Studies in Pyu phonology, I
Onsets

Marc Miyake

The extinct Pyu language was spoken during the first millennium CE and the early centuries of the second millennium CE in what is now Upper Burma. It has been classified as Sino-Tibetan on the basis of basic vocabulary, but its precise position within the family remains unknown. It survives in inscriptions in an Indic script. In this study, the first of its kind, I begin to reconstruct Pyu phonology on the basis of spellings in those inscriptions. I propose that Pyu was a sesquisyllabic language with 7 preinitials and 43 or 44 initials.

Keywords: Pyu, Sino-Tibetan, phonology, reconstruction, sesquisyllabic

1. Introduction

The extinct Pyu language was spoken during the first millennium CE and the early centuries of the second millennium CE in what is now Upper Burma. It has been classified as Sino-Tibetan on the basis of basic vocabulary, but its precise position within the family remains unknown.

Pyu survives only in two types of written records: (1) inscriptions, primarily on stone, in an Indic script; and (2) transcriptions in Chinese historical records. The latter are few in number and difficult to interpret, so I shall postpone their study until after laying down a foundation based on the former which are much more numerous. This study, the first of its kind, is based solely on Indic-script Pyu texts apart from a reference to a single crucial Chinese transcription (§ 7.3.12). All texts are numbered according to the inventory in Griffiths et al. (2017b).

https://doi.org/10.1075/lali.00077.miy
Language and Linguistics 22:1 (2021), pp. 28–70. issn 1606-822x | e-issn 2309-5067

2. Are Pyu phonology and phonetics recoverable?

It is tempting to take Pyu spellings at face value and assume that all graphemes had Indic-like phonetic values. For instance, one might think the spelling *paṁḥ* 'give' in line 5 of Pyu inscription (7) was pronounced something like [pãh] or even [pə̃h], if one goes so far as to project a Sanskrit-like value of *a* as [ə] onto Pyu.

However, it is unlikely that two unrelated languages like Pyu and Sanskrit (or Pali) had identical phonetic inventories. Moreover, it is also unlikely that an unmodified Indic script would happen to have an inventory of graphemes sufficient for a straightforward phonemic or phonetic representation of a language outside the subcontinental linguistic area. Thus we should expect some degree of adaptation of Indic writing to Pyu: old symbols used in new ways and/or new symbols to write non-Indic sounds.

Sound changes potentially add another layer of complexity for interpreting the Pyu script. An IAST (International Alphabet of Sanskrit Transliteration) or ISO (International Organization for Standardization) 15919-style transliteration of the Indic scripts in use in modern Southeast Asia often bears only a partial resemblance to actual pronunciation: e.g.

- Burmese ကြက် <krak·> [tɕɛʔ] 'chicken'
- Mon ဝ္ဂံ <gvaṁ> [kə̤ʔ] 'to obtain'
- Thai กษัตริย์ <kaṣaťariya˟>[1] [kaˌ satˌ] 'king' (< Skt *kṣatriya-*)
- Lao ຊາຕົກ <jātŏka>[2] [saːᄀᄀ dokᄀᄀ] 'jātaka'
- Khmer ធម៌ <dharma> [tʰoa] 'dharma'

This is partly because the orthographies of those languages are largely historical and do not reflect sound changes that occurred after those languages were first written: e.g. *kr-* > [tɕ] in Burmese, *g-* > [k] in Mon, *dz-* (or *ɟ-*) > [s] in Lao, and *dh-* > [tʰ] in Khmer.

But there are also idiosyncratic uses of symbols unrelated to sound changes: e.g. the Mon use of *anusvāra* <ṁ> for nonnasal segments: [ʔ] in 'to obtain' and [h] and [ɔ] in other words.

If a future linguist were to attempt to reconstruct the phonologies of those languages based solely on their 21st century orthographies and a knowledge of the

---

1. The Thai script has no conjunct consonants or *virāma*s, so I have mechanically transliterated all consonants with final <a> unless accompanied by a vowel symbol. I use <ť> to indicate the letter ต derived from ด <t>, originally used for an earlier Thai implosive [ɗ] (now [d]). <˟> transliterates the ทัณฑฆาต *thanthákhâat* letter silencer.

2. I transliterate Lao as ຕ <t> since it is derived from Indic <t> even though it actually represents Lao [d] from an earlier implosive [ɗ]. The Lao ໄม้ ก็ ๅ *mâi kòng* vowel symbol ◌ົ<ŏ> is a Lao innovation. Like the Thai script, the Lao script has no *virāma*s, so Lao final [k] is written as ກ <ka>.

scripts' Indic origins, the results would be far from accurate. Errors introduced by an ignorance of sound changes and orthographic idiosyncracies would be compounded by the inability to reconstruct the tonal contours of Burmese, Thai, and Lao,[3] or the registers of Mon.

The task facing me today is not unlike the assignment of that hypothetical future linguist. How much of Pyu phonology and phonetics can be recovered solely from the script? Although the answer is certainly not "all", it is not zero either.

Suppose one were to try to recover English phonology from English spelling. Unless one had access to variant spellings like <da> for <the> implying a voiced pronunciation of <th>, the use of <th> for both voiceless /θ/ and voiced /ð/ would go undetected. And no amount of variant spellings would permit the reconstruction of unaspirated and aspirated allophones of English voiceless stops. Nonetheless, one could still make phonotactic observations going beyond a mere listing of graphemes: e.g. the maximal syllable structure is CCCVCCC as in <strengths>, /s/ cannot precede voiced obstruents, the cluster /srt/ is not possible, etc.

My goal is to make such observations about Pyu – to recover what Nishida Tatsuo called *sonus grammae*, the sound system implied by a writing system (Yabu 2014). This implied system can only be a part of the lost whole, yet it is preferable to nothing.

3. Methodology

3.1 Corpus

All known legible Pyu texts have been transliterated by Arlo Griffiths, Julian K. Wheatley, and Marc Miyake. This corpus is available online at Griffiths et al. (2017a), a website which is being continuously updated.

Some texts have subscript final consonants; others do not, and one of the copies of the ဂူပြောက်ကြီး <gūprok·krīḥ> Kubyaukgyi (often anachronistically called မြစေတီ <mracetī> Myazedi) inscription (8) only has subscript final consonants in its first three lines. The presence or absence of subscript final consonants does not seem to correlate with geography or chronology.

---

3. It is possible to reconstruct tonal categories to the degree that they are indicated in the script by tone markers, but it is not possible to reconstruct mergers or splits: e.g. the merger of Thai tones B2 (written with ไม้เอก *máy èk*) and C1 (written with ไม้โท *máy thoo*) or the three-way split of the unmarked Lao tone A. That of course assumes tone markers are properly identified as such. A future linguist unaware of Southeast Asian phonological typology might guess that they were consonant or vowel diacritics.

There are only seven texts that have dates which are only *termini post quem*. Two are the Kubyaukgyi inscription (7) and (8) mentioning a lunar year corresponding to 1112 or 1113 CE and the မြစ်သား <mrac·sāḥ> Myittha inscription (39) mentioning 1078 CE. Those three texts have unusual spelling characteristics that may reflect geographical variation as well as sound changes in what Shafer (1943: 316) called Late Pyu.[4] Most Pyu texts are from the ruins of the city now known as သရေခေတ္တရာ <sarekhettarā> Sriksetra, but orthography otherwise does not seem to vary with geography: e.g. an extreme geographic outlier, the တုံးတော် <tuṁḥtau> Tondaw inscription (23) in what is now Rakhine State, has no unique spelling characteristics.

The remaining four inscriptions with dates (3), (4), (5), and (6) refer to a Pyu calendar whose interpretation is uncertain. I provisionally regard all but one of these inscriptions as being in what Shafer (1943: 316) called Old Pyu since they lack the orthographic characteristics of Late Pyu in inscriptions (7), (8), and (39). See Shafer (1943: 356–357) for a proposal to distinguish between Old Pyu and Late Pyu on the basis of differences in word order. Shafer also speculated on the possibility of phonetic differences between the two stages, and I shall present evidence for his speculation below (§ 7.3.6).

The only undated inscription that may not be in Old Pyu is the ဥသျှစ်ပင် <°u shyac· paṅ·> Ouk Shit Pin sculpture inscription (37) whose orthography may be transitional between Old and Late Pyu. I provisionally consider (37) to be in Middle Pyu (§ 7.3.8).

The last Pyu inscription may be the bilingual Chinese-Pyu Tharaba gate inscription (11) which must postdate the first Mongol invasions of Burma in 1277 CE.

3.2 Conventions

All quoted Pyu forms are followed by their source's inventory number from Griffiths et al. (2017b) in parentheses. Arabic numerals following inventory numbers refer to lines. Roman numerals following inventory numbers refer to sections of texts. Letters after inventory numbers refer to faces on an inscription. Letters for odd-numbered faces are capitalized: e.g. A for the first face but not b for the second. Citations are not comprehensive; only one example is given per form unless I am discussing frequency. Old Mon and Old Burmese forms in Pyu multilingual texts are also cited by inventory numbers using the same conventions.

---

4. Shafer used the term "Late Pyu" to refer to the language of the Kubyaukgyi inscription (7) and (8). He was unaware of the Myittha inscription (39) discovered in November 2013.

Pyu transliterations are in an ISO 15919-style romanization for Sanskrit with one modification: middle dots follow subscript consonants for codas and precede transliterations of the subscript dot (*ṃ*, distinct from *anusvāra*), *anusvāra* (*ṁ*), and *visarga* (*ḥ*): e.g. *del·ṃṁḥ* (16.4A) has the final consonant *l*·.

I italicize Pyu transliterations but distinguish between italics for conventional non-IPA transcriptions and angle brackets for transliterations of all languages other than Pyu: e.g. Thai ทัณฑฆาต *thanthákhâat* <dăṇaḍaghāta> (cf. IPA [tʰan˦ tʰa˥ kʰaːt˥˩]).

3.3 Scope

The Pyu lexicon has two major components: Indic and non-Indic. The Indic loans are principally if not entirely from Sanskrit and Pali. The possibility of loans from non-Pali Middle Indic varieties (i.e. Prakrits) cannot be discounted. These loans are in a mix of etymological and nonetymological, presumably nativized spellings which are both quite unlike those of the rest of the lexicon and hence deserving of a separate study. I shall only refer to those loans on occasion whenever they shed light on native Pyu phonology.

The non-Indic component is almost certainly not entirely native, as there is at least one potential Mon loanword (*tha* 'golden', possibly /tʰar/ with an unwritten /r/; 7.10).[5] There may be other borrowings that have so far eluded detection.

Indic loans are usually polysyllabic, whereas the majority of non-Indic morphemes that have been identified so far, mostly by Blagden (1911; 1919) and Griffiths et al. (2017b), are written as monosyllabic *akṣara*s. I shall argue in § 5 that at least some of those written monosyllables represented spoken sesquisyllables.

This study is based on an Excel file of all 1,702 unique *akṣara*s found in the Pyu corpus (Miyake 2017b). Many texts are poorly preserved, so 421 (24.7%) of those *akṣara*s are partly illegible with editorial restorations whenever possible in

5. The corresponding Old Mon word is a noun <thar·> 'gold'. The Old Mon and Pyu words were first identified by Blagden (1909; 1911), who was also the first to propose that the Pyu word was a loan from Mon. The fact that the etymon is attested in Nyah Kur (Diffloth 1984: 137) far to the east of Pyu may indicate that it was in Mon prior to contact with the Pyu in the west. However, the word is not attested outside the Monic branch of Austroasiatic, so it may not be of Austroasiatic origin. Pyu and Mon may have independently borrowed it from a third language, possibly a substratum language of the region. Benedict (1972: 9) saw a "correspondence of 'loan-word' type" between Pyu *tha* 'golden' (which he regarded as a noun 'gold') and Karen /tʰaʔ/ 'iron', but I reject the comparison on semantic grounds. A reviewer proposed that the Old Mon word may be associated with written Tibetan *gser* 'gold' which Jäschke (1881: 590) derived from Persian *zar* 'gold'. I think the resemblance between these words is fortuitous, as there is no known sound change of **s* or **z* to *th* in either Pyu or Old Mon. Nor am I aware of any other known loanwords of Tibetan or Persian origin in Pyu or Old Mon.

parentheses or have uncertain readings indicated by brackets and/or capital C and V for unknown consonants and vowels. Such *akṣara*s may only be trivially unique because of the patterns of their illegibility: e.g. an uncertain *[pa]* (8.24) may be identical to a definite *pa* (8.14), and *pVṅ·* (32.4) might have been identical to an attested *paṅ·* (16.1A) rather than an otherwise unattested †*piṅ·*, †*puṅ·*, †*peṅ·*, or †*poṅ·*. The dagger symbol (†) indicates expected but unattested forms. I have excluded partly illegible and uncertain *akṣara*s from my analysis unless their definitely legible portions contain truly unique patterns.

3.4 Asemantic phonology

Proper phonology requires a knowledge of semantics to identify minimal pairs. If a text in an unknown language contains <pa> and <ba>, we cannot assume that /p/ and /b/ are distinct phonemes unless we know that <pa> and <ba> are semantically distinct: i.e. that they were different words. Ideally, a minimal pair of <pa> and <ba> would be attested in the same phonemic environment. Otherwise, it is possible that <ba> reflects a sandhi variant of /p/ after voiced phonemes as in Korean. Of course, it is difficult to find matching environments for minimal pairs in running texts as opposed to elicited data.

