALS

2.1 Column-2 Rhymes and Prenasalized Initials in Mediaeval Hexi Chinese

Huáng Kǎn 黃侃 famously said that the essence of philology lies in “uncovering” fāmíng 發明, i.e. of hidden connections between well-known materials, and not “discovering” fāxiàn 發現, i.e. of new materials. In a twist reminiscent of the eccentric, a robust correlate to the paired columns, which has long eluded the search of Tangutologists, reveals itself in an all too familiar place. As this section will demonstrate, the key to understanding the nature of Tangut “vowel length” lies in the Chinese-to-Tangut evidence.

A starting point for examining this question is Gong Hwang-cherng’s article from 1991, which contains a useful table containing all the Chinese syllables transcribed or borrowed into Tangut in the Lèilín, ordered by the Tangut rhymes of the target syllables. Once one examines specifically the column-2 rhymes enlisted to render Chinese words, one could not fail to notice that Tangut syllables there mostly render Chinese syllables with the Middle Chinese nasal initials 明 m-, 泥 n-, 娘 ň-, and 疑 ŋ-. Among the rhymes containing Chinese-to-Tangut syllables in Lèilín, thirteen major-cycle rhymes are reconstructed by Gong Hwang-cherng with “long vowels”—that is, assigned as column-2 rhymes. Among these thirteen rhymes, eleven of them, shown in Table 2, almost exclusively transcribe Chinese syllables with the Middle Chinese nasal initials 明 m-, 泥 n-, 娘 ň-, and 疑 ŋ-.

Among the rhymes containing Chinese-to-Tangut syllables in Lèilín, thirteen major-cycle rhymes are reconstructed by Gong Hwang-cherng with “long vowels”—that is, assigned as column-2 rhymes. Among these thirteen rhymes, eleven of them, shown in Table 2, almost exclusively transcribe Chinese syllables with EMC nasal initials. The remaining two rhymes, R.21 and R.59, exhibit
the diametrically opposite behavior: as shown in Table 3, they exclusively transcribe Chinese syllables with non-nasal initials. Insofar as we understand the column 1/2 distinction as a unitary phonemic distinction, we can safely reject them as column-2 rhymes, given their completely column-1 behavior; their actual nature will be discussed in forthcoming papers.

**Table 2.** Tangut column-2 rhymes that predominantly transcribe Chinese nasal initials in Gong’s *Xīxià yǔwén yǎnjiū lùnwéngjí*, 454–57.

<table>
<thead>
<tr>
<th>Rhyme</th>
<th>GHC reconstruction</th>
<th>MC nasal initials</th>
<th>MC non-nasal initials</th>
</tr>
</thead>
<tbody>
<tr>
<td>R.5 (1.5–2.5)</td>
<td>{-uu}</td>
<td>奴 nu 軋 muX 至 miuw 謀 miuw 母 muwX 母 muwX 穆 miuwk 五 guX 伍 guX 慕 muH 蕃 mak 茂 muwH 謀 miuw 嫣 mu</td>
<td>騎 dzawk 禾 yiuX 余 jia 禾 yiuX 猶 juw</td>
</tr>
<tr>
<td>R.7 (1.7–2.6)</td>
<td>{-jwu}</td>
<td>玉 iiaX 園 iaX 語 iaX 漁 ia 去 jiu 御 iaH 女 iaX</td>
<td>蜀 dzawk 禾 yiuX 余 jia 禾 yiuX 猶 juw</td>
</tr>
<tr>
<td>R.12 (1.22–2.11)</td>
<td>{-ee}</td>
<td>默 mak 墨 mak</td>
<td></td>
</tr>
<tr>
<td>R.14 (1.14–2.12)</td>
<td>{-ii}</td>
<td>儀 iie 毅 iijH 凝 ij 密 mit 慘 mieX 汨 mek 慘 mie</td>
<td></td>
</tr>
<tr>
<td>R.22 (1.22–2.19)</td>
<td>{-aa}</td>
<td>末 mat 熱 jaw 昇 jawH 納 nap</td>
<td></td>
</tr>
<tr>
<td>R.23 (2.20)</td>
<td>{-iaa}</td>
<td>牙 iæ 頭 iæn 懸 maen 起 maex 馬 maex 雅 iæX</td>
<td>晏 ?ænH</td>
</tr>
<tr>
<td>R.24 (1.23–2.21)</td>
<td>{-jaa}</td>
<td>政 iæp 懽 lienX</td>
<td></td>
</tr>
<tr>
<td>R.33 (1.32–2.29)</td>
<td>{-ii}</td>
<td>呂 iip</td>
<td></td>
</tr>
<tr>
<td>R.38 (1.37–2.34)</td>
<td>{-ej}</td>
<td>艾 iajH 肉 nwajH</td>
<td>哀 ?aj</td>
</tr>
<tr>
<td>R.54 (1.52–2.45)</td>
<td>{-oo}</td>
<td>穆 miuwk 摩 ma 芳 manX 茂 muwH</td>
<td></td>
</tr>
<tr>
<td>R.55 (1.53–2.46)</td>
<td>{-iio, -joo}</td>
<td>樂 iæwH 岳 iæw</td>
<td></td>
</tr>
</tbody>
</table>
Table 3. Tangut column-2 rhymes that exclusively transcribe Chinese non-nasal initials in Gong’s Xīxià yǔwén yánjiù lùnwénjí, 454–57.

<table>
<thead>
<tr>
<th>Rhyme</th>
<th>GHC reconstruction</th>
<th>MC nasal initials</th>
<th>MC non-nasal initials</th>
</tr>
</thead>
<tbody>
<tr>
<td>R.21 (1.21–2.18) {−aa}</td>
<td></td>
<td></td>
<td>葛 kat 照 teewH 少 siewH 昭 teew 藥 kawX 瑤 jew 陶 jew 豐 lew 邵 dzewH 紹 dzewX</td>
</tr>
<tr>
<td>R.59 (1.57) {−ioow}</td>
<td></td>
<td></td>
<td>叔 ewk 屬 dzawk 蠱 dzawk 築 tjuwk</td>
</tr>
</tbody>
</table>

The significance of this becomes clear once it is taken into account that the Tangut syllables used to render Chinese nasals have voiced stop initials instead. They reflect the Héxī reflexes of EMC nasals as prenasalized voiced stops:

- Tangut {b-} for Héxī *mb- < EMC m- 明: 𘓣₂₇₃₆ {biaa²} for 马 mæX, 蠻 mæn, 莊 mæwX
- Tangut {d-} for Héxī *nd- < EMC n- 泥: 𘕧₅₂₅₀ {daa²} for 納 nʌp
- Tangut {dź-} for Héxī *ndź- < EMC ȵ- 娘: 𘗗₄₇₀₆ {dźjuu²} for 女 ȵi X
- Tangut {g-} for Héxī *ŋg- < EMC ŋ- 疑: 𘤡₃₅₉₀ {gjii¹} for 儀 ȵi, 毅 ȵi ɨjH, 凝 ȵi

The Sino-Tangut transcription in Lèilín reveals an affinity between the column-2 rhymes of Tangut and prenasalized voiced stops in MEDIAVAL Héxī Chinese. Could these supposedly “long vowel” rhymes indicate some kind of prenasalization? This question will be further discussed in §2.2. In the meanwhile, the exceptions need to receive a brief examination.

Two kinds of exceptions to this generalization exist. The first category, more apparent than real, concerns the Chinese syllables with weak, zero, or zero-like, initials: ʔ- 影, yi- 云, or j- 以, which are transcribed in Tangut with the initial consonant ɕ-. This can be reasonably accounted for by an internal change in source Chinese dialects, from weak initials into *ŋ- (影喻入疑 yǐng–yù ʔ ŭu):

- 鏘₄₀₃₁ {gjuu²} renders the Chinese syllable 于 yiù with a weak initial, but also Chinese syllables with the expected 疑 ȵ- initial such as 玉 ȵiowk and 御 ȵiAH.
- 鏘₀₇₇₅ {gjuu¹} transcribes the Chinese syllables 于 yiù, 羽 ȵiuX, 猶 juw, 禹 ȵiuX, and 瑀 ȵiuX with weak initials, but also Chinese syllables with the expected 疑 ȵ- initial such as 玉 ȵiowk, 御 ȵiAH, 虞 ȵiu, 言 ȵiAX, 圉 ȵiAX, and 漁 ȵia.
• 𗿖₃₃₂₃ {ŋia²} “goose” (the reconstruction of which should be revised to {gia²}, cf. §4.2.2) transcribes the Chinese syllable 晏ʔænH with a weak initial, but also Chinese syllables with the expected 疑ŋ-initial such as 雅ŋæX, 牙ŋæH, 頭ŋæH, and 牙ŋæ.

• 𗕥₁₀₀₉ {geej²} transcribes the Chinese syllable 悲ʔʌj with a weak initial, but also the Chinese syllables with the expected 疑ŋ-initial such as 艾ŋajH.

The case of 𗭢₃₇₄₅ {dzjaa²}, which transcribes 輰lienX, might also reflect variant pronunciation in the source Chinese dialect. The Tangut form reflects a likely non-standard pronunciation *ŋienX, which had since become mainstream in contemporary Chinese dialects, cf. Modern Běijīng niǎn, Suzhou nǐ, etc.

There are, however, two indisputably genuine exceptions, where the Tangut form does not have a voiced initial consonant. They are left unaccounted for in this essay.

• 𗖨₀₁₅₁ {šjuu¹} transcribes the Chinese syllable 蜀ʑowk. However, 蜀ʑowk is also transcribed in the Lèilín as 𗗡₄₄₂₅ {šju¹} and 𗗡₅₂₉₇ {šioow¹}, the latter of which belongs to rhyme R.59, mistakenly assigned by Gong Hwang-cherng to column 2.