Unfortunately, the vast majority of words in the running texts of the Pyu corpus are unidentified. We do not even know where most words begin or end because the Pyu script has no word spacing. The only clear units are *akṣara*s which may or may not correspond to morphemes. There is a strong correlation between *akṣara*s and non-Indic morphemes in the comprehensible fraction of the Pyu lexicon as glossed by Blagden (1919) and Griffiths et al. (2017b), but it is not absolute: e.g. Old Pyu *tar· dav·ṃḥ* 'king' (27.4) and Late Pyu *mayaḥ* 'wife, consort' (7.3), possibly a loan from Old Burmese *mayā* 'id.', contain two *akṣara*s and may only be partly analyzable.[6] When I am unable to identify words or morphemes, I shall use the terms "*akṣara*s" to avoid judging the semantic status of monosyllables and sesquisyllables. The term "*akṣara*s" is also less unwieldly than "monosyllables and sesquisyllables".

I shall attempt what I call "asemantic phonology": the identification of phonemes on distributional grounds without reference to semantics. I shall provisionally regard any two graphemes in the same position in an *akṣara* as phonemes unless there are distributional and/or typological reasons for doubt. For concrete

---

6. The *ma* of 'wife' may be 'female' (see the discussion of Proto-Lolo-Burmese *m-ya² 'wife' at the Sino-Tibetan Etymological Dictionary and Thesaurus (Matisoff 1987–Ongoing)), and the *dav·ṃḥ* /ðaʍ/ of 'king' may be cognate to written Burmese <tau> < *d- 'suffix indicating sacred, royal, or official status' and/or Old Chinese 主 *toʔ 'master' (Baxter & Sagart 2014b: 157), but the other halves of those words remain unidentified.

examples of this methodology, see § 6.2 and § 7.4 in which I reject literal interpretations of the script that would require postulating many dubious phonemes.

Asemantic phonology is inevitably even less reliable than the pursuit of the *sonus grammae* (§ 2) of a language with known semantics. An asemantic phonologist studying early transcriptions of Hawaiian would not be able to determine that, for instance, <p> and <b> both represented /p/, as both letters occurred in identical environments.[7] Still, an asemantic phonologist would correctly determine that Hawaiian has a (C)V syllable structure.

4. From *akṣara*s to syllables

Before delving into the arrangement of phonemes in Pyu syllables, I shall describe the arrangement of graphemes into *akṣara*s in the Pyu script. I shall explain how these graphemes correspond to consonant phonemes in § 5–7.

4.1 *Akṣara* structures

4.1.1 *Structure of akṣaras without independent vowel symbols*

Nearly all Pyu *akṣara*s are built around a *Ca* consonant symbol with an inherent *a* vowel. Other symbols within an *akṣara* are either attached to this *Ca* symbol or are in 'orbit' around it. The left-to-right arrangement below roughly follows the vertical arrangement of graphemes within an *akṣara* turned 90 degrees counterclockwise[8] rather than the transliteration. Optional components are in parentheses.

$(\dot{m})(r)C_1a(C_2)(C_3)(V)(\underline{m})(\underline{h})(C\cdot)$

Two exceptions to that formula are a single instance of double vowel marking (*pṛi*; 32.5) and a single instance of double subscript consonants (*rlar·r·*; 55).

---

7. English speakers sometimes perceived unaspirated Hawaiian [p] as voiced (Schütz 1994: 81). There was a tendency to write Hawaiian /p/ as <b> intervocalically prior to the standardization of Hawaiian orthography (Schütz 1994: 80) which might lead an asemantic phonologist to conclude that Hawaiian had only /p/ but also had a rule of intervocalic voicing. There is no such rule in modern Hawaiian, though the phenomenon might have existed in the late 18th century.

8. Vowel symbols that cancel the inherent *a* of C_1a vary in position: *ā, i, ī, e, ai, o*, and *au* are atop the first consonant symbol (superscript *r-* or C_1a), whereas *u, ū*, and *ṛ* are below the last consonant in a stack. *Visarga* is always to the right of C_1 or an independent vowel symbol.

4.1.2 *Structure of akṣaras with independent vowel symbols*

Other exceptions to the formula in § 4.1.1 have independent vowel symbols (°*V* = °*a*, °*i*, °*o*, and possibly °*ū*;[9] § 7.3.1) instead of C_1a:

(*ṁ*)°*V*(*ḥ*)(*C*·)

The transliteration symbol ° differentiates independent vowel symbols from the inherent *a* of *Ca* consonant symbols and the dependent vowel symbols *i*, *o*, and *ū*. I do not intend that symbol to imply that the independent vowel symbols share a common graphic component or that they are combinations of such a component with the dependent vowel symbols.

Unlike *Ca* consonant symbols, independent vowel symbols cannot be attached to consonant symbols (superscript *r*- and conjunct C_2 and C_3), though they may have unattached subscript consonants (*C*·) beneath them.

Independent vowel symbols never have subscript dots beneath them since subscript dots modify consonant symbols to indicate fricatives.

4.2 Inventories of graphemes in each position

Only positions with more than one possible grapheme are listed below. Graphemes are displayed in an arrangement based on the Sanskrit phonemic inventory table in Bucknell (1994: 73).

4.2.1 C_1a

There are 28 possible C_1a (Table 1). This does not mean that Pyu had 28 initial or preinitial phonemes, as consonant graphemes do not necessarily match consonant phonemes (§ 7.1). C_1a unique to Middle and Late Pyu are in parentheses.

Table 1. Pyu C_1a graphemes

					ha		
	ka	*kha*		*ga*	*gha*	*ṅa*	
	ca	*cha*		*ja*		*ña*	*ya*
	ṭa			*(ḍa)*			*ra, (ḷa)*
sa	*ta*	*tha*		*da*	*dha*	*na*	*la*
	pa	*pha*	*ḅa*	*ba*		*ma*	*va*

9. Previously published tables of the Pyu script (e.g. Tha Myat 1963: 1) contain more independent vowel symbols. However, symbols such as °*e* that are not listed here are absent from Pyu-language texts in the Pyu script, though they may be present in Sanskrit and Pali texts written by the Pyu in special scripts for Indic languages distinct from the script they used for their own language. The conflation of these scripts as a single "Pyu script" is common in the literature.

All standard Sanskrit and Pali C_1a are represented with the exceptions of:

- the voiced aspirates *jha* and *bha*
- the retroflexes *ṭha, ḍha*,[10] *ṇa*, and *ṣa*
- the palatal fricative *śa*

ḍa is unique to Late Pyu (7, 8, and 39; § 7.3.6).
ḷa is unique to Middle Pyu (37; § 7.3.8).[11]

There is one C_1a not found in Indic: *ḅa*, which looks like *ba* with a stroke through the middle (Figure 1). In inscription (12), Pyu *ḅ* has a dot in the center like Old Mon *ḅ* (Figure 2). Blagden (1911: 368) regarded this character as "evidently related to and perhaps borrowed from the Talaing [i.e. Mon] *ḅ*". I discuss *ḅ* further in § 7.3.12.

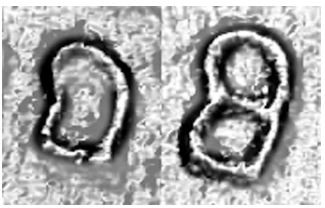

Figure 1. Pyu *ba* (7.15) and *ḅa* (7.15). Extracted from a photograph by James Miles licensed under a Creative Commons Attribution 4.0 International license (CC BY 4.0)

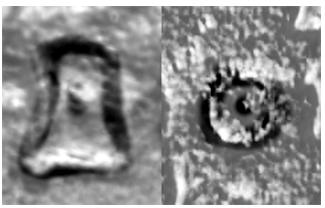

Figure 2. Pyu *ḅa* (12.3) and Old Mon *ḅa* (7.2). Extracted from photographs by James Miles licensed under a Creative Commons Attribution 4.0 International license (CC BY 4.0)

10. Griffiths et al. (2017a) read Blagden's *ḍh* in the Kubyaukgyi inscription (7) and (8) as *p*. Shafer (1943: 316) expressed doubts about *ḍh* and considered reading it as *p*, but ultimately retained Blagden's reading.

11. Griffiths et al. (2017a) read Blagden's *l* in C_1 position in the Kubyaukgyi inscription (7) and (8) as *ḍ* or in two instances as a C_1C_2 sequence *kḍ* (7.9 and 8.9). Blagden (1913–1914) suggested that his *de* 'month' in inscriptions (3–6) might also be read as '*ḷe*, or something else'; Griffiths et al. (2017a) read it as *rla* which is closer to other Sino-Tibetan words for 'month' such as written Burmese <la>.

4.2.2 C_2

There are 28 possible C_2 (Table 2). C_2 unique to Middle and Late Pyu are in parentheses.

Table 2. Pyu C_2 graphemes

					h		
	k			*g*		*ṅ*	
	c	*ch*		*j*		*ñ*	*y*
ṣ	*ṭ*	*(ṭh)*		*(ḍ)*			*r, (ḷ)*
s	*t*	*th*		*d*	*dh*	*n*	*l*
	p	*ph*	*ḅ*	*b*		*m*	*v*

The only C_1 graphemes that lack C_2 counterparts are:

– the voiceless aspirate *kh*
– the voiced aspirate *gh*

The only C_2 graphemes that lack C_1 counterparts are both retroflexes:

– the stop *ṭh* which only occurs once under an unknown consonant (37; § 7.4.6)
– the fricative *ṣ* which only occurs once under *k* (56; § 7.4.14)

ḍ is unique to Late Pyu (7, 8, and 39; § 7.3.6).
ṭh and *ḷ* are unique to Middle Pyu (37; § 7.3.8).[12]

4.2.3 C_3

There are only 3 possible C_3: *ñ* which appears twice under *rj-* in a possible Indic loanword on a pottery fragment (161; § 7.4.4), *r*, and *l*.

4.2.4 *V*

There are 9 vowel symbols (Table 3). All symbols for Sanskrit and Pali vowels are represented except for long *r̥̄*, *l̥*, and the theoretical long *l̥̄*.[13]

12. Griffiths et al. (2017a) read Blagden's *ḷ* in C_2 position in the Kubyaukgyi inscription (7) and (8) as *ḍ* or in one instance as *l* (7.24).

13. Although there are no actual Sanskrit words with long *l̥̄*, the independent vowel symbol ឦ <l̥̄> represents the Khmer word [liː] 'to hear', and ฦๅ <l̥̄> is an obsolete spelling of Thai ลือ [lɯː] 'to rumor', so *l̥̄* would not be entirely unexpected for [l] followed by a nonrounded high vowel in a Southeast Asian context.

Table 3. Pyu *V* graphemes

a	*ā*		
i	*ī*	*e*	*ai*
r̥			
u	*ū*	*o*	*au*

Table 3 excludes the aforementioned single case of double vowel marking: *pṛi* (32.5).

4.2.5 *C·*

There are only 10 possible *C* (Table 4).

Table 4. Pyu *C·* graphemes

k·	*ṅ·*	
		y·
		r·
t·	*n·*	*l·*
p·	*m·*	*v·*

Table 4 excludes the aforementioned single case of double subscript consonants: *rlar·r·* (55).

5. (Sesqui) syllabic structure

If I simply mapped Pyu graphemic structure onto Pyu syllable structure, I could claim that each *akṣara* represented a Pyu syllable. However, improbable consonant clusters in the script suggest that at least some *akṣara*s could not represent monosyllables. Blagden (1919: 60) was the first to propose that Pyu *akṣara*s might in fact represent what Matisoff (1973) was the first to call "sesquisyllables" (emphasis mine):

> It [Pyu] allowed a limited number of combinations of two consonants as initials, the second member of such combinations being a semi-vowel or a liquid. But even in such cases it is not quite certain that a **short neutral vowel** may not have been inserted in between the two consonants, for **the spelling often uses conjuncts in combinations (such as *td*) where such a vowel must of necessity be introduced.** It shares this peculiarity with the contemporary Môn usage, a usage that has survived in modern Môn also.

It would not be surprising if Pyu shared an areal trait of sesquisyllabism with Mon and Mon-Khmer languages in general as well as with modern Burmese whose "minor syllables" are presyllables.

The fact that the Pyu script allows three-consonant initial sequences ($C_1a(C_2)(C_3)$) absent from the small corpus Blagden had at hand would seem to strengthen the case for sesquisyllabism, but as I shall explain in § 7.3.10 below, in all or nearly all cases, such graphemic sequences represented only two phonemic consonants. Hence Blagden was still correct about Pyu having "allowed a limited number of combinations of two consonants as initials", albeit at the phonemic rather than the graphemic level. But he was not correct about "the second member of such combinations being a semi-vowel or a liquid", for his own readings contain counterexamples which remain in the current readings of Griffiths et al. (2017a): e.g. *td* which he mentioned later in that very paragraph.

The graphemic sequence *td* and its velar and labial counterparts *kg* and *pb*[14] are not absolute proof of sesquisyllabism. In theory they could have been digraphs for single phonemes lacking Indic script symbols such as implosive /ɗ ʄ ɓ/ or tense /t̼ k̼ p̼/ as in Korean rather than ways to write sequences like [CəC]. I doubt that was the case because of their frequency and distribution.

pb occurs only once in the corpus (*pban·ḥ* '?' 20.3) and hence is an extremely implausible candidate for a digraph representing a consonant phoneme.

kg is a better candidate, as it occurs in 5 unique *akṣara*s: *kgaṃḥ* '?' (12.1), *kgam·* '?' (20.5), *kgin·ṃṁḥ* '?' (20.2), *kgoṃ* '?' (25.4), and *kgoy·* '?' (27.1).

td is the best candidate of the three, as it occurs in 29 unique *akṣara*s, including the high-frequency word *tdav·ṃḥ* 'king' (12.1; also spelled *tdaṃḥ* in texts without final consonants such as 3).

A frequency-based argument against a unit phoneme interpretation of the common character sequence *td* will not work, and such an argument would be shaky in the case of *kg*. There is, however, a stronger line of argument that applies equally to all three.

If *td, kg*, and *pb* represent unit consonants, they should have the same distribution as less controversial unit stop consonants such as *t, k*, and *p* in onsets. But they do not: there are no †*Ctd*, †*Ckg*, or †*Cpb* clusters corresponding to *Ct, Ck*, and *Cp* clusters such as *kt, pk*, and *tp*. Nor are there †*tdC*, †*kgC*, or †*pbC* clusters corresponding to *tC, kC*, and *pC* clusters such as *ty, kr*, or *pl*. Although it is possible that *td, kg*, and *pb* represent unit consonants that could not form clusters, it is simpler to interpret them as presyllable-syllable onset sequences like [CəC], even though such sequences lack parallels in Mon which had a constraint against homorganic stops within sesquisyllables. This argument also applies to *tc* which

14. The palatal and retroflex sequences *cj* and *ṭḍ* are not attested.

does not appear in †*Ctc* or †*tcC* clusters corresponding to *Cc* and *cC* clusters such as *mc* or *cm*.[15]

The aforementioned word *tdav·ṃḥ* 'king' has a disyllabic variant *tar· dav·ṃḥ* (12.1). If *td* represents a presyllable-syllable onset sequence, then *tdav·ṃḥ* would be a sesquisyllabic compression of *tar· dav·ṃḥ*. A monosyllabic compression with clusters such as *td* representing unit consonants cannot be completely ruled out, but I prefer a typologically more plausible solution involving sesquisyllables which are widely attested in Southeast Asia unlike an implosive /ɠ/[16] or tense /t̤ k̤ p̤/.

If *td, kg, pb*, and *tc* represent presyllable-syllable onset sequences, does that mean all written consonant clusters represent such sequences? Or could other written clusters represent spoken clusters (e.g. *ty* = [tj]) or be ambiguous (e.g. *ty* = [tj] ~ [təj])? If Pyu had sesquisyllables, did they have optional monosyllabic pronunciations: e.g. *tyaṅ·* '?' (20.1) = [təjaŋ] ~ [tjaŋ]? Or were there minimal pairs of sesquisyllables and monosyllables with identical spellings: e.g. *tyaṅ·* = /təjaŋ/ and /tjaŋ/? Did one have to know Pyu to know whether any given consonant cluster-initial *akṣara* was pronounced with or without a presyllabic vowel such as schwa?

None of those questions are answerable at this point, though if Pyu verse is discovered in the future, the meter may reveal whether consonant cluster-initial *akṣara*s were sesquisyllables or monosyllables.

To avoid committing to an interpretation of written consonant clusters apart from (near-)homorganic stop sequences, I shall use the agnostic term "preinitial" to refer to the first consonant in a Pyu (sesqui)syllable. A preinitial was probably pronounced as a presyllable with a predictable short neutral vowel after it, if it followed a (near-)homorganic consonant: e.g, /td/ = [təd]. Other preinitials may or may not have been pronounced as presyllables; those that were pronounced as presyllables might have had phonemic short neutral vowels. Sonorant presyllables may have been syllabic: e.g. *mcuḥ* /m.cuh/ '?' (25.2) may have been pronounced [ṃcuh].

Pyu has the following phonemic (sesqui)syllable structure:

/(C).(C)V(C)/

I use a period to separate the preinitial from the initial consonant. If Pyu had minimal pairs of the type /təjaŋ/ : /tjaŋ/, that period could be replaced by (ə):

/(C)(ə)(C)V(C)/

---

15. I discuss this unusual cluster in § 6.2.2.

16. Implosive /ɗ ɓ/ are widely attested in Southeast Asia. In § 7.3.12 I propose that Pyu *ḅ* represents /ɓ/.

There is a single instance of an *akṣara* with two preinitials: *rcṅa(C·)ḥ* (24 bottom). I am hesitant to add a second preinitial to the preceding formulae on the basis of only one example that may be a nonce spelling for a rapid, fused pronunciation of two syllables or a loanword from an unknown language.[17]

6. Preinitials

As in other languages of the region such as Mon, the set of preinitials is a subset of the set of initials (§ 7). I discuss possible preinitial-initial combinations after introducing the initials in § 7.

6.1 Core preinitials

The preinitials in Table 5 are the most common and are hence the least controversial.

Table 5. Pyu core preinitials

	/k./		
			/r./
/s./	/t./	/n./	
	/p./	/m./	

The stop preinitials /k. t. p./ may have had voiced and/or fricative allophones:

- [x] (§ 7.4.1)
- [d] ~ [ð] ~ [θ] (§ 6.2.3, § 7.4.8, § 7.4.9, § 7.4.11, § 7.4.13)
- [b] ~ [β] ~ [v] ~ [w] (§ 6.2.4–6.2.6)

Although one might expect voiced velar allophones [g] ~ [ɣ] to parallel voiced dental allophones [d] ~ [ð] and voiced labial [b] ~ [β] ~ [v] ~ [w] allophones, there are no spellings with preinitial *g* or *gh*.

Conversely, although one might expect voiceless labial fricative allophones [f] ~ [ɸ] to parallel voiceless velar [x] and voiceless dental [θ], there are no cases of *pC-ṃ*-sequences in which *ṃ* might be a diacritic indicating a fricative value for *p*: e.g. †*pṅ-ṃ* whose *ṃ* could modify *p* rather than *ṅ* which cannot be read as a fricative.

Those possible allophonic gaps could be artifacts of the small number of anomalous spellings hinting at the pronunciation of Pyu.

---

17. The cluster *rcṅ-* is unknown in Mon which is the most likely known source of borrowings with complex initials.

6.2 Peripheral preinitials

The phonemic status of the rare preinitials in Table 6 is disputable. I supply transliterations in italics whenever they differ from the phonemic symbols.

Table 6. Pyu peripheral preinitials

			ṅ /ŋ./?	
/c./?				
		/d./?		
	ḅ /ɓ./?	/b./?		*v* /w./?

6.2.1 *ṅ /ŋ./?*

There is only a single instance of *ṅr* in *ṅraḥ* '?' (20.4). /ŋ/ is a typologically unlikely presyllabic initial, as it is absent from presyllables in Mon, Khmer, Old Chinese (Baxter & Sagart 2014a), and pre-Tangut (Miyake 2012) and from Old Tibetan preinitials. *ṅraḥ* may be a nonce spelling for a rapid, fused pronunciation of two syllables or a loanword from an unknown language.[18]

6.2.2 */c./?*

There are only two instances of preinitial *c*: the aforementioned *rcṅa(C·)ḥ* '?' (24 bottom; § 5) and *cmol·* '?' (25.2).

Palatal-initial presyllables are common in Mon and Khmer but are unknown in Old Chinese (Baxter & Sagart 2014a), and pre-Tangut (Miyake 2012). Moreover, Old Tibetan has no palatal preinitials. Thus it is unlikely that Pyu would have palatal preinitials in native words. Like *rcṅa(C·)ḥ, cmol·* may be a loanword, though it could not be from Old Mon which has no †*cmol·*.

Another possibility is that *cmol·* should be read *khmāl·*, as *ca* and *khe* are similar in shape (Griffiths et al. 2017b), and the right half of the symbol for *o* may be *ā*. The *c* in *rcṅa(C·)ḥ*, however, clearly does not resemble *kh*, and *ā* rarely appears in non-Indic Pyu words.

6.2.3 */d/?*

There are only five instances of preinitial *d: dvaṃ* '?' (20.3), *dviC·* '?' (23.3), *dmaC·* '?' (25.1), *dṅok·* '?' (64.3), and perhaps *[dṅey·]ḥ* (the first half of *[dṅey·]ḥ du[r]·* 'footstool' in 16.2) whose *d* is uncertain.

Preinitial *d* may represent a voiced allophone of preinitial /t/ before a voiced consonant or a presyllable like [də] with secondary initial voicing (cf. [də] from

18. The cluster *ṅr-* does exist in written Burmese, but inscription (20) is in Old Pyu and hence may predate the arrival of Burmese in the Pyu-speaking area.

*t-syllables in Burmese words like တံခါး <taṁkhāḥ> [dəgaˀ˥] 'door'). If so, then *dṅok·* (64.3) may have been phonologically identical to *tṅok·* /t.ŋok/ '?' (27.3). Unfortunately, there are no †*tC-akṣara*s in the corpus corresponding to the other *dC-akṣara*s.

6.2.4 */b./?*

There is only a single instance of preinitial *b: bro* (7.25), which Blagden (1919: 67) glossed as "meaning undetermined, but possibly the phrase which it begins contains the idea of 'violence,' 'harm' ".

Preinitial *b* may represent a voiced allophone of preinitial /p/ before a voiced consonant or a presyllable like [bə] with secondary initial voicing (cf. [pə] from *p-syllables in Burmese words like ပုဂံ <pugaṁ> [bəgã] 'Pagan'). Unfortunately, there is no †*pro* in the corpus corresponding to *bro*.

6.2.5 *ḅ /ɓ./?*

There is only a single instance of preinitial *ḅ* which I interpret elsewhere as an implosive /ɓ/ (§ 7.3.12): *ḅnuḥ* '?' (20.5).

Implosives are absent from Mon and Khmer presyllables and are hence unlikely to be in Pyu preinitials. The characters *ḅ* and *b* are only distinguished by the presence or absence of a stroke in the middle. Perhaps *ḅnuḥ* is an error for †*bnuḥ* which might have been phonologically identical to an unattested †*pnuḥ* /p.nuh/. See § 6.2.4 on preinitial *b*.

6.2.6 *v /w./?*

There are only three instances of preinitial *v: vḅo* '?' (39.3), *vra* '?' (39.6), and perhaps *[v]r[el]·* (the first half of *[v]r[el]· ndrom·* 'to be kind' or 'kindness' in 16.2) whose reading is uncertain.

Preinitial *v* may represent a voiced allophone of preinitial /p/ such as [β], [v], or [w] before a voiced consonant or a presyllable such as [βə], [və], or [wə] with secondary initial voicing. The development of Pyu [βr] from *p(ə)r would be typologically parallel to the development of Japhug [βr] from Proto-rGyalrongic *pr (Jacques 2004: 331).

7. Initials

As with the preinitials, I divide initials into two groups: common core initials and rarer peripheral initials.

7.1 Core initials

There are far more initial phonemes in Pyu (Table 7) than there are consonant symbols in the Pyu script. Many of the 43 or 44 phonemes are written with digraphs or with subscript dots. I supply transliterations in italics whenever they differ from the phonemic symbols. Combinations of consonant symbols with a subscript dot are written with *-ṃ*: e.g. *g-ṃ* represents *g* with a dot under it. Initials unique to Middle and Late Pyu are in parentheses. /ʍ/ has two spellings: *hv* for the allophone [ʍ] and *hv-ṃ* for the allophone [f].

Table 7. Pyu core initials

	/°/ʔ	/h/										
	/k/	/kʰ/		/g/		*g-ṃ* /ɣ/	*hṅ* /ŋ̊/	*ṅ* /ŋ/				
hy-ṃ /ç/	/c/	*ch* /cʰ/		*j* /ɟ/		*y-ṃ* /j̥/	*hñ* /ɲ̊/	*ñ* /ɲ/	*hy* /ɟ̊/	*y* /j/		
				(*ḍ* /ɖ/)	(*hḍ* /D/)				*hr* /ɽ̥/	/r/ (*ḷ* /l̥/)	*ṭr* /R̥/	*dr* /R/
/s/	/t/	*th* /tʰ/		/d/		*d-ṃ* /ð/	*hn* /n̥/	/n/	*hl* /l̥/	/l/	*ṭl* /L̥/	*dl* /L/
	/p/	*ph* /pʰ/	*ḅ* /ɓ/	/b/		*v-ṃ* /v/	*hm* /m̥/	/m/	*hv, hv-ṃ* /ʍ/	*v* /w/		

7.2 Combination of core preinitials and core initials

I list all attested combinations of core preinitials and core initials in Table 8. I supply transliterations in italics whenever they differ from the phonemic symbols. Combinations unique to Middle and Late Pyu are in parentheses.