• 𗞐₅₈₂₁ {·juu¹} renders the Chinese syllables 瑜ju, 餘ja, 佘ja. However, these syllables are also transcribed in the Lèilín as 𗂏₃₅₁₉ {·ju¹} (佘) and 𘋺₁₇₇₈ {·ju²} (瑜).

In conclusion, in the corpus of the Sino-Tangut transcription in Lèilín, Tangut column-2 rhymes almost exclusively contain Tangut syllables with a voiced stop initial, which render Chinese syllables with Northwest-type prenasalized voiced reflexes of EMC nasals.

2.2 Paired Columns in Chinese Loanwords in Tangut

This affinity between Tangut column-2 rhymes and prenasalized voiced stops in Sino-Tangut seems rather telling. Could column-2 rhymes actually express not long vowels, but the presence of prenasalization on the initials instead? Only by recourse to a contrastive scenario with minimal or near-minimal pairs could we determine the exact nature of this contrast.

We immediately encounter a problem: no attested or major reconstructed variety of Middle or Late Mediaeval Chinese contrasts plain voiced b- with prenasalized mb-. We can then remind ourselves that loanwords in the target language freeze the source form at the precise time and place of borrowing. As we saw in §1.2, plain voiced b- exists in older forms of Chinese continuing EMC
b- 並, while prenasalized mb- exists in Mediaeval Héxī Chinese < EMC m- 明. A synchronically non-existent b-:mb- contrast can be collaged, so to speak, from the different chronological strata of borrowings. In other words, a Tangut syllable with the initial b- used to render something Chinese could be an example of either of the following two cases:

- If the Tangut initial b- renders a Chinese syllable in the older layer of established Chinese loanwords, in which case we speak of an old voiced stop, the source form would have the original plain voiced value of the EMC voiced stop b- 並.
- If the Tangut initial b- occurs in one of the newer layers of Tangut-to-Chinese material, such as the newer layer of established Chinese loanwords or Late Sino-Tangut, in which case we speak of a new voiced stop, the source form would have the prenasalized Héxī reflex mb- of the EMC nasal m- 明 instead.

If the actual nature of Tangut paired columns of rhymes does involve Northwest-type prenasalization, one would expect loanwords with old voiced stops (borrowed from EMC b- 並, etc.) to occur exclusively in column-1 rhymes, and loanwords with new voiced stops (borrowed from Héxī *mb- < m- 明, etc.) to occur exclusively in column-2 rhymes. As we shall soon see, this hypothesis is confirmed by an exhaustive investigation of established Chinese loanwords in Tangut with voiced stop initials.

2.2.1 Tangut Reflexes of Chinese Borrowings with Old Voiced Stops

We first examine all established Chinese loanwords in Tangut covered in Gong Hwang-cherng’s “Xīxiàyǔ zhōng de Hànyǔ jièci” that belong to a rhyme in the major cycle (R.1–R.60) with an old voiced stop, i.e., those that have a Tangut voiced stop initial which renders a voiced stop initial in the Chinese source. We expect them all to belong to Tangut column-1 rhymes.

First, we examine the Tangut syllables in {b-} borrowed from Chinese etyma in EMC b- 並. There are four of them. As expected, all four belong to column-1 rhymes.

- Tangut 菰 0876 {bã¹} “tray, plate” is borrowed from Chinese 盤 ban “id.”
- Tangut 菰 3806 {bu¹}, in the disyllable 菰𢀜 {bu¹lọ¹} “weed,” is borrowed from Chinese 蒲 bu “cattail.”
- Tangut 菰 1971 {bia²} “to crawl, to creep” is borrowed from Chinese 爬 bæ “id.”
• Tangut \( \text{𘒚} \) \( \_1508 \) \( \{ \text{bej}^1 \} \) “to lose and flee in a war” is borrowed from Chinese \( \text{敗} \text{bæi}H \) “to lose, to fail.”

Four Tangut syllables in \( \{d-\} \) are borrowed from Chinese etyma in EMC \( d- \text{定} \).

Three belong to column-1 rhymes as expected, but there is one exception.

• Tangut \( \text{𗁐} \) \( _{3098} \) \( \{ \text{djij}^2 \} \) “to stop, to rest” is borrowed from Chinese \( \text{停} \text{den} \text{H} \) “id.”

• Tangut \( \text{𗉰} \) \( _{2833} \) \( \{ \text{djij}^2 \} \) “tranquility, certainly” is borrowed from Chinese \( \text{定} \text{den}H \) “id.”

• Tangut \( \text{𗁠} \) \( _{0712} \) \( \{ \text{du}^2 \} \) (rhyme 2.4), in the disyllable \( \text{𗁠} \text{𗁠} \) \( \{ \text{bə}^2\text{du}^2 \} \) “襪肚 \( \text{wàdù} \), a kind of scarf worn as loincloth” is borrowed from Chinese \( \text{肚} \text{duX} \) “belly.”

• Tangut \( \text{𗔜} \) \( _{2799} \) \( \{ \text{dwəə}^1 \} \) “protruding, concave” is considered by Gong Hwang-cherng to be borrowed from Chinese \( \text{凸} / \text{突} \text{du} \text{ʌt} \) “id.” The Tangut word belongs to a column-2 rhyme, and thus shows an exceptional correspondence. One might be tempted to dismiss this exception as a non-borrowing. This course should nonetheless not be taken too lightly, given that the rhyme correspondence, Tangut \( \{-a\} \) for Chinese \( \text{没} -\text{ut} \), is corroborated in late Sino-Tangut pronunciation: \( \text{𗔜} _{5233} \) \( \{ \text{phə}^1 \} \) transcribes Chinese \( \text{渤} *\text{ph}^- < \text{bu} \text{ʌt} \).

Two Tangut syllables in \( \{dz-\} \) are borrowed from Chinese etyma in EMC \( d^- \text{澄} \).

As expected, both belong to column-1 rhymes.

• Tangut \( \text{𗔘} \) \( _{0443} \) \( \{ \text{dzjo}^1 \} \) “long” is borrowed from Chinese \( \text{長} \text{diaŋ} \) “id.”

• Tangut \( \text{𗔗} \) \( _{4411} \) \( \{ \text{dzjwã}^1 \} \) “rafter” is borrowed from Chinese \( \text{椽} \text{diwen} \) “id.”

Four Tangut syllables in \( \{g-\} \) are borrowed from Chinese etyma in EMC \( g^- \text{禽} \).

As expected, all four belong to column-1 rhymes.

• Tangut \( \text{𗔗} \) \( _{5503} \) \( \{ \text{gju}^1 \} \) “canal, ditch” is borrowed from Chinese \( \text{渠} \text{gi} \text{ʌ} \) “id.”

• Tangut \( \text{𗔗} \) \( _{0005} \) \( \{ \text{gjow}^1 \} \) “to win” is borrowed from Chinese \( \text{強} \text{giaŋ} \) “strong.”

• Tangut \( \text{𗔗} \) \( _{3879} \) \( \{ \text{gju}^2 \} \) “utensil, container” is borrowed from Chinese \( \text{具} \text{giuH} \) “utensil, tool.”

• Tangut \( \text{𗔗} \) \( _{5501} \) \( \{ \text{gju}^2 \} \) “tool” is borrowed from Chinese \( \text{具} \text{giuH} \) “utensil, tool.”

Five Tangut syllables in \( \{dz-\} \) are borrowed from Chinese etyma in EMC \( dz^- \text{從} \).

As expected, all five belong to column-1 rhymes.
 Nasal Preinitials in Tangut Phonology  •  457

- Tangut 𗅒 ₂₉₈₂ {dzwa¹} “short in stature” is borrowed from Chinese 矮 dzwa “id.”
- Tangut 𘔭 ₁₆₀₄ {dzjɨ} “money” is borrowed from Chinese 錢 dzien “id.”
- Tangut 𘍒 ₅₀₉₇ {dzwej¹} “crime,” agha “is borrowed from Chinese 罪 dzwaX “id.”
- Tangut 𘉅 ₂₅₄₉ {dza¹} “mixed” is borrowed from Chinese 杂 dzap “id.”
- Tangut 𗛥 ₄₁₇₀ {dza¹} “to chisel” is borrowed from Chinese 鑿 dzak “id.”

Finally, two Tangut syllables in {dź-} are borrowed from Chinese etyma in EMC dʐ- 崇. As expected, both belong to column-1 rhymes.

- Tangut 𘁑 ₄₉₆₄ {bioo¹} “cat” is borrowed from Chinese 貓 mæw “id.”
- Tangut 𗽰 ₂₁₃₈ {buu²} “tomb” is borrowed from Chinese 墓 muH “id.”
- Tangut 𘌫 ₃₄₀₇ {dźiow²} “official report” is borrowed from Chinese 狀 dzianH “id.”

There are twenty-three established loanwords with old voiced stops. Twenty-two among them belong to column-1 rhymes. Only one exception belongs to a column-2 rhyme, which is however unexplained.

2.2.2 Tangut Reflexes of Chinese Borrowings with New Voiced Stops

We now turn our attention to established major-cycle Chinese loanwords with a new voiced stop, i.e., those borrowed from Chinese etyma with a nasal initial. First, we examine the Tangut syllables in {b-} borrowed from Chinese etyma with Héxī *mb-, reflecting a Northwest-type outcome of EMC m- 明. There are two of them. As expected, both belong to column-2 rhymes.

- Tangut 𘌫 ₄₉₆₄ {bioo¹} “cat” is borrowed from Chinese 貓 mæw “id.”
- Tangut 𘌫 ₂₁₃₈ {buu²} “tomb” is borrowed from Chinese 墓 muH “id.”

One Tangut syllable in {d-} is borrowed from a Chinese etymon with Héxī *nd- < EMC n- 泥. As expected, it belongs to a column-2 rhyme.

- Tangut 𘌫 ₂₆₃₇ {duu¹} “slave” is borrowed from Chinese 奴 nu “id.”

One Tangut syllable in {dź-} is borrowed from a Chinese etymon with Héxī *ndʐ- < EMC ȵ- 娘. As expected, it belongs to a column-2 rhyme.

- Tangut 𘌫 ₄₇₀₆ {dźjuu²} “woman” is borrowed from Chinese 女 ȵiaX “id.”
Two Tangut syllables in \{g-\} are considered by Gong Hwang-chereng to be borrowed from Chinese etyma with Héxī *ŋg- < EMC ŋ- 疑. One, however, is likely not to be a loanword. The other is exceptional, belonging to a column-1 rhyme.

- Tangut 𘃺⁰₇₅₀ \{gjii¹\} “to chew, to hold by teeth” is considered by Gong Hwang-chereng to be borrowed from Chinese 龤 ŋet “to bite.”¹⁸ I previously argued that this is an inherited word, cognate to Japhug 剋-ŋka.¹⁹
- Tangut 𗷲₁₄₇₈ \{gjĩ¹\} “to examine, to check” is borrowed from Chinese 驗 ŋiemH “id.” This exception is expected, given that nasalized rhymes do not have a distinction of paired sequences.

There are five safe Chinese loanwords with a new voiced stop. Four among them belong to column-2 rhymes. Only one exception belongs to a column-1 rhyme, which can be explained by the rhyme of the syllable. These examples show the same behavior as the Late Sino-Tangut pronunciation, shown in Table 2.

### 2.3 Conclusion

The evidence discussed in §2.1 and §2.2 suggests a radical revision of the reconstruction of Tangut paired columns of rhymes. Rather than treating the distinction between column 1 and column 2 as one between short and long vowels, it is far more natural to view it as one between the absence and presence of an initial nasal element.

Table 4 shows near-minimal pairs and other contrastive examples between the different chronological strata of Sino-Tangut materials. A column-1 syllable like 𗻍₃₈₀₆ \{bu¹\} is borrowed from EMC 蒲 bu; a column-2 syllable like 𗽰₂₁₃₈ \{buu²\} is borrowed from Héxī 墓 *mbuH < EMC muH, with a Héxī prenasalized stop *mb- corresponding to an EMC nasal. This immediate correspondence allows us to project the Chinese situation straight onto Tangut: column-2 rhymes actually indicate prenasalization, which is absent in column-1 rhymes.

---

¹⁸ Gong, “Xīxiàyǔ zhōng de Hànyǔ jiècì.”
Table 4. Proposed revision of Gong Hwang-cherng reconstruction with prenasalization.

<table>
<thead>
<tr>
<th>Column-1 rhymes</th>
<th>Column-2 rhymes</th>
</tr>
</thead>
<tbody>
<tr>
<td>short vowels → absence of prenasalization</td>
<td>long vowels → presence of prenasalization</td>
</tr>
<tr>
<td>Tangut character</td>
<td>revision</td>
</tr>
<tr>
<td>¨3806 {bu¹}</td>
<td>bu² → bu²</td>
</tr>
<tr>
<td>¨0712 {du²}</td>
<td>du³ → du³</td>
</tr>
<tr>
<td>¨0453 {dzö¹}</td>
<td>dzö¹ → dzö¹</td>
</tr>
<tr>
<td>¨3879 {gju²}</td>
<td>gu² → gu²</td>
</tr>
</tbody>
</table>

3. HYPOTHESIS: COLUMN-2 RHYMES HAVE A NASAL PREINITIAL

The marked contrast between older and newer Chinese loanword sources of Tangut voiced stop initials, shown in Table 4, heavily implies that initial prenasalization is the distinguishing element that sets apart column-2 from column-1 rhymes. However, column-2 rhymes occur not only with voiced stop initials, but with other types of initials too. It is therefore necessary to sketch a complete theory of column-2 syllables with different initial types.

Directly generalizing prenasalization is precluded by the existence of column-2 syllables with a nasal initial, such as ¨3830 {mjiij¹} “dream” or ¨4902 {ŋwuu¹} “speech.” If we understand prenasalization as a timing effect of soft palate raising, nasals, by definition, cannot be prenasalized. Revising the reconstruction of ¨2138 {buu²} from buu² to some kind of mbu², but for syllables with a nasal initial, does not make any sense unless understood as one of the following possibilities:

- An initial consonant cluster of which the first element is a nasal consonant. Thus, ¨3030 {mjiij¹} “dream,” reconstructed as mee¹ under the uvularization hypothesis, is to be revised as mme¹, with an initial “geminate” consonant cluster of mm-;
- A minor syllable in a sesquisyllabic phonotactics. For example, ¨3030 {mjiij¹} is to be understood as mme¹, with a demi-syllabic preinitial m preceding a syllable me¹.

The difference between these two treatments is neither knowable in principle nor consequential with regard to other aspects of Tangut synchronic and diachronic linguistics. Following general usage in Sino-Tibetan, East Asian, and
Mainland Southeast Asian linguistics, syllables previously reconstructed with a “long vowel” should be analyzed as having a nasal preinitial.

The interaction of the nasal preinitial hypothesis with the problem of the reconstruction of the Tangut voiced series needs to be addressed briefly. Gong Hwang-cherng proposes that this voiced stop series of initial consonants should be reconstructed as non-prenasalized plain voiced consonants.20 Guillaume Jacques, on the other hand, demonstrates that the Tangut voiced series come from the Pre-Tangut prenasalized voiced stop, a result that raises the possibility that even during the Tangut empire period, the voiced series remained prenasalized.21 The nasal preinitial hypothesis favors the non-prenasalized value of plain voiced consonants, as the contrast between 𗻍₃₈₀₆{bu¹} and 𗽰₂₁₃₈{buu²} would be less awkward as one between buⁿ¹ and mбуⁿ² than one between ṱbuⁿ¹ and mⁿbuⁿ². The latter scenario is not impossible, as Swahili features a contrast between trisyllabic /m.bu.ni/ “coffee bush” and disyllabic /ⁿbu.ni/ “ostrich,” but presupposes a sesquisyllabic treatment.22

One final question pertains to the identity of the nasal preinitial. Etymologically (see especially §4.3.1–3), the nasal preinitial predominantly derives from earlier dental *n-. However, before stops and nasal, there are strong reasons to consider the nasal preinitial as homorganic. Thus, the closest typological parallel is the nasal preinitial of modern West Rgyalrongic languages, likely the closest relatives to Tangut, such as Khroskyabs24 and Geshiza25. In these languages, the nasal preinitial N- is homorganic before stops: Khroskyabs ṱɡ, Geshiza ṱɡə “nine.” However, before other types of consonants, it can surface as a dental n- instead: Khroskyabs ṱnɡə, Geshiza ṱnɡə “soft.”

Based on the Sino-Tangut data, etymological comparisons, and typological parallels, I propose the following distribution before different initial types:

- Before stops or affricates, a homorganic nasal preinitial is reconstructed: mp-, nd-, nk-, ng-, nstsh-, ndz-… For example, 𗻍₂₁₃₈{buu²} “tomb,” reconstructed under the uvularization hypothesis26 as buuⁿ², should be revised into mбуⁿ².

---

23 Lai et al., “Tangut as a West Gyalrongic Language.”
26 Gong, “Xīxiàyǔ zhōng de Hànyǔ jièci.”
• Before nasals, Khroskyabs dialects show a dental n-27. In Tangut, however, the mixed fânpìè behavior (§4.2.1) and etymological origin in *ɣ- and *ʁ- (§4.3.4) favors an interpretation as a homorganic nasal preinitial. The preinitialed nasal initials can be labeled *geminate nasals: mm-, nn-, ŋŋ-, NN-. Thus, the reconstruction of 復5702 {mjaa¹} “ulcer” should be revised from maa¹ to mma¹.

• Before sibilants, the preinitial n- is reconstructed: ns-, nś-, nz-, nź-. For example, the reconstruction of 復0716 {sti¹} “to butcher” should be revised from śii¹ to nśi¹, cf. Khroskyabs ɲɕ.

• Before laterals, the preinitial n- is reconstructed: nl-, nlh-. For example, the reconstruction of 復5522 {liij²} “to wait” should be revised from lee² to nle², cf. Khroskyabs njé.

• Before glides, the nasal preinitial n- is reconstructed28: n·w-, n·j-. For example, the reconstruction of 復0320 {·wəə¹} should be revised from wəə¹ to n·wə¹, cf. Geshiza nvo.

• Few instances of attested Tangut syllables in column-2 rhymes have initials x-, ɣ-, and ·- (zero initial). Purely as a notation, one could write nx-, nɣ-, and n·- for such cases. The only examples of such syllables with nx-, etc. will be argued in §4.2.3 to be spurious, lacking the nasal preinitial in reality.

The proposed reconstruction will be presented again in Table 5 in the conclusion.

4. DISCUSSION

In this section,29 I will discuss the interaction of the nasal preinitial hypothesis with other sources of evidence on Tangut phonology, starting with other transcription materials (§4.1), followed by the fânpìè practice in native Tangut rhyme books (§4.2), and concluding in a comparison with modern Rgyalrongic and other Burmo-Qiangic languages (§4.3).