Table 8. Combinations of Pyu core preinitials and core initials

	/k./	/t./	/n./	/p./	/m./	/r./	/s./
/k/	/k.k/	/t.k/	/n.k/	/p.k/	/m.k/	/r.k/	/s.k/
/g/	/k.g/	/t.g/	/n.g/	/p.g/	/m.g/		/s.g/
/ɣ/	*kg-ṃ* /k.ɣ/	*tg-ṃ* /t.ɣ/	*ng-ṃ* /n.ɣ/	*pg-ṃ* /p.ɣ/	*mg-ṃ* /m.ɣ/	*rg-ṃ* /r.ɣ/	*sg-ṃ* /s.ɣ/
/ŋ̊/	*khṅ* /k.ŋ̊/	*thṅ* /t.ŋ̊/	*nhṅ* /n.ŋ̊/	*phṅ* /p.ŋ̊/			

Table 8. *(continued)*

/ŋ/	*kṅ* /k.ŋ/	*tṅ* /t.ŋ/	*nṅ* /n.ŋ/	*pṅ* /p.ŋ/			
/c/	/k.c/	/t.c/	/n.c/	/p.c/	/m.c/		/s.c/
/cʰ/	/k.cʰ/						
/ɟ/	*kj* /k.ɟ/			*pj* /p.ɟ/			
/ɟ̰/	*ky-ṃ* /k.ɟ̰/						
/ɲ̊/	*khñ* /k.ɲ̊/						
(/ɖ/)	(*kḍ* /k.ɖ/)			(*pḍ* /p.ɖ/)	(*mḍ* /m.ɖ/)		
/t/	/k.t/		/n.t/	/p.t/	/m.t/	/r.t/	/s.t/
/tʰ/	/k.tʰ/						*s[th]* /s.tʰ/?
/d/	/k.d/	/t.d/	/n.d/	/p.d/	/m.d/		/s.d/
/ð/	*kd-ṃ* /k.ð/	*td-ṃ* /t.ð/	*nd-ṃ* /n.ð/	*pd-ṃ* /p.ð/	*md-ṃ* /m.ð/		*sd-ṃ* /s.ð/
/n̰/	*khn* /k.n̰/	*thn* /t.n̰/				/r.n̰/	
/n/	/k.n/	/t.n/		/p.n/		/r.n/	/s.n/
/p/	/k.p/	/t.p/	/n.p/			/r.p/	
/pʰ/							/s.pʰ/
/ɓ/	*kḅ* /k.ɓ/	*tḅ* /t.ɓ/	*nḅ* /n.ɓ/	*pḅ* /p.ɓ/			*sḅ* /s.ɓ/
/b/	/k.b/	/t.b/	/n.b/	/p.b/	/m.b/		/s.b/
/v/	*kv-ṃ* /k.v/	*tv-ṃ* /t.v/		*pv-ṃ* /p.v/	*mv-ṃ* /m.v/	*rv-ṃ* /r.v/	
/m̰/	*khm* /k.m̰/	*thm* /t.m̰/	*nhm* /n.m̰/	*phm* /p.m̰/			
/m/	/k.m/	/t.m/	/n.m/			/r.m/	/s.m/
/j̊/	*khy* /k.j̊/	*thy* /t.j̊/		*phy* /p.j̊/			
/j/	*ky* /k.j/	*ty* /t.j/	*ny* /n.j/	*py* /p.j/		*ry* /r.j/	*sy* /s.j/
/r̰/		*thr* /t.r̰/	*nhr* /n.r̰/				
/r/	/k.r/	/t.r/	/n.r/	/p.r/	/m.r/		/s.r/
(/l̰/)				(/p.l̰/)			
/ʀ̰/	*kṭr* /k.ʀ̰/			*pṭr* /p.ʀ̰/	*mṭr, mtr* /m.ʀ̰/		
/R/	*kdr* /k.R/	*tdr* /t.R/	*ndr* /n.R/	*pdr* /p.R/	*mdr* /m.R/		*sdr* /s.R/
/l̰/	*khl* /k.l̰/	*thl* /t.l̰/	*nhl* /n.l̰/	*phl* /p.l̰/			
/l/	/k.l/	/t.l/	/n.l/	/p.l/	/m.l/	/r.l/	/s.l/

Table 8. *(continued)*

/L̥/	*kṭl* /k.L̥/					*rṭl* /r.L̥/	*sṭl* /s.L̥/
/L/	*kdl* /k.L/	*tdl* /t.L/	*ndl* /n.L/	*pdl* /p.L/			*sdl* /s.L/
/ʍ/				*phv, phv-ṃ* /p.ʍ/			
/w/	*kv* /k.w/	*tv* /t.w/		*pv* /p.w/	*mv* /m.w/	*rv* /r.w/	*sv* /s.w/
/ç/	*khy-ṃ* /k.ç/						
/s/	*ks, kṣ* /k.s/	/t.s/	/n.s/	/p.s/	/m.s/		
/h/			/n.h/		/m.h/	/r.h/	

7.2.1 *The problem of /s.C^{h}/ clusters*

I am not sure whether /s.t^{h}/ belongs in Table 8. It is not clear whether what might be read as *sthuy·* 'ʔ' (32.1) should be read as *sp̣uy·* or *svuy·*. In any case, *sth* is at least theoretically possible given that *sph* is attested once in *sphir·* 'ʔ' (20.4).

The rarity of *sph* and perhaps *sth* may imply that they reflect aspirated allophones of stops after /s./ and should be phonemicized as /s.p/ and /s.t/ rather than as /s.p^{h}/ and /s.t^{h}/. But that could only be confirmed if we knew what *sthuy·* and *sphir·* meant and if variant spellings †*stuy·* and †*spir·* existed.

7.2.2 *Initials not attested after preinitials*

The only initials other than /°/ that are not attested with preinitials are of low frequency: /k^{h}/ (§ 7.3.3), /ɲ/, and /D/. The absence of /C.k^{h}/ and /C.ɲ/ may be accidental since preinitials are found before other aspirates and nasals. If /D/ was the voiceless counterpart of /ɖ/ (§ 7.3.7) and if /ɖ/ is a hardened liquid (§ 7.3.6), then the absence of /C.D/ may also be accidental since preinitials are found before voiceless liquids.

7.3 Commentary on specific core initials

7.3.1 *The problem of phonemicizing*

Indic scripts lack a means to distinguish between zero initials and glottal stops. Indic initial vowel symbols can stand for bare vowels or glottal stop-vowel sequences depending on the language.[19] Hence one could phonemicize Pyu *°V akṣaras* as

19. I am disregarding liquid-initial readings of the syllabic liquid symbols in languages lacking syllabic liquids: e.g. Hindi [ɾɪ] for Devanagari ऋ <ɾ̥>, etc.

either /V/ or /ʔV/. I am unable to formulate a strong argument in favor of either analysis.

I could argue that /V/ is correct because if Pyu had /ʔV/ syllables, I would expect *C°V* /CʔV/ syllables with preinitials before glottal stops as in Old Mon. No such syllables exist in the corpus. However, we do not know how prefixation worked in Pyu or even if it existed as a productive process.[20] Perhaps there are no *C°V* because glottal stops were lost between consonant preinitials and vowels.

I could also argue that /ʔV/ is correct because it would simplify the syllabic structure: there would be no zero-initial syllables, so I could drop the parentheses around the initial (C) in my formula for Pyu syllables. However, that is merely the imposition of a modern aesthetic, notational preference that tells us nothing about whether native Pyu speakers produced or perceived initial glottal stops centuries ago.

/ʔV/ is more defensible on typological grounds since it is typical of languages of the region (e.g. Mon), but typology only points to probable solutions which are not necessarily correct. In Kammu, /ʔV/ syllables have the low tones characteristic of syllables with voiced initial consonants (Svantesson 1983: 51). Those tones may have developed during a stage when those syllables began with voiced vowels rather than with voiceless glottal stops.

If Pyu had a modern descendant or even a close surviving relative which had true initial vowels or glottal stops, I could simply project those initials back into Pyu, but no such successors exist.

To avoid choosing one or the other solution, I carry over the non-IPA symbol /°/ from my transliteration into my phonemic transcription. /°/ resembles a zero and may be interpreted as a zero initial or as a glottal stop.

/°/ occurs almost entirely before /o/. There are only three definite *akṣara*s with /°/ before other vowels, and only one appears more than once:

- *°at·ṁḥ* '?' (25.7)
- *°ik·* 'one' (16.1A)
- *°ip·* '?' (32.2, 32.4 [2×], 32.5 [2×], 32.6 [2×])

An *akṣara* which might be read *°ar·ṁḥ* '?' (32.5) could be added to that short list. Another potential addition is an *akṣara* which might be read *°ū* (170). Although there are tables of the Pyu script containing independent vowel symbols for *°ā °ī*

20. Old Chinese, Old Tibetan, and Pre-Tangut are rich in prefixation, so it is likely that their common ancestor Proto-Sino-Tibetan was also rich in prefixation. But to what degree Pyu inherited that wealth is an open question. The prefixes that have been identified in Pyu may have been fossilized: e.g. the *p-* in the motion verb *pgau[t]·* /p.got/ (16.4b, 16.6A) which may be cognate to *got·* /got/ (16.4A), another motion verb in the same text.

°u °e °ai °au (e.g. Tha Myat 1963: (1), such characters are not in Griffiths et al.'s (2017a) corpus. Those characters are taken from Pyu scripts for Indic languages. Their inclusion reflects a long tradition of regarding the Pyu scripts for Pyu and Indic languages as a single script. The Pyu characters for *°ā °ī °u °e °ai °au* in Pyu are unknown; they may not exist.

The rarity of /°/ in Pyu may imply that nearly all original zero or glottal stop-initial words developed a secondary initial consonant with the major exceptions of the high-frequency morpheme /°o/ marking possession of a following noun and its homophone /°o/ marking nominalization of a following verb.

7.3.2 */h/*

In Table 1, the *akṣara ha* was in the voiced aspirate column since it originally represented Indic voiced /ɦ/, but in Table 7, Pyu /h/ is in the voiceless aspirate column. I reconstruct /h/ as voiceless because /h/ is more typical of languages of the region (e.g. Mon and Burmese),[21] and the Pyu used *h* in digraphs for voiceless sonorants (§ 7.3.4). It would make less sense for the Pyu to use *h* to write voiceless sonorants if it represented voiced /ɦ/, though *h* would be the best available choice out of the standard Indic script character set.[22]

7.3.3 *Voiceless aspirated obstruents*

Although voiceless aspirated obstruents are common in the Pyu script, voiceless aspirated obstruent phonemes are far less common. Most voiceless aspirated obstruent symbols appear in character clusters for sequences of voiceless unaspirated stops followed by voiceless sonorants (Table 8 and § 7.3.4). The frequencies of the remaining written aspirated obstruents (excluding uncertain readings other than *Cṭha* [§ 7.4.6] and *s[th]uy·*) are in Table 9.

Table 9. Pyu voiceless aspirated obstruent frequency

Aspirate	***kh*** **/kʰ/**	***ch*** **/cʰ/**	***ṭh*** **/ʔ/**	***th*** **/tʰ/**	***ph*** **/pʰ/**
Unique *akṣara*s	7	9	1	5–6?	9
Without preinitial	18	25	0	23	16
With preinitial	0	2	1	1?	1
Total	**18**	**27**	**1**	**23–24?**	**17**

21. Moulmein Sgaw Karen, however, has voiced [ɦ] but no voiceless [h] (Jones 1961: 7). Jones phonemicized [ɦ] as /h/, but one could just as easily use the symbol /ɦ/ instead.

22. *Visarga* does represent voiceless /h/ but can only appear in *akṣara*-final position. If the Pyu pronounced *h* as /ɦ/ as in Indic languages and if they chose not to limit themselves to the existing Indic character set as much as possible, they could have created new characters or diacritics for voiceless sonorants.

The figures for *kh* are probably slightly lower than they should be because of the difficulty of distinguishing *kha* from *ce*. But even the correct figure would not alter the fact that voiceless unaspirated obstruents are far rarer in Pyu[23] than they are in other early Sino-Tibetan languages. I shall explore the diachronic implications of that fact in future studies of Pyu phonology.

7.3.4 *Voiceless sonorants*

Sequences of *h* atop sonorant symbols (e.g. *hn*) are very common in the corpus. These could be interpreted as /h.C/ preinitial-initial sequences or as phonemic voiceless sonorants /m̥ ŋ̊ n̥ m̥ j̊ r̥ l̥ ʍ/.

I favor the latter interpretation for three reasons.

First, there are no instances of *h* before nonsonorant symbols. If Pyu had /h.n/, why would it not also have /h.t/?[24] Modern spoken Mon has /h/ in presyllables in both types of environments: e.g. ဗနက် <banak·> /hənèak/ 'immersion' and ဗ္ဒန် <bdan·> /hətòn/ 'to hide'.

Second, *h* (128×) is more common than any uncontroversial preinitial with the exception of *k*- (171×) before sonorants (Table 10).

Table 10. Frequency of *h*- and preinitials before sonorant symbols

	h-	*k-*	*t-*	*n-*	*p-*	*m-*	*r-*	*s-*
-ṅ-	8	11	14	2	10	0	0	0
-ñ-	2	0	0	0	0	0	0	0
-n-	49	42	22	0	4	0	3	17
-m-	9	6	15	7	0	0	17	4
-y-	9	7	20	3	2	0	10	2
-r-	29	37	57*	8	58	43	0	29
-l-	16	45	21	2	45	14	15	4
-v-	6	23	20	0	5	4	6	4
Total	**128**	**171**	**91**	**22**	**124**	**61**	**51**	**60**

* I have excluded one instance of *tr* from the total since I consider it to be an error for *ṭr* /ʀ̥/ (§ 7.3.10).

23. Shafer (1943: 325) was the first to observe the low frequency of voiceless aspirated obstruents in Pyu. My subsequent interpretation of written *Ch*-sonorant clusters as voiceless unaspirated stop-voiceless sonorant clusters has further reduced that already low frequency.

24. It is possible that */h.t/-type sequences became /tʰ/, but even if that were the case, a voiceless sonorant interpretation of *hn*-type sequences is also still possible.

Third, phonemic voiceless sonorants are genealogically and typologically plausible: they are in Old Chinese (Baxter & Sagart 2014a) and Old Tibetan[25] and in Pyu's neighbor Old Burmese.

Sequences of voiceless aspirated obstruent consonant symbols atop sonorant symbols (e.g. *thn*) are also very common in the corpus. These coexist alongside sequences of voiceless unaspirated obstruent consonant symbols followed by sonorant symbols (e.g. *tn*). Although pairs of *akṣara*s differing only in aspiration exist: (e.g. *thna* '?' 25.7 : *tna* '?' 27.6 and *phn̓an·* '?' 32.4 : *pn̓an·* '?' 20.3), nothing is known about their semantics. We do not know if *thna* and *tna* were different morphemes (or even if they were merely parts of polysyllabic morphemes). Nor do we know if *phn̓an·* and *pn̓an·* were variant spellings of the same morpheme or the same part of a morpheme.