28 Given that the Gong Hwang-cherng reconstruction does not admit a reliable distinction between initial glides w-, j- and zero initial followed by glide medial -w-, j-, Gong Hwang-cherng’s zero initial symbol ·- can be repurposed as an orthographical separator, thus n·w-, n·j-. Also, with regard to the initial j- question, it is assumed that any zero-initial Grade III syllable with a nasal preinitial has an initial yod.
29 Tangut-language literature referred to in this section, unless otherwise indicated, came from the Tangut collection of the Institute of Oriental Manuscripts of the Russian Academy of Sciences, St. Petersburg, published as Institute of Oriental Manuscripts of the Russian Academy
4.1 Nasal Preinitials in the Other Transcription Materials

Tibetan transcription evidence is rather limited in utility, as most cases of the Tibetan nasal preinitial འ‘-> occur before voiced stops indiscriminately in both column-2 rhymes with a Tangut nasal preinitial and column-1 rhymes without one. As discussed in the previous section, there are reasons to prefer both a prenasalized and non-prenasalized value for Tangut voiced stops. The nasal preinitial hypothesis is only moderately in favor of the latter hypothesis. Accordingly, under either belief, this situation can be analyzed as reflecting some residual prenasalization of the Tangut voiced series or reflecting a deprenasalization in the Bde dialect of Tibetan in parallel to the Tangut sound change.

However, it is still noteworthy that two of the only three examples of <‘-> preceding a voiceless aspirated stop in the Tibetan transcription involve Tangut column-2 rhymes:

- མོ༣༢ mpo² {phjoo²} “to combine,” transcribed as འཕོ ‘pho).
- ཐོ༢༢ nmo¹ {thjoo¹} “wonderful,” transcribed as འཐོ ‘tho).

The one exception is ལེཤ བ ར འ ‘kha. However, this could be understood as one of the orthographical devices used to transcribe in Tibetan letters the Tangut uvular initial.

Sanskrit transcriptions, on the other hand, provide a much more solid basis for discussion. A particularly interesting point to consider is a transcription practice discovered by Arakawa Shintarō in the Tangut version of Mahāmāyūrīvidyārājñī. From this collection of mantras, Arakawa documented...
the existence of six nasal-CV characters. It turns out that five of the proposed nasal-CV characters belong to a column-2 rhyme and therefore, under the nasal preinitial hypothesis, have nasal preinitials:

- 細5792 nda¹ {djaa¹}, used to transcribe ratnakaraṇḍake.
- 細1512 mba²t {baa¹}, used to transcribe ambare ambarāvati.
- 細5416 ndu¹t {djuu¹}, used to transcribe vindūpati.
- 細5388 mbo²{boo²}, used to transcribe kambu.
- 細3889 mbe² {bjiij²}, used to transcribe duṃbe doduṃbe.

Arakawa also considers the column-1 syllable 細5743 gi¹ {gji¹} to be a “nasal-CV” character.36 This character appears in the segment 細9574 phə²rir²gi¹rir²ka¹·je² {phjɨ¹rjir²gji¹rjir²kjaa¹·jij²}, used to transcribe bhṛṅgārīkāya, which poses too much of an irregularity to warrant a conclusion one way or the other. There is, however, another possible candidate of a “nasal-CV” character for Sanskrit ngi/ṅgī in the Mahāmāyūrīvidyārājñī. In the same dḥāraṇī, the word utṭiṅgirī38 is transcribed as ཆུས་འཛིན་འབྱོར་འཁོར་གིས་ngi {gji²} {wu²ti²nggi²rir²}, which transcribes the segment ngī, thereby constituting

---

36 Arakawa, Seikago tsūin jiten, 115.
37 The character 細3984 {kjaa¹} belongs to the rhyme R.21 (1.21–2.18), whose status as a “long vowel” rhyme assigned by Gong Hwang-cherng is mistaken, cf. §2.1.
38 Based on a majority of Sanskrit mss., all written in the Rañjanā script, Takubo Shūyo reads daṭāṇ 것이ni for this word, which he proposes to emend to daṭiṅgi. (see Takubo, Bonbun Kujaku myōōkyō / Ārya-Mahā-Māyūrī Vidyā-Rājñī, 46.) The Sanskrit texts that served as the basis for Takubo’s edition, however, do not agree with most of the mediaeval Chinese and Tibetan testimonies—a point that Takubo failed to note. The Chinese translations read 僧伽婆羅, i.e., utṭiṅgiri or uddingiri (Sanghabala 僧伽婆羅 tr., ~520CE, Taishō 0984), 呼微祗哩, i.e., utṭiṅgiri (Yijing 圓浹 tr., ~700CE, Taishō 0985), and 呼微; 僧伽婆羅 tr., i.e., utṭiṅgiri (Amoghavajra 不空 tr., ~760CE, Taishō 0982). The standard Tibetan version (Derge 0559, Peking 0178) reads  Ngbu ngini; the re-Sanskritized form, a rather awkward one since Skt -ṣṭ- would have yielded an aspirated -ṭṭh- instead, points to an underlying shape similar to that of Amoghavajra. Takubo also cited Serge Oldenbourg’s earlier 1899 Sanskrit edition. The nature of this reference must have been rather bibliographical, however, since Oldenbourg, too, reads utṭiṅgiri—see Ol’denburg, Serge, “Otryvki kašgarskix i sanskritskix rukopisej iz sobranija N. F̀. Petrovskogo [Some Kashgarian and Sanskrit Manuscripts from N. F. Petrovsky’s Collection]”, Zapiski vostočnago otdelenija imp. russ. arch. obščestva 11 (1899), 252. Oldenbourg’s Sanskrit ms., written in a mixed Gupta script with “Kashgar” (= South Turkestan Gupta?) elements (p. 208), must lie rather close to the ancestral text of the Chinese and Tibetan editions.
a possible “nasal-CV” character not discussed in Arakawa. This syllable is also a column-2 syllable as expected. Thus, all currently known reliable “nasal-CV” characters have nasal preinitials under the nasal preinitial hypothesis.

By way of summary, although the Tibetan evidence favors the nasal preinitial hypothesis only slightly, the consistent column-2 status of “nasal-CV” characters in the Sanskrit-to-Tangut transcription of the Mahāmāyūrīvidyārājñī strongly supports the nasal preinitial hypothesis.

4.2 Nasal Preinitials and Fǎnqiè Evidence

4.2.1 Fǎnqiè Behavior of Syllables with Nasal Preinitials

If Tangut column-2 rhymes indeed involve a nasal preinitial, we may expect that native speakers conceive of the presence of the nasal preinitial as part of the initial consonant rather than rhyme. We can catch a glimpse of the native-speaker psychology concerning syllable structure from the practice of fǎnqiè 反切 in native dictionaries, a type of phonetic spelling which functions by phonologically segmenting a syllable into the initial, indicated by the initial speller (反切上字 fǎnqiè shàngzì), and the rhyme, indicated by the rhyme speller (反切下字 fǎnqiè xiàzì). Thus, we would predict that the initial speller of a column-2 rhyme should itself belong to a column-2 rhyme.

It is beyond the scope of this article to attempt an exhaustive examination of all fǎnqiè xìlián 反切系聯 sequences in Sofronov in the light of the nasal preinitial hypothesis. I content myself with presenting in Figure 2 all the fǎnqiè ancestors of the character 鼓nsi¹ ʂjii¹ “to butcher.” As we can see in Figure 2, all its fǎnqiè ancestors of both sides themselves belong to a column-2 rhyme, reconstructed as having a nasal preinitial.

---

39 Arakawa, Seikago tsūin jiten, 115.
40 Sofronov, Mikhail V., Grammatika tangutskogo jazyka [Grammar of the Tangut Language], vol. 2 (Moscow: Nauka, 1968).
This generalization of matched presence of nasal preinitials—that both fānqiè spellers of a syllable with nasal preinitials have nasal preinitials—is not without exceptions. “Geminate” nasal initials (mm-, nn-, ŋi-, ŋn-), in particular, can show unmatched spellers. For example, the character 畏₄₉₀₂ ŋn⁴₁ wū¹ {ŋwuu¹} “speech” is spelt 畏₂₀₂₆ nву⁴¹{ŋwuu¹} + 畏₅₆₂₅ nthwuu¹{thwuu¹}. 畏₄₉₀₂ nву⁴¹{ŋwuu¹} belongs to the column-2 rhyme 1.5 in Wénhǎi. The column-2 character of this word is also supported by the fact that it is indicated as non-homophonous to column-1 畏₂₀₂₆ nву⁴¹{ŋwuu¹} in Homophones and other dictionaries. Under the nasal preinitials hypothesis, 畏₄₉₀₂ nву⁴¹{ŋwuu¹} “speech” does have a nasal preinitial.

Thus, in the case of nasal-initial syllables like 畏₄₉₀₂ nву⁴¹{ŋwuu¹}, the unmatched spellers 畏₂₀₂₆ nву⁴¹{ŋwuu¹} + 畏₅₆₂₅ nthwuu¹{thwuu¹} are resolved in preference to the rhyme speller 畏₅₆₂₅ nthwuu¹{thwuu¹}. This fānqiè behavior is probably unsurprising as mm- resembles m- acoustically in a way that mb- does not resemble b-. In §4.2.3, We shall see that this is not always the case with other unmatched spellers.