Ignorant of semantics, I see three possible analyses:

A. Aspiration is phonemic in clusters: e.g. *thna* /tʰ.na/ : *tna* /t.na/.
B. Aspiration is not phonemic in clusters: e.g. *thna* ~ *tna* /t.na/.
C. The aspiration in the first written consonant reflects an aspirated allophone of a preinitial obstruent before a phonemic voiceless sonorant: e.g. *thna* /t.n̥a/ : *tna* /t.na/.

Analysis A is improbable because phonemic aspirated preinitials seem to be rare in languages of the region. Thomas' (1992: 209) survey of Southeast Asian sesquisyllabism contains only a single example of a minor syllable with an aspirated initial (Northeastern Thai /'pʰi"ti?/ 'pretend'). Unlike Pyu preinitials, that minor syllable has a vowel /i/. I suspect that Thomas's minor syllables with "nearly full vowel contrasts" (Thomas's type (iv)) may have a wider range of consonant phonemes than presyllables with more restricted vocalism (Thomas's types (i–iii)).

Analysis B looks initially promising because it avoids positing obstruent preinitial-voiceless sonorant sequences such as /t.n̥a/, and it has a typological parallel in Khmer, which generally has aspirated allophones of unaspirated phonemes before sonorants other than /r/:[26] e.g. ខ្មែរ <khmɛra> /kmae/ [kʰmae] 'Khmer'.

However, Pyu allowed preinitials to combine freely with initials, and if Pyu had voiceless sonorant initials, there should be preinitial-voiceless sonorant initial combinations like /t.n̥/. We know nothing about how Pyu prefixation worked. Perhaps voicing in sonorant-initial roots was neutralized before stop prefixes: e.g. /t./ + /n̥/ and /t/ + /n/ would both be *thn* ~ *tn* /t.n/. Such a rule would require examples involving roots with glosses, and at present I know of no examples.

---

25. Old Tibetan has a single voiceless sonorant transliterated as *lh*.

26. The sole exception is /k/ in the homorganic cluster /kŋ/ which is pronounced as [kəŋ] instead of †[kʰŋ] (Huffman 1970: 8).

Analysis C entails positing many obstruent preinitial-voiceless sonorant sequences like /t.n̥a/, but on the other hand it requires no speculative morphophonemic rules and is therefore preferable to analysis B which may require such a rule. At this early stage of the study of Pyu phonology, I prefer overdifferentiation unless it is typologically implausible as in the case of *mṭr* and *mtr* for /m.R/ (§ 7.3.10).

7.3.5 *Palatal sonorants*

Although I use palatal stop symbols /c cʰ ɟ/ for the phonemes corresponding to *c, ch*, and *j* in the script for parallelism with /ɲ/, these phonemes could have been pronounced as palatal affricates [tɕ tɕʰ dʑ], retroflex affricates [tʂ tʂʰ dʐ], or palato-alveolar affricates [tʃ tʃʰ dʒ]. If *ts* represents /t.s/ (§ 7.4.7) pronounced as a presyllable-syllable onset sequence [təs], then *c, ch,* and *j* could have been alveolar affricates [ts tsʰ dz].

7.3.6 */ḍ/*

ḍ /ɖ/ is exclusively in the three Late Pyu texts (7, 8, and 16). In several instances it corresponds to *l* (and in one case to *t*) in both foreign and Pyu words:

- Skt *triloka-* 'three worlds' : Pyu *triḍoga-* (7.4, 8.3)
- Pali *muggali* (a name) : Pyu *mugaṃḍu-* (7.15), *mugaṃtu-* (8.15)
- Old Burmese *sak·munalon·* (7.30) and Mon *sak·munalor·* (a name; 7.25) : Pyu *[sa]manarḍo[ḥ]* (7.20), *samanarḍoṃṁḥ* (8.20)
- Pyu *pli* 'grandson' (7.24) : Pyu *pḍi* 'id.' (8.25)
- Pyu *la* 'be it' (in the sense of 'either … or'; 7.24) : Pyu *ḍa* 'id.' (8.25)

Pyu *pli* and *la* have Old Burmese cognates with *l: mliy·* 'grandson' and *lañ·* 'be it' (7.36).

Clearly Old Pyu /l/ hardened to a /ɖ/-like consonant written as *ḍ* and perhaps in one instance as *t* (*mugaṃtu-*; 8.15) in Late Pyu.[27] Middle Pyu *l̤* /l̠/ represents an intermediate step in that shift (§ 7.3.8).

Yet Late Pyu texts also still have *l*: e.g. 'four' appears in (39.3) as *plaṁ* rather than as †*pḍam*. It is not clear whether *ḍ* and *l* represented allophones in free variation or if *l* in Late Pyu was a purely historical spelling. In other words, could, for instance, 'grandson' have been pronounced as both [pɖi] and [pli] by Late Pyu speakers? Or was [pɖi] the only pronunciation of /pɖi/ which could either be written as currently pronounced (i.e. as *pḍi*) or according to its archaic or defunct pronunciation [pli] (i.e. as *pli*)?

---

27. Late Pyu /t/ may have been voiced in intervocalic position. If so, *t* would have normally represented [d] in that position and would have been a close though imperfect match for [ɖ].

It is also not clear if Late Pyu had neologisms and/or recent borrowings that reintroduced the phoneme /l/ after earlier /l/ had shifted to /ɖ/. I would expect such words to be consistently spelled with *l* unlike earlier words which were spelled with both *l* and *ḍ*. There are no *akṣara*s with *l* unique to the available Late Pyu texts, but of course those three texts contain only a tiny fraction of the Late Pyu lexicon.

For now I regard /ɖ/ as the Late Pyu consonant corresponding to Old Pyu /l/ and Middle Pyu /l̥/. I could write all three phonemes with the same symbol /l/, but I prefer to use symbols approximating their pronunciation in each period. Moreover, I cannot rule out the presence of a new /l/ in Late or even Middle Pyu.

dl /L/ (§ 7.3.10) is absent from Late Pyu texts. If that absence is not due to chance and the small size of the Late Pyu corpus (only three legible inscriptions), perhaps Old Pyu /L/ became another source of Late Pyu /ɖ/, though I cannot demonstrate that without sets of *dl*- and *ḍ*-spellings for the same morphemes.

/ɖ/ may have phonetically been something other than a retroflex stop: e.g. an implosive [ɗ] which is the Mon value of the symbol *ḍ*. An agnostic symbol like /D/ might be preferable to /ɖ/, but I have already reserved /D/ for another phoneme (§ 7.3.7).

7.3.7 */D/*

The Late Pyu initial *hḍ* is unique to the Kubyaukgyi inscription where it occurs in only three morphemes:

- *hḍiṁḥ* '?' (7.7, 8.8)
- *hḍiṁḥ* 'to dedicate to' (7.25, 8.26)
- *hḍ[ī]* (7.20), *hḍī* (8.20) 'dedication formula'

The last two may belong to the same word family, though there are no other known cases of an *iṁḥ* ~ *ī* (or *i*) alternation.

hḍ cannot be a voiced aspirated retroflex stop /ɖʱ/ which would have been written as *ḍh*. Nor can it be a preinitial-initial sequence /h.ɖ/ since I have rejected /h./ as a preinitial (§ 7.3.4). I shall write this phoneme as /D/ to indicate that it was similar to a voiced retroflex stop. (There is no capital D with a retroflex tail in Unicode, so I use a simple capital D.)

Another possible phonemic symbol is /ɖ̊/ to indicate that *hḍ* /D/ may be to *ḍ* /ɖ/ what Old Pyu *hl* /l̥/ was to *l* /l/. There is no *hl* in the three Late Pyu texts known so far (7, 8, and 39) which is what I would expect if *hḍ* is a fortition of /l̥/. It is not, however, possible to dismiss the possibility that Late Pyu had reacquired /l̥/ in neologisms and loans, and/or retained /l̥/ after preinitials in words that happen not to be in those three texts.

Middle Pyu *ṭh* may represent a transitional stage between /l̥/ and /D/ (§ 7.4.3).

If Old Pyu /L/ was a source of Late Pyu /ɖ/ (§ 7.3.6), then Old Pyu /L̥/ may have been another source of Late Pyu /D/. *ṭl* /L̥/ is absent from Late Pyu texts, though that is not sufficient evidence of the absence of /L̥/ from Late Pyu due to the paucity of material. The three Late Pyu words with /D/ do not have potential alternate spellings with *ṭl* in Old Pyu.

7.3.8 /ɭ/

The initial *ḷ* is unique to inscription (37) in which it appears in three *akṣara*s: *ḷo* '?' (2×), *pḷaṁḥ* '?', and *ḷiṁ* '?'. *pḷaṁḥ* may be the same word as *plaṁḥ* 'base' (16.1b, 16.2b, 16.2C). *ḷo* and *ḷiṁ* may correspond to *lo* '?' (25.3) and *liṁ* '?' (32.3), though no matching collocations have been found. These parallels suggest that *ḷ* may be from an earlier /l/. I shall take the spelling *ḷ* at face value and phonemicize it as retroflex /ɭ/.

The retroflexion of /l/ might have been a dialectal phenomenon, as (37) was found in Ouk Shit Pin in Lewe Township in Naypyidaw Union Territory, approximately 120 km northeast of Sriksetra, the source of most Pyu texts. There is, however, no guarantee that the location of an inscription necessarily correlates with the local language: e.g. the Tondaw inscription (23) was found in Rakhine which was probably never Pyu-speaking.

A retroflex /ɭ/ may also have been an intermediate stage between /l/ and the retroflex stop /ɖ/ found in the Late Pyu texts of Pagan (7 and 8) and Myittha (39) which are both about 200 km away from Ouk Shit Pin. So I tentatively consider (37) to be in Middle Pyu. But if /ɭ/ never hardened to /ɖ/ in the dialect of (37), (37) could actually be contemporary with the Late Pyu texts. Conversely, if the retroflexion in the dialect of (37) might be an old phenomenon, then (37) could actually be contemporary with the Old Pyu texts. In either case, the term "Middle Pyu" would be inappropriate. A more neutral term may be "Retroflex L Pyu" which avoids chronology.

Perhaps the retroflexion of /l/ was an innovation of northern Pyu distinguishing it from the southern Pyu of Sriksetra. If so, then the fortition of /ɭ/ to /ɖ/ was a later innovation which further distinguished the northernmost dialects of northern Pyu from the others.

There is also a single instance of regular *l* in (37): *le* '?' an *akṣara* also found in Old Pyu texts (e.g. 27.5), though not in collocations matching *kra le nga* '? ? ?' in (37). It is not clear whether the *le* of (37) was a historical spelling, a phonetic spelling reflecting an allophone of /ɭ/, or a phonetic spelling of a neologism or loanword with a new /l/ postdating the retroflexion of original /l/. I tentatively phonemicize the *le* of (37) as /le/ until further discoveries clarify the history of laterals in Pyu.

7.3.9 */l̥ l/*

The status of these phonemes in Middle and Late Pyu is uncertain.

It seems that at least some /l/ backed to *l̤* /ɭ/ in Middle Pyu (§ 7.3.8) and then hardened to *ḍ* /ɖ/ (§ 7.3.6), but it is unclear whether the *l* coexisting with *l̤* and *ḍ* in those texts represents an allophone.

Perhaps at least some /l̥/ backed and hardened to *hḍ* /D/ in Late Pyu (§ 7.3.6) and possibly to *ṭh* after a now-illegible preinitial in Middle Pyu (§ 7.4.6).

hl /l̥/ is absent from Middle and Late Pyu texts, but that could be an artifact of a limited corpus of only four inscriptions.

7.3.10 */R̥ R L̥ L/*

Clusters of three written consonants such as *kṭr* and *pdl* are frequent in Pyu texts. How can they fit into the (sesqui)syllable structure /(C).(C)V(C)(m̤)(h̤)/ that I proposed in § 5? A clue to the solution lies in the word "written"; consonant symbols do not necessarily correspond one to one with consonant phonemes.

I originally thought Pyu (sesqui)syllables had a /(C).(C)(C)V(C)(m̤)(h̤)/ structure. Then I noticed that all the three-consonant sequences that did not include digraphs for voiceless sonorants had preinitials followed by only five possible two-consonant combinations: *ṭr, tr, dr, ṭl*, and *dl*. There were no combinations such as *knp-* or *syv-* without coronal stops followed by liquids.

A very literal interpretation of the spellings would necessitate three-consonant sequences in the syllable structure with only three possibilities in the second slot (*ṭ, t, d*) and two in the third (*r* and *l*). But such limited possibilities are suggestive of digraphs for single phonemes. In the *quôć ngữ* orthography for modern Vietnamese, the only possible consonant clusters in native words are *ch, gh, kh, ng, ngh, nh, ph, th*, and *tr*. If one knew nothing about Vietnamese, one might propose that Vietnamese had six /Ch/ clusters, a single /ng/ cluster, a single /ngh/ cluster, and a single /tr/ cluster. However, all nine sequences represent eight single phonemes /c ɣ x ŋ ɲ f tʰ tʂ/.[28] Similarly, the five Pyu sequences *ṭr, tr, dr, ṭl*, and *dl* may also represent four single phonemes /R̥ R L̥ L/.

I use capital letters to avoid committing to specific phonetic values. The four phonemes may have been pronounced as clusters [ʈɽ ~ tr dr tl dl] or as affricates [tʂ dʐ tɬ dɮ] lacking characters in Indic scripts.