It is worth pondering, as one anonymous reviewer suggests, whether the fact that the rhyme speller for a column-2 syllable is itself column-2 implies that the distinction of parallel columns does not only involve nasal preinitials but is also associated with some vocalic feature. The current evidence does not allow us to decide the question either way, as the absence of vocalic cor-

---

**Figure 2:** All fānqiè ancestors of the character 畏₄₀₇₁₆ nši¹ {ʃiʃi¹} “to butcher.”
relates is plausible given the formal rules of fānqiè. If there is some kind of vocalic correlate, however, I would lean toward weak nasalization rather than Gong Hwang-cherng’s hypothesis of vowel length, given the general direction of compression in which Tangut phonology has developed.\textsuperscript{41}

4.2.2 Revising the Tangut Reconstruction of 鵝₃₃₂₃ NNʕa²{ŋiaa²} “Goose”

The rule of matched presence of nasal preinitials contributes to our growing understanding of the behavior of Tangut fānqiè, which in turn can help determine the validity of fānqiè spellings in sources. An example is the character 鵝₃₃₂₃, in the disyllable 鵝₃₆₂₃ 鵝₃₃₂₃{bā¹-} “goose.” Its pronunciation given in all versions of the Tangut-Chinese Dictionary\textsuperscript{42} is NNʕa²{ŋiaa²}. It is justified by the following fānqiè spelling given in the dictionary entry: 鵝₂₇₇₇ Ne⁎r¹{ŋewr¹} + 鵝₅₇₆₆ Nʕa²{siaa²}. This fānqiè spelling is doubly suspicious given that:

- 鵝₂₇₇₇{ŋewr¹} is a retroflex syllable, which usually only enters into a fānqiè relationship with other retroflex syllables. 鵝₃₃₂₃{ŋewr¹} belongs to Sofronov’s fānqiè xīlián series velar-20, which contains the following characters in Sofronov (1968:81), which are all retroflex: 鵝₁₃₅₇ Nwi²r¹{ŋwer¹} “to equate,” 鵝₀₃₃₃ Nwi²r²{ŋwer²} “keen,” 鵝₄₄₂₃ Nu²r¹{ŋur¹} “head,” 鵝₁₃₅₇ Nwi²r¹{ŋwer¹} “hill,” 鵝₁₄₂₃ Nwə²r¹{ŋwər¹} “seven,” 鵝₁₂₁₉ Nwi²r²{ŋwer²} “slack.”
- Similarly, as §4.2.1 argues, prenasalized syllables mostly enter into a fānqiè relationship with other prenasalized syllables, though in the case of ŋ- the rule is not as strict.

This fānqiè spelling of the Tangut-Chinese Dictionary is obviously not taken from the Wénhǎi, of which only the píng-toned volumes have survived to this day. A review of the sources reveals that it originates from the Combined Edition of Wénhǎi and Homophones (original title lost; assigned the title 同音文海宝韵合编 by Hán Xiǎománg)\textsuperscript{43}, which survives chiefly in Inv. No. 4153/4781/6685/8179, a particularly challenging manuscript written in a careless semi-cursive hand.

on the reverse side of another document. The dictionary entry of "goose" is reproduced in Figure 3.

**Figure 3:** The dictionary entry for the character ‘goose’ in the *Combined edition of Wénhǎi and Homophones* (Inv. No. 4153/4781/6685/8179, 24-15), zooming in on the initial speller in question ( Engineers?), with the diagnostic -shape highlighted.

The actual pronunciation of this character thus depends on the reading of the initial speller in the *Combined Edition*, namely  Engineers, which is also shown enlarged in Figure 3. Hán Xiǎománg retains the doubly dubious reading of the *Tangut-Chinese Dictionary*, namely  Engineers and  Engineers. I propose that the initial speller  Engineers should instead be read as  Engineers. This reading is not only compatible with other sources of evidence for the pronunciation of "goose" but is also paleographically more convincing.

The single most diagnostic difference between  Engineers and  Engineers is between the upper portions of the middle components: between Engineers and Engineers. As a matter of fact, the contrast between Engineers and Engineers is among the most robust in the notoriously chaotic semi-cursive and cursive styles of hand-

---

44 Hán, Tóngyín Wénhǎibàoyùn hébiān zhènglǐ yǔ yánjū, 144.
written Tangut. In the following examples, semi-cursive characters are taken from the manuscript version of the Art of War\textsuperscript{45}; cursive characters are taken from Sūn Yīngxīn’s study of the Eight Upavāsa Precepts. (2015)\textsuperscript{46}.

- \textsuperscript{3} is usually written with a joined \(\angle\)-shape, cf. semi-cursive \(\us\) for \(\text{𘟙}_3830\), \(\us\) for \(\text{𘆨}_40646\); cursive \(\us\) for \(\text{_TypeDef}_2559\), \(\us\) for \(\text{_TypeDef}_2640\).

- \(\text{𓊄}\), on the other hand, always conserves its regular shape with the left dot often independent, cf. semi-cursive \(\text{𓊄}\) for \(\text{_TypeDef}_5751\), \(\text{𓊄}\) for \(\text{_TypeDef}_4797\), cursive \(\us\) for \(\text{_TypeDef}_5215\), \(\us\) for \(\text{_TypeDef}_3452\).

- Characters like cursive \(\text{𓊄}\) for \(\text{_TypeDef}_5388\) tellingly contrast both graphical elements within a single character.

The initial speller \(\text{𓊄}\) contains the characteristic \(\angle\)-shape indicative of the \(\text{𓊄}\) component, and hence should be read as \(\text{𓊄}_{2776}\) \(\text{ŋu}^2\) \{\text{guiu}\} instead of \(\text{𓊄}_{2777}\) \(\text{ne}^\text{wr}\) \{\text{newr}\}. With this revision of the initial speller, the reading of the character \(\text{_TypeDef}_3323\) should be revised from \(\text{nŋa}^2\) \{\text{niaa}\} to \(\text{nŋa}^2\) \{\text{giaa}\}.

This proposed revision can be generalized into the following conjecture: whenever a Tangut syllable in a column-2 rhyme is used to transcribe a Chinese syllable with an EMC nasal initial, the syllable is likely to have a voiced stop initial rather than a nasal initial.\textsuperscript{47}

4.2.3 Transcription Characters with Apparent Nasal Preinitials Before Zero Initial and \(\text{x}\)-

Staying on the subject of R.23 (2.20), I conclude this section by discussing the characters \(\text{𓊄}_{4623}\) \(\text{nŋa}^2\) \{\text{niaa}\}, \(\text{𓊄}_{40871}\) \(\text{nŋa}^2\) \{\text{niaa}\}, and \(\text{𓊄}_{2856}\) \(\text{ŋxâ}^2\) \{\text{xiaa}\}, which constitute the only reliable examples of column-2 syllables with the initials \(\text{GetMethod}\) and \(\text{x}/\text{ӵ}\)-. There are no reliable examples that start with \(\text{ŋ}/\text{ŋ}/\text{k}\)-. All these characters are special characters presumably created for the purpose of transcribing Sanskrit. Their fānqiè spelling and assumed Sanskrit target of transcription are as follows:

\footnotesize
\begin{itemize}
\item Note that this hypothesis does not generalize to the opposite Tangut-to-Chinese direction: in the Pearl in the Hand, Chinese syllables with EMC nasal initials happily transcribe both nasal and voiced stop initials in Tangut.
\end{itemize}
the pronunciation of both problems can be eliminated by removing the nasal preinitial, i.e., revising \( ^{2} \). Given that there are no syllables reconstructed with \( ^{n} \cdot a \), \( \chi \) and \( n \cdot - \) featuring the guttural initials with the nasal preinitial \( n \cdot - \), \( \chi \) and \( n \cdot - \), respectively. Equally unnatural is the fact that these characters are the only examples of Sanskrit syllables actually transliterated with this character, but it would probably denote Sanskrit \( h \). As a concluding remark, this revision does complicate the picture of the treatment of unmatched \( f \) spellings with regard to the nasal preinitial:

- \( n \cdot a^{\prime} \) has no surviving \( f \) spelling, but it is a \( f \) character made up of \( n \cdot - \) and \( a^{\prime} \). I am not aware of any Sanskrit syllables actually transliterated with this character, but it would

In the framework of the nasal preinitial hypothesis, \( n \cdot a^{\prime} \) as a transcription of \( a \) or \( a \) would be quite unnatural, as would \( n \cdot a^{\prime} \) as a transcription of \( h \). Equally unnatural is the fact that these characters are the only examples featuring the guttural initials with the nasal preinitial \( n \cdot - \), \( n \cdot - \), and \( n \cdot - \). Given that there are no syllables reconstructed with \( a^{\prime} \) and \( n \cdot a^{\prime} \), both problems can be eliminated by removing the nasal preinitial, i.e., revising the pronunciation of \( n \cdot a^{\prime} \) and \( n \cdot a^{\prime} \) to \( a^{\prime} \), and that of \( n \cdot a^{\prime} \) to \( a^{\prime} \). As a concluding remark, this revision does complicate the picture of the treatment of unmatched \( f \) spellings with regard to the nasal preinitial:

- \( n \cdot a^{\prime} \) is spelt \( n \cdot a^{\prime} \) {\( j i^{2} \) + \( t \cdot i^{2} \)}. It likely used to denote the Sanskrit syllable \( ã \), as it is glossed in the \( Wénhài \) as \( a^{\prime} \) and \( a^{\prime} \) “one of the four major seed syllables (\( bi \), \( jì \) ).” (Inv N° 211 212 213:100–108).
- \( n \cdot a^{\prime} \) is spelt as \( n \cdot a^{\prime} \) {\( t \cdot i^{2} \) + \( t \cdot i^{2} \)}. It refers transparently to the Sanskrit syllable \( ã \), as it is graphically derived from \( a^{\prime} \) “Sanskrit syllable \( a^{\prime} \)” and \( t \cdot a^{\prime} \) {\( d \cdot i^{2} \) “long.”}
- \( n \cdot a^{\prime} \) {\( x i a^{\prime} \)} has no surviving \( f \) spelling, but it is a \( f \) character made up of \( n \cdot -x \) {\( j u^{2} \) + \( t \cdot n x \)}. I am not aware of any Sanskrit syllables actually transliterated with this character, but it would probably denote Sanskrit \( h \).

As a concluding remark, this revision does complicate the picture of the treatment of unmatched \( f \) spellings with regard to the nasal preinitial:

- \( n \cdot a^{\prime} \) “speech” is spelt \( n \cdot a^{\prime} \) {\( n w u^{2} \) + \( n \cdot a^{\prime} \) \( t h w u^{1} \)}. This syllable, by virtue of belonging to rhyme 1.5 (R.5) and contrasting phonologically with \( n \cdot a^{\prime} \) \( n w u^{2} \) in the native dictionaries, does indeed have a nasal preinitial. In this case, the conflict between the initial speller without a nasal preinitial \( n w u^{2} \) and the rhyme speller with a nasal preinitial \( t h w u^{1} \) is resolved in preference of the latter.
- In the case of \( n \cdot a^{\prime} \) \( n \cdot a^{\prime} \) \( n w u^{2} \) \( t h w u^{1} \), spelt \( n \cdot a^{\prime} \) {\( j i^{2} \) + \( t \cdot i^{2} \)}, systematic and transcriptional considerations discussed above suggest the absence of the nasal preinitial. Thus, the conflict be-

---

48 One anonymous reviewer raised the question whether the fact that these syllables all seem to transcribe Sanskrit long vowels could not support Gong Hwang-cherng’s hypothesis of vowel length instead. I consider this an unlikely possibility. 

49 I am not aware of any Sanskrit syllables actually transliterated with this character, but it would probably denote Sanskrit \( h \).

---
between the initial speller \( ji \) and the rhyme speller \( nts\text{ī}a\text{ŋ} \) is resolved in favor of the initial speller.

This lack of consistency, while worrisome, should be maintained nonetheless given the strong rationales for both treatments in the cases mentioned. There is a serious need for a wholesale reinvestigation of Tangut \( f\text{ā}nqiè \) behavior addressing the problems raised in this section.

4.3 Comparative Problems

Jacques’ *Esquisse de phonologie et de morphologie historique du tangoute*, our primary authority on Tangut etymology, features no discussion regarding the origin of Tangut column-2 (“long-vowel”) rhymes; nor did the subsequent literature address the issue. Revising the value of column-2 rhymes from vowel length to the presence of a nasal preinitial enables meaningful hypotheses to be postulated as to the origin of these rhymes. In this section, we discuss all the etymologies postulated in the *Esquisse* of Tangut words belonging to a column-2 rhyme in the major cycle,\(^{49}\) as well as a few other cognates not proposed there. The comparison is made, as usual, mostly against modern Rgyalrongic languages, especially Japhug and occasionally its fellow Upper Rgyalrong languages Tshobdun and Zbu, as well as modern West Rgyalrongic languages, Khroskyabs, and an assortment of Stau-Horpa lects, likely the closest relatives to Tangut.\(^{50,51}\) The Pre-Tangut and Proto-Rgyalrong(ic) forms are provisional and liable to further changes.

---

\(^{49}\) However, the tentative comparison given by Jacques in the *Esquisse* between \( \text{mf}_2\text{mb}_2 \) \( \{\text{mji}_1\} \) “to give” and Japhug \( \text{m}\text{mb}_1 \) “id.” judged as “pas certaine” (p. 37) and “très problématique” (p. 97) on account of its sui generis correspondence between Tangut \( m\) - and Japhug \( \text{mb}_1 \), is not discussed. One anonymous reviewer suggested that \( \text{mf}_2\text{mb}_2 \) \( \{\text{mji}_1\} \) could be related to Geshiza \( \text{m}\text{ə} \) “to feed” instead.

\(^{50}\) Lai et al., “Tangut as a West Gyalrongic language.”

\(^{51}\) Japhug data is cited from the *Esquisse* and checked against the latest version of Jacques’ dictionary (*Dictionnaire Japhug-chinois-français, Version 1.1.*). Zbu Rgyalrong data are cited from Gong’s *Le rgyalrong zbu, une langue tibéto-birmane de Chine du Sud-ouest : une étude descriptive, typologique et comparative*. Khroskyabs data are cited from Lai’s *Grammaire du khroskyabs de Wobzi*. “Stau” data, referring to the Stau dialect of Khanggsar, are kindly provided by Guillaume Jacques, Lai Yunfan, Anton Antonov, and Lobsang Nyima (cf. the authors 2017). “Geshiza” data, referring to Eastern Geshiza of Balang, are cited from Honkasalo’s *A grammar of Eastern Geshiza: A culturally anchored description*. Other Stau-Horpa lects, as well as a few forms in Zbu and Tshobdun, are cited from the r*Gyalrongic Languages Database*, ed. Yasuhiko Nagano and Mariëlle Prins. Entries of the *Database* are annotated with the locality, the four-letter locality code, and the numerical entry code. Whenever there is a retranscription, the original form is also left in parentheses: Mda·mdo \( \text{kājə} \) (\( \text{kān’jə}, \text{DB-dand-1993} \)).
4.3.1 Root Dental Preinitial n- in Rgyalrongic Languages

Several Tangut words with a nasal preinitial show good evidence for a nasal preinitial *n- in Pre-Tangut.

- 𘙣₀₇₁₆ nśi¹ {śjii¹} “to butcher” is a cognate of Japhug -ntsha “to butcher,” from proto-Rgyalrong *-nə. Some West Rgyalrongic languages preserve a form close to the revised Tangut reconstruction, such as Khroskyabs ɲɕa.
- 𗡡₀₃₂₀ n·wə¹ {·wəə¹} “soft” is a cognate of Japhug -mpɯ from proto-Rgyalrong *-nəpu. Some West Rgyalrongic languages preserve a form close to the revised Tangut reconstruction, such as Khroskyabs ɲvə and Stau-Horpa forms such as Geshiza nvo, Tag-gsum ḭvə (DB-dasa-1714).
- 𗢭₃₁₁₃ ɲɡə¹ {gjɨɨ¹} “nine” is a cognate of Japhug kɯng. While all modern Rgyalrongic evidence points to a nasal preinitial *n-, Sino-Tibetan comparanda such as Tibetan dgu suggest a Pre-Proto-Rgyalrong form **-təⁿɡu. Modern West Rgyalrongic languages, such as Khroskyabs Ḧgā, Geshiza Ḧgæ, share the place assimilation hypothesized for Tangut.
- 𗭻₅₅₂₂ nle² {ljiij²} “to wait” is a cognate of Japhug -njo, Zbu -npe'ŋdʒè, from a proto-Rgyalrong root akin to *-nalaŋ. Khroskyabs njé similarly preserve the nasal preinitial.

The n- preinitials in two cases are not of obvious Pan-Rgyalrongic pedigree but must be reconstructed using evidence specifically from modern West Rgyalrongic languages.

- 𘞌₀₉₀₂ źər¹ {źjɨ r¹}, leaves the slim but enticing possibility of a separate Proto-Rgyalrongic root shared by Japhug -rzi.
- 𘞊₂₆₂₁ nse² {sjiij²} “to think” is analyzed by Jacques as cognate to Japhug -susə, however, Khroskyabs ntʃa, Stau and Geshiza ntʃə seem to present a better candidate for cognacy.

---

52 Gong, Le rgyalrong zbu, une langue tibéo-birmane de Chine du Sud-ouest, 303–9.
53 An alternative Tangut word for “heavy, weight,” 仩₀₀₀₂ žər¹ {źjir¹}, leaves the slim but enticing possibility of a separate Proto-Rgyalrongic root shared by Japhug -rzi.
54 Jacques, Esquisse de phonologie et de morphologie historique du tangoute, 180.
4.3.2 Autobenefactive *nə-

In one case, the Tangut nasal preinitial clearly derives from the Rgyalrongic autobenefactive prefix *nə-.\(^{55}\)

- \(\text{ŋh}^{\text{4040}} \text{qhu}^{\text{51}} \{\text{khu}u\} \) “to greet”\(^{56}\) is a cognate of Japhug -qr. Zbu and West Rgyalrongic uniformly prefer a form with the autobenefactive *nə-: Zbu -nəqrə, Khroskyabs ḳʰrə, Stau nərə.

Three other verbs have an unexplained nasal preinitial, which probably also reflects the Rgyalrongic autobenefactive prefix *nə-. It is worth noting, however, that none of these etymologies seems particularly solid.

- \(\text{ŋ}^{\text{2621}} \text{nse}^{\text{2}} \{\text{sjiij}^{\text{2}}\} \) “to think” is analyzed by Jacques as cognate to Japhug -sus.\(^{57}\) While in §4.3.1 I propose that it could also be analyzed as cognate to Khroskyabs nštə, retaining Jacques’s etymology would suggest that this word is an example of autobenefactive *nə-.

- \(\text{ŋ}^{\text{0369}} \text{nθu}^{\text{1}} \{\text{θjuu}^{\text{1}}\} \) “to inspect” is judged by Jacques\(^{58}\) to be “potentielle-ment […] rapproché” to Japhug -θu “to ask.”

- \(\text{ŋ}^{\text{5612}} \text{nθʃe}^{\text{1}} \{\text{θʃjiij}^{\text{1}}\} \) “to speak” is judged by Jacques\(^{59}\) to “potentiellement se comparer” to Japhug -ti. The correspondence, while rather poor, cannot be entirely ruled out, as for example Zbu B dialects have comparable forms such as Go·la·thang ké-tse (ka’tse, DB-gele-0904).

4.3.3 Stative *ŋa- Before a Pre-Tangut Acute Prenasalized Voiced Initial

Two instances of the nasal preinitial, both with Tangut voiced stop initials < Pre-Tangut prenasalized voiced initials, derive from the stative prefix *ŋa-, reflected as Japhug a-.

- \(\text{ŋ}^{\text{3149}} \text{nθu}^{\text{1}} \{\text{duu}^{\text{1}}\} \) “to accumulate” < Pre-Tangut *ŋ-nθu is a cognate of Japhug -ajtu or alternatively -ndu, from Proto-Rgyalrong *ŋa-ləntuŋ.