/R̥/ is always spelled *ṭr* except in *mtraḥ* '?' (74.4) which is probably an error for †*mṭraḥ* /mR̥ah/. A phonemic distinction between *mṭr* in *mṭreḥ* '?' (25.1) and *mtr* in *mtraḥ* is implausible, as I know of no language in the region that distinguishes /ʈr/

---

28. *ng* and *ngh* are in complementary distribution in Vietnamese orthography. *ngh* represents /ŋ/ before the front vowel letters *i, ê*, and *e*, and *ng* represents /ŋ/ elsewhere.

and /tr/. All other instances of *tr* are /t.r/, just as *kr* and *pr* are /k.r/ and /p.r/. The retroflex *ṭ* of *ṭr* indicates that the stop element of /Ṛ/ had assimilated to a following retroflex element (e.g. *ṭra* /Ṛa/ '?' 25.4), whereas *tr* represents /t.r/ with a dental preinitial /t./ that did not assimilate to /r/ (e.g. *tra* /t.ra/ 'goods, ornaments' 8.6).

Similarly, *ṭl* represents /Ḷ/ with retroflex *ṭ* indicating that the *t*-element of the phoneme had assimilated to a following retroflex *l*-element (e.g. *ṭlan·* /Ḷan/ '?'; 20.3), whereas *tl* represents /t.l/ with a dental preinitial /t./ that did not assimilate to /l/ (e.g. *tla* /t.la/ '?'; 27.6).

There are no parallel spellings *ḍr* or *ḍl* with voiced retroflex *ḍ* since Pyu has no preinitial /d./ (§ 6.2.3) and hence has no ambiguity between /R/ and †/d.r/ or /L/ and †/d.l/; *dr* and *dl* with *d* always represent the unit phonemes /R L/. If Pyu had no [dr], the lack of †/d.r/ in Pyu may explain why Sanskrit *candra-* 'moon' was borrowed into Pyu as *jan·tra-* (16.2d) with *tr* /t.r/.

One objection to /Ṛ R Ḷ L/ is that they raise the number of liquids in Pyu to eight (or nine including /ɭ/). Such a wealth of liquids may be without parallel in Southeast Asia. However, if /Ṛ R Ḷ L/ were not actually liquids but clusters or affricates with retroflex and lateral features such as /tʂ dʐ tɬ dɮ/, then they would be comparable to the four retroflex affricates (*ts nts tsh ntsh* /tʂ ⁿtʂ tʂʰ ⁿtʂʰ/) and four lateral cluster unit phonemes (*d ndl dh ndlh* /tl ⁿtl tḷ ⁿtḷ/) of Green Mong, albeit without prenasalization or aspiration. A rich inventory of *r*- and *l*-like sounds may have been partly inherited, as the eight Pyu phonemes written with *r* and *l* may correspond to the four to seven "flowing wind sounds" (liquids) of the Tangut native phonetic tradition or the eight liquids that Baxter & Sagart (2014a: 69) reconstructed for Old Chinese.[29]

ṭl /Ḷ/ is absent from Middle and Late Pyu. If that lacuna is not due to chance, Old Pyu /Ḷ/ could have become a stop written as *ṭh* in Middle Pyu (§ 7.4.6) and as *hḍ* /D/ in Late Pyu (§ 7.3.7).

dl /L/ is absent from Middle and Late Pyu. If that lacuna is not due to chance, Old Pyu /L/ could have become Middle Pyu /ɭ/ (§ 7.3.8) which hardened to Late Pyu /ɖ/ (§ 7.3.6).

---

29. "Flowing wind sounds" are also known as 來日音 *lái rì yīn* '*l*- and *r*-sounds' in Chinese Tangutological literature. There is no consensus on the number of "flowing wind sounds" in reconstructions. Fanwen Li's (1986: 127) table compares inventories of "flowing wind sounds" in five reconstructions.

Starostin (1989) reconstructed a set of four lateral affricates (*tl- *tlh- *dl- *dlh-) for Old Chinese, but his proposal has not been widely adopted. See Sagart (1999: 36–40) for a critique.

7.3.11 */w/*

There is no way to distinguish between /w β ʊ v/ in an Indic script without diacritics or special characters, so all that can be said for certain is that (1) *v* represented a Pyu consonant similar to the phonetic value of *v* as they heard it from speakers of Indic languages and (2) *v* represented a consonant distinct from the consonant written as *v-ṃ*. The values of *v* and *v-ṃ* are dependent on one another; if I assign one value to *v*, I cannot also assign it to *v-ṃ*, and vice versa. Similarly, the phonetic values of voiceless *hv* (§ 7.3.4) and *hv-ṃ* are dependent on the values of their voiced counterparts *v* and *v-ṃ*.

v occurs in both initial and final subscript positions. [w] is a more likely coda than [β ʊ v] in a Southeast Asian language. Hence I write the phoneme represented by *v* as /w/. Although the possibility that /w/ had positional allophones cannot be ruled out, I retain *v* in transliteration to preserve the spellings of Indic loans.

7.3.12 */ɓ/ and other proposed glottalized consonants*

I have already mentioned the graphic resemblance between Pyu *ḅ* and Old Mon *ḅ*, both modifications of Indic *b*, in § 4.2.1. Does this resemblance extend to the phonemic level? In Old Mon, *ḅ* represents the implosive /ɓ/. It would not be surprising if Old Mon's neighbor Pyu also employed *ḅ* for /ɓ/, but neighboring languages may have different values for the same modified letter: e.g. *ä* is ideally /ɛː/ in German but /æ/ in conservative Slovak.[30] Therefore additional evidence is needed to identify the phonemic value of *ḅ* in Pyu.

Long after Blagden (1911: 368) had observed the graphic similarity of Pyu *ḅ* to Old Mon *ḅ* /ɓ/, Luce (1985: 63) was perhaps the first to phonetically identify *ḅ* in IPA as [ɓ], and twenty years later, Katō (2005) also identified *ḅ* as *ɓ[31] without citing Luce. Neither Luce nor Katō explicitly stated their reasoning, though both did provide clues. Luce mentions areal parallels in "Old Mon, Northern Dai, Southern Chin, and Southern Karen" and a Middle Chinese transcription of what he assumed to be one of the Pyu *ḅ*-words for 'Buddha' (e.g. *ḅut·dha* 20.1) as 沒馱 *muət-dʻâ in Karlgren's (1957) reconstruction. Karlgren's *m is likely to have been prenasalized [ᵐb] in the prestigious northwestern court dialect of the period (Coblin 1994: 58). Katō's Pyu-West Pwo comparisons (Table 11) suggest that he

---

30. Many German speakers pronounce *ä* as [eː] (Hall 2003: 84). Nearly all modern Slovak speakers have merged /æ/ with /e/ [ɛ] (Short 1993: 534). That has nothing to do with German influence; /æ/ only occurs after labials, whereas /e/ has no distributional limitations and is hence far more common.

31. I have starred all of Katō's phonemic reconstructions to distinguish them from his transliterations which I have italicized and from my phonemic interpretation of Pyu in slashes.

was motivated by Pyu *ḅ* corresponding to West Pwo /ɓ/ in his proposed cognate sets. Katō's cognate sets are dependent on his proposed meanings for Pyu words in the Kubyaukgyi inscription and are not accompanied by any discussion of regular correspondences.

Even if Katō's cognate sets are valid, ideally /ɓ/ should be supported by Pyu-internal evidence. A key clue in the Pyu script is the fact that *ḅ* is almost never accompanied by the subscript dot that is frequently found with voiced consonant characters.[32] In Southeast Asian languages, implosive /ɓ/ often patterns with voiceless consonants even though it is voiced: e.g. in Mon, *ɓ words developed clear register like *p-words rather than breathy register like *b-words. Similarly, Pyu *ḅ* – with only one exception – avoids subscript dots like *p* instead of welcoming them like *g* and *d*.

Pyu *b* is never accompanied by a subscript dot unlike *g* and *d*. When I thought that the subscript dot indicated breathy phonation after nonimplosive voiced consonants, I briefly considered the possibility that *b* could have been implosive /ɓ/[33] whereas *ḅ* was something else: e.g. a fricative like /β/ (Miyake 2017a). But there is no evidence such as *ḅ* ~ *v* alternations indicating a fricative pronunciation of *ḅ*. The aforementioned Chinese transcription of a Pyu word for 'Buddha' is ambiguous: it has a prenasalized stop *ᵐb that might have been the only available approximation for either Pyu *ḅ* or *b* [34] since Late Middle Chinese lacked an oral stop *b or voiced labial fricatives like *β or *v (Pulleyblank 1991: 10).

Another possibility is that *ḅ* and *b* represent allophones [ɓ] and [b] of a single phoneme that did not pattern like a voiced stop and therefore may have been implosive /ɓ/. Modern Thai and Lao /b/ from *ɓ is no longer implosive, but it is associated with series 1 tones that distinguish it from historically voiced stops that are associated with series 2 tones. Pyu spelling may reflect a period when *ɓ was losing its implosion. Such a loss would not have been conditioned, as both *ḅ* and *b* appear before the same vowels with one exception.[35]

---

32. The sole instance of *ḅ* with a subscript dot (*ḅaṃ* '?'; 151) may be an error for *ḅaṁ* '?' (98) with a superscript dot (i.e. *anusvāra*) or a loanword not conforming to native Pyu phonotactics.

33. The use of unmodified consonant characters for implosives has parallels in Mon which has retroflex <ḍ> for implosive /ɗ/ and Khmer which has retroflex <ṭ> and <p> for implosive [ɗ] and [ɓ].

34. Although Luce assumed that the Chinese transcription was probably of a Pyu word for 'Buddha' beginning with *ḅ*, he was also aware of spellings with *b* (e.g. *budha*; 8.11), and there is no way to know for sure which version of the Pyu word the Chinese transcriber had in mind.

35. *b* is never attested with *e*, whereas there are two *akṣara*s with *ḅe: pḅe* '?' (20.4) and *ḅeṅ·* '?' (22.7). The absence of †*be* is probably accidental: either no such syllable happened to exist in Pyu, or it did exist but happened not to be in the limited corpus. In either case, there is no phonetic reason for *ḅ* to be favored over *b* before *e*.

At first that hypothesis is appealing because of *ḅ* ~ *b* alternations in spellings of the same morpheme. The preverbal realis marker *ḅin·ṁḥ* (16.2C) ~ *ḅiṁḥ* (7.1) ~ *biṁḥ* (7.10)[36] and the honorific *ḅay·ṁḥ* (16.2A) ~ *ḅaṁḥ* (7.12) ~ *baṁḥ* (7.13) are spelled with both characters, as is the loanword 'Buddha': *ḅuddha* (24 II) ~ *ḅut·dha* (20.2) ~ *ḅudha* (8.10) ~ *budha* (8.11). However, there are also reasons to curb one's enthusiasm for a single voiced labial stop phoneme, at least at an early stage of Pyu.

First, the characters for *ḅ* and *b* are so similar that it would be easy to accidentally write one instead of the other. Such an error might have been especially common among Late Pyu speakers influenced by Old Burmese which had neither /ɓ/ nor /b/ in native words. Sometimes it is difficult to tell the two characters apart. Blagden (1919) read *ḅ* in the Kubyaukgyi inscription where Griffiths et al. (2017a) read *b* in the honorific and 'Buddha'.

Second, the *b*-spellings of the aforementioned three morphemes have a limited distribution possibly implying a merger of /ɓ/ and /b/ in later texts and sporadic errors or intentionally Indicized spellings of 'Buddha' with *b* in earlier texts.

The only secure cases of the preverbal realis marker *biṁḥ* with *b* instead of *ḅ* are in Late Pyu texts (7 and 8). There is a *biṁḥ* in (21), a stone slab from သဲကုန်း <saikun·ḥ> Thegon township near Sriksetra, but that text has subscript final consonants, so that *biṁḥ* most likely represented /bïh/ rather than the preverbal realis marker *ḅin·ṁḥ* /ɓïn̰/. However, it is not impossible that *biṁḥ* is an error for *bin·ṁḥ* since *bin·ṁḥ tor·* '?' (165.1) might be another spelling of *biṁ tor·* '?' (165.3). But even if *bin·ṁḥ* had been intended in (21), that still does not insure that the actual spelling *biṁḥ* in (21) was a preverbal marker since the text following *biṁḥ* has been lost, and it is not certain that *bin·ṁḥ* (165.1) was the preverbal marker as I explained in footnote 36.

baṁḥ is only in the Late Pyu texts (7), (8), and (39), the Middle Pyu text (37), and the Old Pyu texts (24) and (55b), a silver reliquary and piece of silver foil from the ခင်္ဘ <khaṅ·bha> Khin Ba mound of Sriksetra, and (36), a Buddha sculpture from the Mandalay region.

The function of *baṁḥ* is uncertain in all texts other than (7), (8), (24), and (39) where it is an honorific accompanying revered nouns such as *hra* 'sacred image'.

baṁḥ is the only spelling of the honorific in (24), where it appears eight times before Indic names. It may be an Indicized spelling of *ḅaṁḥ* with *b* instead of

36. This preverbal marker was glossed by Julian K. Wheatley (p.c.). I exclude *bin·ṁḥ* (165.1) because it is not clear from context whether it is another spelling of the preverbal marker. It is not known whether the *tor·* '?' following *bin·ṁḥ* is a verb, the first syllable of a verb, or another part of speech.

non-Indic *ḅ*. That Indicization hypothesis, however, fails to account for why *baṁḥ* thrice precedes *ḅuddha* with non-Indic *ḅ* (24.II, 24.III, 24.IV).

(36) (which may not be in Pyu; Julian K. Wheatley, p.c.) and (37) have no final subscript consonants, so their *baṁḥ* may represent morphemes with a final consonant other than /j̊/: e.g. *bam·ṁḥ* /bäm̥/ '?' (20.5).

(179) does have final subscript consonants, so its two *baṁḥ* may be a /bäh/ '?' unrelated to the honorific /ɓäj̊/. The four *ḅaṁḥ* in (179) may be the same morpheme. Both *baṁḥ* and *ḅaṁḥ* are in the same context in (179): i.e. following the numeral 1 and preceding one or more *akṣara*s of unknown meaning. If *baṁḥ* and *ḅaṁḥ* were the honorific, the scribe may have deliberately abbreviated the word (cf. English *St.* for *Saint*) while retaining subscript final consonants in other words.

Although 'Buddha' appears in the corpus with both *ḅ* and *b*, only two texts (7 and 8) contain both *ḅ* and *b*-spellings, and each only has a single *b*-spelling. The *b*-spelling in (7.19) is questionable, as the *b* is so poorly formed that Blagden (1919:63) read it as *g*. (8), whose text is nearly identical to (7), has *ḅudha* in line 19. Moreover, the only other texts containing *b*-spellings are (36) from the Mandalay region, and a molded tablet (74) from Pagan like (7) and (8).

The *b*-spellings on those Buddhist objects may be etymological spellings reflecting Indic *b* rather than evidence for the loss of implosion. They may also reflect a different borrowing of the same Indic word. Multiple borrowed versions of a word are not evidence for synchronic allophony: e.g. Japanese *ku* and *kyū* 'nine' were borrowed from two different dialects of Chinese in different periods and are not evidence for *u* ~ *yū* allophony.

The use of *b* for *ḅ* in a single loanword in (36) cannot be a dialectal trait of the Pyu of what is now the Mandalay region, as (32), a much longer text from the general area, contains a mix of the two characters. (The seven legible *akṣara*s of (36) do not contain *ḅ*, and (66), the only other text from the Mandalay region, is short and contains neither character.) Unfortunately, none of the *ḅ* and *b akṣara*s in (32) can be glossed with certainty, so I do not know if *ba* (in the collocation *tin·ṁ ba* '?'; 32.7, 32.8 [2×], 32.10) is an alternative spelling of the reduced negative morpheme *ḅa* (16.4C). What may be the full negative morpheme *ḅaḥ* (16.3A) may be spelled consistently in (32); there are no instances of *baḥ* there or anywhere in the corpus except (21.5) where its meaning is unknown.

For now I phonemicize Pyu /ɓ/ and /b/ according to spellings and acknowledge the possibility that the two may have merged in Late Pyu or even as early as Middle Pyu if the *baṁḥ* of (37) is the honorific.

If Pyu had /ɓ/ – with or without a [b] allophone – did it also have other implosives? Katō (2005) answered this question in the affirmative; he reconstructed *ɠ and *ɗ corresponding to *g* and *d* with subscript dots in the script. He

also reconstructed *ʔj[37] and *ʔw corresponding to *y* and *v* with subscript dots in the script. As with *ɓ, he did not explictly state his reasoning for any of those reconstructions, though his Pyu-West Pwo comparisons imply he was projecting West Pwo initials back into Pyu (Table 11).

Table 11. Katō's glottalic consonant comparisons

Griffths' et al.'s Pyu transliteration	Katō's transliteration	Katō's reconstruction	Katō's gloss	Blagden's gloss	West Pwo
kdiṃ	*kḍi*	*kăɗi	causative	(unknown)	*ɗɤɯʔ*
tduṃ	*tḍū*	*tăɗu	water	water	*thì*
tḅaḥ	*tƀa:* *	*tăɓa:	again	auxiliary	Geba *ɓā*
daṃḥ	*ḍa:*	*ɗa:	attributive	(unknown)	*dà*
daṃṁ	*ḍȧ* **	*ɗaʔ	and then	probably a verb or auxiliary indicating the past	*taʔ*
diṃṁ	*ḍi̇* ***	*ɗī	to assemble	to assemble	East Pwo *dèiɴ*
doṃ	*ḍo*	*ɗo	great	(unknown)	*dò*
pdaṃ	*pḍa*	*păɗa	flat	(unknown)	*dá*
ḅa	*ƀa*	*ɓa	to believe	(unknown)	*ɓá*
ḅaḥ	*ƀa:*	*ɓa:	should	possibly an optative negative	*ɓà*
ḅuḥ	*ƀū:*	*ɓu:	offering (n.)	to do (?)	*ɓòuɴ*
rvaṃḥ	*ẉe:*	*ʔwe:	beloved	second half of a word possibly meaning 'queen'	*ʔwì*

* I have substituted ƀ for Katō's non-Unicode b with a stroke through the bottom half. Katō transliterates *visarga* as a colon. He does not provide a phonetic interpretation of *visarga*, though he does provide correspondences between *visarga* and tones in Karenic languages.

** Katō transliterates the *anusvāra* as a single superscript dot and interprets it as a glottal stop or vowel nasalization.

*** I am unable to replicate Katō's dot over a dotted *i*, so I have substituted Blagden's (1919) similar *i̇* for that combination.

There are several problems with Katō's data.

37. I have converted Katō's non-IPA ʔy to *ʔj for consistency with my own IPA-based notation.

First, the basis of his Pyu glosses differing from those of Blagden (1919) is unclear. At least one of Katō's glosses is definitely incorrect: in a Sanskrit-Pyu bilingual inscription (16.3A), *ḅaḥ* is a gloss for Sanskrit *na* 'not' (Griffiths et al. 2017b).

Second, he did not gloss the Karenic forms, so it is impossible without a knowledge of those languages to judge the semantic gaps between their words and his proposed Pyu cognates.

Third, there are unexplained irregular correspondences in his cognate sets: e.g. his reconstructed Pyu *ɗ corresponds to West Pwo *th* ('water') and *t* ('and then') as well as *ɗ*, and his reconstructed Pyu *u corresponds to West Pwo *i* ('water') as well as *ouɴ*.

Fourth, his data provide no support for his typologically improbable *ɠ, a sound not in any Southeast Asian language to the best of my knowledge,[38] or his *ʔj. I presume those reconstructions are by analogy with his reconstructions for other consonants written with subscript dots.

There are two Pyu-internal arguments against Katō's set of glottalized consonants.

The first is statistical. One-third of *g* in the corpus (54/157) was written with a subscript dot. Whatever that combination indicated was probably less marked than Katō's *ɠ. 56% of *v* in the corpus (72/129) was written with a subscript dot. It is unlikely that a glottalized *ʔw was more common than a plain *w. In Baxter & Sagart's (2014a: 69) Old Chinese reconstruction, there is no †ʔʷ, and pharyngealized *ʔˤʷ is "rare" compared to *ʔ. There are no forms with *ʔˤʷ in Baxter & Sagart's (2014a: 69) list of nearly 5,000 Old Chinese reconstructions.

The second is orthographic. On the one hand, it is plausible that the Pyu wrote glottalized *ʔj and *ʔw with subscript dots like implosives because they regarded glottalized sonorants and implosives to belong to the same phonemic class, and in fact I argue in my article on Pyu rhymes (Miyake 2018) that all consonants written with a subscript dot belong to a single phonemic class (fricatives). On the other hand, if Katō is correct, it is puzzling why the Pyu wrote *ɓ with a special character *ḅ* instead of *b* with a subscript dot. *b* has no strokes beneath the line that would impede the writing of a subscript dot. Why not write all consonants of the same class with the same diacritic? If the Pyu had borrowed *ḅ* for /ɓ/ from the Mon script, why didn't they also borrow the Mon practice of writing /ɗ/ as retroflex *ḍ*? The Pyu did not use *ḍ* to write non-Indic words until a late period, perhaps as recently as the 11th century, long after *ḅ* was created (§ 7.3.12). Hence one cannot argue that retroflex *ḍ* was not available for /ɗ/ at an earlier period.

38. The phonetically and geographically closest consonant I know of is /ʔg/ in the speech of some Sui speakers in Guizhou (Edmondson 2004).

The uniqueness of *ḅ* in the Pyu script suggests that *ḅ* represented a unique phoneme. I am hesitant to identify that phoneme as /ɓ/ because /ɓ/ is almost always paired with /ɗ/ in Southeast Asia to the best of my knowledge. The same argument applies to /ʔb/. The only languages with no implosive other than /ɓ/ in the UCLA Phonological Segment Inventory Database are in Africa with the exception of Jacaltec in Guatemala (Maddieson & Precoda 1984). However, I shall continue to regard *ḅ* as a symbol for /ɓ/, as I cannot think of any better alternative. Pyu may have been like the Myeik dialect of Burmese which has [ɓ] as an allophone of /b/ but has no other implosives as phonemes or allophones (Katō and Khin Pale 2012).

7.3.13 *The fricatives /ɣ ç ʝ ð f v/*

I discovered these six phonemes when trying to resolve the question of whether Pyu had tones or register. Hence I discuss them in the section on suprasegmentals of my article on Pyu rhymes (Miyake 2018) even though fricatives are, of course, segments.

7.4 Peripheral initials

These rare written initials are probably anomalous spellings of the phonemes listed in Table 7.

7.4.1 *k-ṃ*

There is only one *akṣara* with *k-ṃ: kkut·ṃ* '?' (20.3). There are no other instances of voiceless stops with subscript dots. I could mechanically interpret *k-ṃ* as /x/ by analogy with *g-ṃ* for /ɣ/, but I would rather not posit a phoneme on the basis of a single spelling. Perhaps *k-ṃ* represents a voiceless allophone [x] of initial /ɣ/ after preinitial /k./ or of the preinitial /k./ immediately before initial /k/.

If the subscript dot is an error, its motivation is obscure. It cannot be an error for a superscript dot (i.e. *anusvāra*) since there is no phoneme written *uṁ*; all instances of *uṁ* in the corpus are probably errors for *uṃ* with a subscript dot. The *akṣara*s immediately surrounding *kkut·ṃ* (*pay·ṁḥ* and *tiṁ*) have *ṁ* but not *ṃ*, so a subscript dot was not accidentally carried over from a preceding or following *akṣara*.

7.4.2 *gh*

There are only two *akṣara*s with *gh: gha* '?' (17.1 and 20.2) and *ghu* '?' (100). Inscriptions (17) and (100) lack final subscript consonants. Hence in theory *gha* and *ghu* might represent borrowings of Indic words with the shapes *ghă̆C(C)a* and *ghŭC(C)a*. However, the Pyu typically retained final *-a* in borrowings from

Indic, so Indic disyllables would most likely have retained their disyllabicity in Pyu. There are no common Indic words of the shape *ghŭC(C)a*, and *gha* cannot be an Indic borrowing in inscription 20, a text with final subscript consonants. There is no Sanskrit or Pali word *gha* or *ghā* other than the Sanskrit particle *gha* 'indeed' which would be an extremely unlikely candidate for borrowing. Therefore I suspect that the *gha* in 17.1 and *ghu* are both native or at least non-Indic.

I reject *gh* as a symbol for /gʱ/ on typological grounds. Although there are a few *akṣara*s with *dh* (§ 7.4.10), there are none with *bh*, and I know of no language with /gʱ/ and /dʱ/ but not /bʱ/. Also, the extremely low frequency of *gh* makes me reluctant to posit a phoneme to account for it. Finally, Indic /gʱ/ appears as *g* in Pyu *saga-* for *saṁgha-* 'monastic community' (7.17 and 8.17), implying that /g/ was the closest Pyu equivalent of Indic /gʱ/.

gh in *gha* and *ghu* may be akin to <gh> in Thai ฆา <ghā¹> /kʰâː/ < Proto-Tai *qaːᶜ 'kill' (Pittayaporn 2009: 357): a pseudo-Indic spelling for a native phoneme that is normally written with another character. In the case of Thai, the /kʰ/ of "kill" would have been etymologically spelled <kh>,[39] and in the case of Pyu, *gh* might have been etymologically spelled *g* or *g-ṃ* /ɣ/. However, without knowing what *gha* means, it is impossible to equate it to attested *g-* and *g-ṃ* forms such as *ga* 'if' (7.25) and *gaṃ* 'to plunge into' (16.3A). *ghu* may or may not be a variant spelling of *gu* '?' (39.3). The *akṣara* †*guṃ* is not attested.

7.4.3 *j-ṃ*

There is only one *akṣara* with *j-ṃ: jaṃ* '?' (167). *j-ṃ* may be a variant spelling of *y-ṃ* /ɟ/, but there is no *akṣara* †*yaṃ* /ja/ in the corpus. The choice of *j* instead of *y* may indicate that *j-ṃ* represents an allophone of /ɟ/ that was phonetically closer to /ɟ/ than /j/: e.g. [ʑ] (if /ɟ/ was [ɟ] or [dʑ]), [ʐ] (if /ɟ/ was [dʐ]), [ʒ] (if /ɟ/ was [dʒ]), or [z] (if /ɟ/ was [dz]).

7.4.4 *jñ*

The *akṣara rjña* '?' appears twice in a row in (161) which has no subscript (i.e. *akṣara*-final) consonants. There are no other instances of *jñ* in the corpus. Although there are a few instances of voiced stop-nasal sequences in the corpus (§ 6.2.3 and § 6.2.5), *rjñ* is the sole attested *r*-voiced stop-nasal sequence. It would be unusual if Pyu had /r.ɟɲ/ but not /ɟɲ/ without a preinitial. I conclude that /ɟɲ/ was not in native Pyu words, though it is possible that Pyu speakers trained in Sanskrit could pronounce it.