- \(\text{ŋ}^{\text{2396}} \text{ndθu}^{\text{2}} \{\text{dzuu}^{\text{2}}\} \) “to sit” < Pre-Tangut *ŋ-nθu is a cognate of Japhug -amdzu, Zbu ‘-amdzo’ “id.,” from Proto-Rgyalrong *ŋa-mədzuŋ. This verb


\(^{56}\) Concerning the rhyme of this word R.6 (1.6), cf. Miyake, “Complexity from Compression.”

\(^{57}\) Jacques, Esquisse de phonologie et de morphologie historique du tangoute, 180.

\(^{58}\) Ibid., 50.

Nasal Preinitials in Tangut Phonology

473

has a cognate in Stau-Horpa: Stau ndzo, Geshiza ndzo, which agrees with the Tangut form.

As a matter of fact, the stative *ŋa- does not induce the Tangut nasal preinitial before other initial types. Japhug -astu “straight,” for example, corresponds to 𗹐₁₅₆₉twụ¹{twụ¹} “straight (esp. morally).” The verb 𗨻₂₂₂₆wi²{we²} “to become,” in an etymology not discussed by Jacques in the Esquisse, is a passive formation derived from 𗶠₅₁₁₃wi¹{wji¹} “to do,” parallel to Japhug -apa “to become.” The attested form is wi²{we²}, not †n.wi²{we²}. In addition, in the case of 𘃡₁₆₃₈gi¹{gji¹} “to do,” parallel to Japhug -apa “to become,” derived from -pa “to do.” The attested form is wi²{we²}, not †n.wi²{we²}. In addition, in the case of 𘄎₁₆₃₈gi¹{gji¹} “clear (water)” < Pre-Tangut *ŋa-ⁿɡri, cf. Japhug -amgri, Zbu -emgréy, Khroskyabs ƣgré, the same *ŋa- fails to induce the Tangut nasal preinitial even before a voiced (< *prenasalized) stop.

Thus, an intermediate *n- must be postulated, the outcome of *ŋ- assimilated to the following acute prenasalized initial. Before a grave initial, such as in the case of 𘂴₁₅₆₇₇mme¹{mjiij¹} “tail” is a cognate of Japhug tɤ-jme, Zbu tɤ-lmëʔ, from Proto-Rgyalrong *-ləmɛ.

4.3.4 Geminate Nasal Assimilation

The examples discussed in §4.3.1–3 can all be traced, in one way or another, to a Pre-Tangut preinitial *n-. A large number of examples, however, correspond to a wide range of preinitials in modern Rgyalrongic comparanda. Their only commonality is that they are followed by a nasal initial. In these cases, a rather atypical assimilation, whereby any preinitial is assimilated toward a geminate nasal, i.e., mm- < *lm-, *rm-, *sm-, *km- ... must be postulated.

In two examples, the geminate nasal arises from an earlier preinitial *l-:

- 𘂴₅₆₇₇mme¹{mjiij¹} “tail” is a cognate of Japhug tx-jme, Zbu tx-lmëʔ, from Proto-Rgyalrong *-ləmɛ.
- 𘈞₀₃₃₀mme¹{mjiij¹} “dream” is a cognate of Japhug tuu-jmŋo, Zbu ta-lmáʔ, from Proto-Rgyalrong *-ləmaŋ.

However, earlier *l- before nasals also give reflexes as tense syllables. Known examples are 𘓔₂₃₂₅mə̣²{mjɨ²} “to forget,” cf. Japhug -jmt; 𗡔₄₆₀₀n wụ¹{ŋwụ¹} “oath,” cf. Japhug kɯjŋu.
In two examples, the geminate nasal arises from an earlier preinitial \(r\)-:

- \(\text{mme}^2 \{\text{mjiij}\}\) “name” is a cognate of Japhug \(\text{ty}-\text{rmi}\).
- \(\text{mme}^2 \{\text{mjiij}\}\) “house” is probably a cognate of Japhug \(-\text{rma}\) “to pass a night in someone’s home.”

However, earlier \(r\)- before nasals also give reflexes as plain syllables. Some known examples are:

- \(\text{mme}^2 \{\text{mej}\}\) “name” is a cognate of Japhug \(\text{ty}-\text{rmi}\).
- \(\text{mme}^2 \{\text{mej}\}\) “hair,” Japhug \(\text{ty}-\text{rme}\).

In three examples, the geminate nasal arises from an earlier preinitial \(s\)-:

- \(\text{nni}^2 \{\text{njiij}\}\) “nose” is a cognate of Japhug \(\text{tu}-\text{sni}\).
- \(\text{nni}^1 \{\text{njiij}\}\) “heart” is a cognate of Japhug \(\text{tu}-\text{sni}\).
- Jacques did not discuss the Burmo-Qiangic etymology of \(\text{nni}^2 \{\text{njij}\}\) “day.” However, it is superposable to Khroskyabs \(\text{sn}a\) “a day,” Geshiza \(\text{ba}-\text{sni}\) “today,” etc. Japhug \(\text{sn}i\) and Tshobdun \(\text{sn}i\), both “day,” are also clearly cognate, though slightly irregular.

However, earlier \(s\)- before nasals also give reflexes as tense syllables. Known examples are:

- \(\text{nni}^2 \{\text{njij}\}\) “nasal mucus,” cf. Japhug \(\text{tu}-\text{snm}\);
- \(\text{nni}^2 \{\text{njij}\}\) “sister of a woman,” cf. Japhug \(\text{tu}-\text{snm}\);
- \(\text{nni}^2 \{\text{njij}\}\) “ear (of grain),” cf. Japhug \(\text{ku}-\text{nom}\).

In four examples, the geminate nasal arises from an earlier preinitial \(y\)- or \(K\)-, probably passing through an intermediate stage as \(k\)-:

- \(\text{maa}^1 \{\text{mjaa}\}\) “ulcer, wound” is a cognate of Japhug \(\text{tu}-\text{ymaz}\) “bleisure,” Proto-Rgyalrong \(\text{ka}-\text{mas}\).
- \(\text{mni}^1 \{\text{njiij}\}\) “two” is a cognate of Japhug \(\text{knu}z\), Proto-Upper-Rgyalrong \(\text{q}a\text{nes}\).
- \(\text{mme}^2 \{\text{mja}\}\) “fire” is a cognate of Japhug \(\text{smi}\). The Tangut form itself corresponds with a West Rgyalrongic dialectal root with \(\text{y}\)-: Khroskyabs \(\text{sm}a\), Stau \(\text{ym}a\), Geshiza \(\text{wma}\), which induced the uvularity compression \(\text{ka}^u\text{ma}^u < \text{k}^u\text{ma}^u\).
- Jacques did not discuss the Burmo-Qiangic etymology of \(\text{mme}^1 \{\text{mjiij}\}\) “corpse.” However, it is cognate to Khroskyabs \(\text{jm}a\), thus reflecting a proto-form akin to Proto-Rgyalrong \(\text{ka}^u\text{ma}^u\). This root is clearly ancient,

---

cf. Proto-Lolo-Burmese (Matisoff) *maŋ, etc. Although Matisoff\(^{62}\) considers the Proto-Sino-Tibetan form to have a preinitial *s-, it can be argued that the Khroskyabs preinitial j- < *ka- is earlier. Matisoff’s *s- forms could instead be regarded as a later composition with the pan-Sino-Tibetan verb “to die,” Chinese 死 sijX, etc., cf. Lôngchuân Ngochang ʂɿ˥ ˥ mʐuaŋ˥ ˥ “corpse”.\(^{63}\)

No exceptions to this correspondence are known. The presence of uvularity compression in mməʶ¹{məə¹} “fire” but the lack of it in mma¹{mjaa¹} “ulcer, wound” and mmə¹{mjii¹} “two” should nevertheless be noted.

Finally, in two examples the extant modern Rgyalrongic comparanda do not permit the identification of the Pre-Tangut preinitial in question.

- 𗫴₂₄₃₆ mma¹{mjaa¹} “fruit” is a cognate of Japhug uu-mat “its fruit.” Among modern Rgyalrongic languages, this etymon is only attested in Upper Rgyalrong: Tshobdun tê-me (tu⁵⁵me³³, DB-caob-0318), Zbu va-mêt, always without a preinitial.
- 𗫐₂₁₂₈ mma²¹{məə³¹} “to blow” is a cognate of Japhug kɤɣɤmɯt. This verb is a deverbal from a noun reflected as Japhug tɤmɯt “exhaled breath.” However, the deverbal formant is the *p- one in Upper Rgyalrong: Japhug kɤɣɤmɯt, Tshobdun kewémo (ka³³wa⁴⁴mo³³, DB-caob-1316), Zbu ka-vamŋʷt. In Geshiza, a West Rgyalrongic language like Tangut, one finds wma < *yŋmə instead. It is difficult to tell if the Tangut verb reflects the *p- deverbal in Upper Rgyalrong of the *k- deverbal in Geshiza.

As a conclusion, the geminate nasal in Tangut unambiguously indicates the existence of a preinitial in Proto-Rgyalrongic. However, the same Proto-Rgyalrongic preinitial before a nasal initial can lead either to a geminate nasal or a different result. There is no obvious solution to this problem. For *l- and *s-, nonetheless, one might tentatively suggest a preference for the geminate reflex in open syllables (and quasi-open syllables with *ŋ), and a preference for tense reflex in close syllables. Compare, for example, 𘎞₅₇₀₀ nŋi²{njii²} “nose,” which has a geminate nasal and derives from earlier *-a, cf. Japhug tu-uŋa, with 𘎞₅₇₃₁ na²¹{na¹} “nasal mucus,” which has a tense vowel and derives from earlier *-ap, cf. Japhug tu-uŋaβ.


\(^{63}\) Huáng Bùfán, Xú Shòuchún, Chén Jiāyīng, Wáng Huìyín eds., *Zàngmiǎnyǔzú yǔyán cíhuì* [A Tibeto-Burman Lexicon] (Beijing: Zhongyang minzu daxue chubanshe, 1992), 54.
4.3.5 Unexplained Comparison

One remaining comparison needs to be discussed. 𗧜₂₈₀₁ nlhu⁰² {lhuu²} “marrow” is considered by Jacques in the *Esquisse* (p. 53) to be cognate to Japhug *tuu-pju*. If the words are indeed cognate, the most likely proto-form would be approximatively *-maal*. This correspondence between Tangut *nlh-* and Modern Rgyalrongic *pj-* remains unattested elsewhere.

4.3.6 Cases Not Discussed in the *Esquisse*

One etymology not proposed by Jacques in the *Esquisse* merits some consideration.

- The stem-alternating verb 𘃟₀₇₅₀ ngii¹ {gjii¹} / 𘁂₁₂₄₉ nggø¹ {gjoo¹} “to chew, to bite” is considered by Gong Hwang-cherng⁶⁴ to be borrowed from Chinese 齧ŋet “to bite, to gnaw.” However, I argue that shows that it reflects in fact a pan-Rgyalrongic etymon *-nəka*, cf. Japhug *kɤ-nɤŋka*, the semantics of which have been bleached to “to eat” in Modern Stau-Horpa, cf. Stau and Geshizha ŋə.⁶⁵

The discovery that column-2 rhymes could reflect the Rgyalrongic autobenefactive derivation also allows us to understand the origin of the verb 𗷆₄₄₈₉ mphi¹ {phjii¹}. It is clearly related to the stem-alternating verb 𘃟₀₇₄₉ phi¹ {phji¹} / 𘁁₄₅₆₈ pho² {phjo²}, which is used as an unmarked causativiser “to make, to order.” The most common meaning for the form with nasal preinitial 𗷆₄₄₈₉ mphi¹ {phjii¹} is a more specific one, “to send someone as representative,” cf. (1). It is interesting to note that a common alternative verb in the same context, 𘒫₅₈₇₁ nziw² {zeew²}, also has a nasal preinitial.

(1) 𘃟 𗸈 𘅳 𘟙 𘋩 𘎪 𘁂 𘃟
    dzwo² mphi¹ tshi¹ ne² do² ntshe¹ phi¹
             person send Qí king POST speak make

“He sent someone to the King of Qí to tell (the story).”⁶⁶

The other frequent meaning of 𗷆₄₄₈₉ mphi¹ is “to employ someone as a ser-

---

⁶⁴ Gong, “Xīxiàyǔ zhōng de Hànyǔ jièci.”
Vant,” cf. (2). A common word for the job of “servant,” indeed, is the nomen patientis of this verb, 翦 翦 mphi¹-le w² {phjii¹-lew²}.

(2) thə² ta¹ mo² dzə¹ dzo¹ rar²-mphi¹ qSe² ni² mphi¹ ma¹-wə²-na²

this TOP I time long PFV-order you order NEG-can-2SG

“I had these people as servants (lit. ordered these people) for a long time; you’ll not be able to work with them (lit. order them).”67

Both uses of the preinitialed verb 翦₄₄₈₉ mphi¹ {phjii¹} can be understood as the effect of an autobenefactive prefix. Compared to the unprefixed 翦₀₇₄₉ phi¹ {phji¹} / 翦₄₅₆₈ pho² {phjo²}, which has a general causative meaning of “to make, to order,” both common meanings 翦₄₄₈₉ mphi¹ {phjii¹}, whether “to send someone as a representative”—to represent oneself—or “to engage someone as a servant”—i.e., in one’s own service—strongly imply that the subject of the verb is a beneficiary, and thus can be regarded as autobenefactive derivations from the unprefixed base verb.

Another potential example of the autobenefactive prefix is 翦₅₄₃₅ n.wi¹ {wee¹} “to be born.” This word could be an autobenefactive derivation from 翦₂₂₂₆ wi² {we²} “to become” (for its etymology cf. §4.3.3), parallel to Geshiza ʑæ “to be born,” probably autobenefactive from ʑæ “to come.”

4.3.7 The Origins of the Tangut Nasal Preinitial

In conclusion, the Tangut nasal preinitial seems to have two principal origins:

• Before nasals, it reflects the result of the assimilation of any Pre-Tangut preinitial to a geminate nasal: NN < CN;
• Before other consonants, it reflects a Pre-Tangut dental preinitial n-.

The nasal preinitial hypothesis settles some etymological problems and opens up fruitful possibilities for further etymological research, especially with regard to Tangut reflexes of the pan-Rgyalrongic autobenefactive derivation. On the other hand, the evolution of Proto-Rgyalrongic preinitials in Tangut68,

---

67 Inv N° 616:5, cf. ibid., 29.
which already has a rather chaotic picture, is further complicated by the non-obligatory geminate nasal assimilation. Further research is needed to elucidate the specific conditions of the nasal assimilation, preferably based on further etymological proposals.

5. CONCLUSION

This essay proposes the nasal preinitial hypothesis, recapitulated in Table 5, according to which syllables having a column-2 rhyme have a nasal preinitial instead of a long vowel as Gong Hwang-cherng proposed, whereas column-1 syllables do not.

Table 5: The nasal preinitial hypothesis.

<table>
<thead>
<tr>
<th>initial classes</th>
<th>Column-1 rhyme short vowel → absence of nasal preinitial</th>
<th>Column-2 rhyme long vowel → presence of nasal preinitial</th>
</tr>
</thead>
<tbody>
<tr>
<td>voiced</td>
<td>Tangut character</td>
<td>revision Note</td>
</tr>
<tr>
<td>voiced</td>
<td>bu⁰¹ → bu⁰¹ “cattail”</td>
<td>buu⁰² → mbu⁰² borrowed from Chinese Héxi *mb- &lt; muH</td>
</tr>
<tr>
<td></td>
<td>no revision homorganic nasal preinitial: mb-, nd-, ng-, ndz-, ndź-</td>
<td></td>
</tr>
<tr>
<td>voiceless</td>
<td>pho² → pho² “to make”</td>
<td>phoo² → mpha² Tibetan transcription as ḷpho &lt;’pho&gt;</td>
</tr>
<tr>
<td></td>
<td>no revision homorganic nasal preinitial: mp(h)-, nt(h)-, nj(h)-, nts(h)-, nts(h)-</td>
<td></td>
</tr>
<tr>
<td>nasals</td>
<td>ma⁰¹ → ma⁰¹ “mother”</td>
<td>maa⁰¹ → mma¹ &lt; *jma, cognate to Japhug tɯ-ɣmaz</td>
</tr>
<tr>
<td></td>
<td>no revision geminate nasal preinitial: m-, nn-, nj-, nn-</td>
<td></td>
</tr>
<tr>
<td>sibilants</td>
<td>ši¹ → ši¹ “before”</td>
<td>šjii¹ → nši¹ cognate to Khroskyabs ņš̂</td>
</tr>
<tr>
<td></td>
<td>no revision dental nasal preinitial: ns-, nʃ-, nz-, nź-</td>
<td></td>
</tr>
<tr>
<td>glides</td>
<td>wa⁰¹ → wa⁰¹ “fur jacket”</td>
<td>waa⁰¹ → nwa⁰¹ cognate to Geshizha nva</td>
</tr>
<tr>
<td></td>
<td>no revision dental nasal preinitial: n·w-, nj-</td>
<td></td>
</tr>
<tr>
<td>gutturals</td>
<td>ka⁰¹ → ka⁰¹ “door”</td>
<td>nx-/nx-, ny-/ns- and n- do not exist (cf. §4.2.3)</td>
</tr>
</tbody>
</table>
The “long vowel” problem has remained an open question in Tangut scholarship ever since Gong Hwang-cherng’s article “A Hypothesis of Three Grades and Vowel Length Distinction in Tangut”. It is a testimony to the power of Sino-Tangutica in the reconstruction of Tangut phonology that a definitive solution to this problem can only come from a thorough examination of Chinese-to-Tangut materials, especially from comparing different chronological layers of borrowings.

The revision of 鷺₃₃₂₃ “goose” from ɴɴʕa² to ɴɢʕa² (§4.2.2) draws attention to other potential misreadings of fànqiè spellings from the Combined Edition of Wénhài and Homophones. This work is crucial in the reconstruction of Tangut phonology, preserving the only testimony of the pronunciation of a large number of shǎng-toned characters. Its unique importance, unfortunately, is rivaled only by its paleographic difficulty. It is hoped that further insights into the rules and mechanism of Tangut fànqiè might bring even more emendations to the pronunciation of individual Tangut characters.

Additional research is called for primarily in two directions: a systematic treatment of Tangut fànqiè behavior, especially with regard to the less well-behaving nasals and Sanskrit transcription characters (§4.2), and an investigation of the specific conditions of preinitial assimilation before nasals (§4.3.4).

Among the rhymes assigned by Gong Hwang-cheng with a “long vowel,” i.e., to column 2, the rhymes R.21 (1.21–2.18) and R.59 (1.57) are shown to be unrelated to the phenomenon discussed in this article. Their nature, as well as that of R.60 (2.50), will be addressed in forthcoming articles. The same remark applies to “long vowel” rhymes outside the major cycle. I consider “long vowel” rhymes in the second (R.80–R.98) and third minor cycles (R.99–R.103) to be unrelated to the “vowel length” distinction discussed in this article. In my opinion, only after a thorough revision of the major cycle could the reconstruction of the minor cycles be updated through Gong Hwang-cherng’s method of phonological alternation.

REFERENCES