---

39. The non-etymological spelling of Thai 'kill' with ฆ <gh> and ◌ˋ <¹> (the latter normally for tone *B words) must postdate the devoicing of */g/, the earlier Thai phoneme associated with ฆ <gh>, and the merger of tones *C2 and *B1 into the falling tone.

rjña could be an *r*-prefixed borrowing of a Sanskrit word from √ *jñā* 'know' such as *jñāna-* 'knowledge' or some Middle Indic descendant like Pali *ñāṇa-*: e.g. /r.ɲan/ with an unwritten final /n/. The *jñ* could be a Sanskritized spelling of a permissible Pyu initial /ɲ/.

rjña could also be a native Pyu word /r.ɲa(C)/ respelled to appear as if it were Sanskrit, possibly due to folk etymology if its meaning had something to do with knowledge.

7.4.5 *ṭ*

There are only three *akṣara*s with *ṭ*: *ṭaḥ* '?' (12.1), *ṭi* '?' (160), and *ṭiḥ* '?' (64.2). They do not appear to be part of any polysllabic Indic loanword.

Retroflex stops are a prominent feature of the Indian linguistic area, but Pyu lies outside of it, so a primary retroflex phoneme /ʈ/ is unlikely. Pyu /ɖ/ was a late, secondary development (§ 7.3.6), as were retroflexes in Chinese and Vietnamese.[40]

If *ṭ* is not a pseudo-Indic spelling of /t/,[41] it may represent a [ʈ]-like allophone of /l̥/ (normally spelled *hl*) or /ʟ̥/ (normally spelled *ṭl*). *ṭlaḥ* /Lah/ '?' (27.3) and *hliḥ* /l̥ih/ 'bond' (16.5d) may be alternate spellings of *ṭaḥ* and *ṭiḥ*. This Pyu [ʈ]-like consonant would be similar to Middle Vietnamese *tr* /ʈ/ which partly originated from earlier *tl-clusters.

7.4.6 *ṭh*

There is a single *akṣara Cṭha* '?' (37) with aspirated retroflex *ṭh* after an illegible preinitial. Like unaspirated *ṭ*, aspirated *ṭh* may be an alternate spelling of /l̥/ or /ʟ̥/. It may represent /ɭ̊/, the voiceless counterpart of the retroflex /ɭ/ that is unique to (37) (§ 7.3.8). (There is no *hl̥* in the corpus, though there is a Late Pyu *hḍ* /D/ which may be from /l̥/ and/or /ʟ̥/ [§ 7.3.7].)

The Pali loanwords *mahaṭhe* < *mahāthera-* 'chief monk' and *-disaṭhe* < *-tissatthera-* 'senior monk Muggaliputta' (both 7.15) with hypercorrect retroflex *ṭh* for dental *th* in the Kubyaukgyi inscription imply that Late Pyu lacked /ʈʰ/ and that its speakers perceived Indic /ʈʰ/ as /tʰ/.

---

40. Baxter & Sagart (2014a: 80–81) describe the development of retroflexes in Middle Chinese. Middle Vietnamese *tr* /ʈ/ is from *Cl-clusters and Middle Vietnamese /ʂ/ is from *Cr-clusters (Gage 1985).

41. This would be without parallels in Burmese and Thai which lack pseudo-Indic spellings with <ṭ> for /t/, though they do have pseudo-Indic spellings with other characters: e.g. Burmese ဘုရား <bhurāḥ> /pʰəjá/ and Thai ณ <ṇaḥ> /náʔ/.

7.4.7 *ts*

At first glance *ts* may appear to be a candidate for a unit phoneme /ts/, but *ts* does not pattern like other initials. It has no aspirated or voiced counterparts, and it is never preceded by a preinitial: e.g. †*kts* or †*mts*. I regard *ts* as a representation of a preinitial-initial sequence /t.s/ that may have been pronounced [təs]. /t.s/ is not to be confused with [ts] as a possible phonetic value for /c/.

ts is probably not a spelling reflecting a [ts]-like pronunciation of *c* in bare initial position. None of the completely legible *ts-akṣara*s have near-homographs with *c* instead of *ts*: e.g. there is no †*caṁḥ* corresponding to *tsaṁḥ* 'deed' (16.2C, 16.5C, and possibly also with that meaning in 20.1 and 73), the most common *ts-akṣara*. *ts* is not in complementary distribution with *c*. Both *ts* and *c* occur with *a, i*, and *e*. Hence *ts* cannot represent an allophone of /c/ conditioned by a following vowel. *tsaṁḥ* /t.säh/ 'deed' may be a combination of *saṁḥ* /säh/ 'to do' (7.20 and 8.20) with a nominalizing prefix /t./.

7.4.8 *dr-ṃ*

dr-ṃ appears in five *akṣaras*:

- *draṃ* '?' (67)
- *dran·ṃ* '?' (32.4)
- *droṃḥ* '?' (20.5)
- *kdraṃḥ* '?' (18)
- *ndrom·ṃ* (in *[v]r[el]· ndrom·ṃ* 'kindness' or 'to be kind'?; 16.2A)

If the dots are not errors, *dr-ṃ* in the first three *akṣara*s may be interpreted as *d-ṃ* plus *r* for [ðr], a pronunciation of /t.r/ with a voiced allophone of /t/ like the voiced allophone of /p/ I proposed to account for the anomalous preinitials *b, ḅ*, and *v* (§ 6.2.4–6.2.6).

Such an interpretation cannot apply to the last two *akṣara*s because /k.t.rah/ and /n.t.rom/ would have double preinitials, violating my formula for Pyu sesquisyllabic structure (§ 5).

Given that *ṃ* indicates a fricative, I could interpret *dr-ṃ* as a fricative allophone [z̨] of *dr* /R/ whose default pronunciation may have been an affricate [dz̨] (§ 7.3.10). The last two *akṣara*s would then have single preinitials before /z̨/: /k.Rah/ [kz̨ah] and /n.Rom/ [nz̨om]. If [z̨] was also possible in bare initial position, the first three *akṣara*s may have been /Ra/ [z̨a], /Ran/ [z̨an], and /Roh/ [z̨oh]. Perhaps [dz̨] weakened to [z̨] between voiced segments. Two of the three instances of initial *dr-ṃ* (*dran·ṃ* and *droṃḥ*) were respectively preceded by *tar·* /tar/ and *tin·ṁ* /tïn/ which both end in voiced sonorants. The *akṣara* preceding the third instance (*draṃ*) is illegible.

7.4.9 *dl-ṃ*

dlin·ṃṁ '?' (27.2) is the only *akṣara* with *dl-ṃ*. It occurs in the collocation *tin·ṁ dlin·ṃṁ* '?' which occurs elsewhere as *tin·ṁ dlin·ṁ* '?' without a subscript dot (20.2, 20.4, 27.4). It is likely that a subscript dot was first inscribed by mistake and that the correct superscript dot (i.e. *anusvāra*) was then added.

If the subscript dot is not an error, *tin·ṁ dlin·ṃṁ* could be a play on *tin·ṁ dlin·ṁ*, and the dot may modify *d* rather than *l* or *dl* as a unit: i.e. *dl-ṃ* is equivalent to *d-ṃ* plus *l*. *dl-ṃ* may represent a pronunciation of /t.l/ as [ðl] with the voiced allophone of /t/ that I proposed for *dr-ṃ* (§ 7.4.8). Such an interpretation is more parsimonious than one involving a lateral fricative reading of *l-ṃ* or *dl-ṃ* as [ɮ] that is only attested once.

7.4.10 *dh and dh-ṃ*

There are six *akṣara*s with voiced aspirated *dh: dha* '?' (66.1, 100.1), *dhaṃ* '?' (167), *dhaṅ·* '?' (66), *dhat·* '?' (64.7), *dhiy·ṁ* '?' (158), and *ndha* '?' (27.2, 27.6). *dhat·* may be a borrowing of Sanskrit or Pali *dhātu* 'element'. *dha* in (100) and *dhaṃ* in (167) may be alternate spellings of *dhat·* without final subscript consonants. However, *dha* in (66.1) cannot be an alternate spelling of *dhat·* since (66) contains final subscript consonants, and the other *akṣara*s do not resemble anything in Indic. Neither Sanskrit nor Pali permit initial *ndh*, and *ndha* cannot be a monosyllabic reduction of Sanskrit *nidhana* 'end' since †/n.dan/ would have been written as †*ndhan·* in (27) which contains final subscript consonants.

Although it would not be impossible for Pyu to have /dʱ/ as its sole voiced aspirate like White Hmong, it is more likely that *dh* and *dh-ṃ* are variant spellings of /ð/ or pseudo-Indic spellings of /d/ akin to the <dh> for /d/ in the spelling of the native Burmese word ဓား <dhāḥ> /dá/ 'knife'.

7.4.11 *n-ṃ (or t-ṃ?)*

tniṃṁ '?' (20.3) is the only *akṣara* with *n-ṃ*. Perhaps †*tniṁ* was intended and a subscript dot was accidentally written first followed by the correct superscript dot. But that explanation cannot be verified unless †*tniṁ* is found and the meanings of both †*tniṁ* and *tniṃṁ* are known.

If the subscript dot of *tniṃṁ* was intended, it might modify *t* rather than *m*, and *tn-ṃ* may represent a pronunciation of /t.m/ with a fricative allophone [θ] of preinitial /t/. If preinitial /t/ had voiced allophones before voiced initials (§ 6.2.3), its voiceless fricative allophone [θ] might have occurred before voiceless initials, and *tn-ṃ* might phonemically be /t.n̥/ with a voiceless initial /n̥/.

7.4.12 *ḅ-ṃ*

ḅaṃ '?' (151) is the only *akṣara* with *ḅ-ṃ*. Perhaps *ḅaṁ* '?' (98) was intended, but that explanation cannot be verified until the hapax legomena *ḅaṃ* and *ḅaṁ* are glossed. *ḅaṁ* may be an error for the honorific *ḅaṁḥ*.

If the subscript dot of *ḅaṃ* was intended, *ḅ-ṃ* might have represented a bilabial [ɓ]-like allophone of /v/: e.g. [β]. *ḅaṃ* would then be an alternate spelling of *vaṃ* /va/ 'to go' (16.1A) or a homophone of that word.

7.4.13 *m-ṃ (or t-ṃ?)*

tmay·ṃḥ '?' (25.6) is the only *akṣara* with *m-ṃ*. Perhaps †*tmay·ṁḥ* was intended, but that explanation cannot be verified unless †*tmay·ṁḥ* is found and the meanings of both †*tmay·ṁḥ* and *tmay·ṃḥ* are known.

If the subscript dot of *tmay·ṃḥ* was intended, it might modify *t* rather than *m*, and *tm-ṃ* may represent a pronunciation of /t.m/ with a fricative allophone [θ] of preinitial /t/. If preinitial /t/ had voiced allophones before voiced initials (§ 6.2.3), its voiceless fricative allophone [θ] might have occurred before voiceless initials, and *tm-ṃ* might phonemically be /t.m̥/ with a voiceless initial /m̥/.

7.4.14 *ṣ*

ṣ appears only once in the corpus in the *akṣara kṣat·* '?' (56) which may be a borrowing of Sanskrit *kṣatriya* 'military caste'. Even if *kṣat* is native, it would be risky to posit a retroflex phoneme /ʂ/ based on a single example, and the misspelling of Sanskrit *pauruṣa* 'valor' with retroflex *ṣ* as *paurusa* (16.1A) with alveolar *s* suggests that Pyu lacked a retroflex sibilant.

It is far more likely that *kṣ* is simply a Sanskritized spelling of /k.s/ with a retroflex *ṣ* in place of the alveolar *s* that is not permissible after *k* in Sanskrit. Such a spelling of /ks/ also existed in Old Mon (Shorto 1971: xiv) and persists in Khmer to this day.[42] Unfortunately, no potential alternate spelling †*ksat·* is in the corpus, though there are other *ks-akṣara*s: e.g. *ksa* (20.3) which is in a text with final subscript consonants and therefore cannot be a spelling of /k.sat/ with a final /t/.

A continuation of this article, in which the fricative onsets /ɣ ç ʝ ð f v/ and rhymes or Pyu will be discussed in detail, has been published as Miyake (2018).

---

42. Khmer subscript <s> is clearly derived from the extinct retroflex character ᩋ <ṣ> and not from alveolar ᩈ <s>. This subscript <ṣ> is the sole allograph of <s> in the environment <Cs>. Similarly, in Old Mon, "/s/ in second position is almost invariably written as [retroflex] ṣ" (Shorto 1971: xiv).

Acknowledgments

This study was conducted as part of the 'Beyond Boundaries: Religion, Region, Language and the State' project (ERC Synergy Project 609823 ASIA) under the supervision of Nathan W. Hill with cooperation from Arlo Griffiths and Julian K. Wheatley of the 'From Vijayapurī to Śrīkṣetra? The Beginnings of Buddhist Exchange across the Bay of Bengal as Witnessed by Inscriptions from Andhra Pradesh and Myanmar' project funded by the Robert H.N. Ho Family Foundation.

References

Baxter, William H. & Sagart, Laurent. 2014a. *Old Chinese: A new reconstruction*. Oxford: Oxford University Press.

Baxter, William H. & Sagart, Laurent. 2014b. *Baxter-Sagart Old Chinese reconstruction, version 1.1*. (http://ocbaxtersagart.lsait.lsa.umich.edu/BaxterSagartOCbyMandarinMC2014-09-20.pdf) (Accessed 2014-09-30.) https://doi.org/10.1093/acprof:oso/9780199945375.001.0001

Benedict, Paul K. 1972. *Sino-Tibetan: A conspectus* (Princeton-Cambridge Studies in Chinese Linguistics II). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511753541

Blagden, C.O. 1909. The Talaing inscription of the Myazedi pagoda at Pagan, with a few remarks on the other versions. *Journal of the Royal Asiatic Society of Great Britain and Ireland* 41(4). 1017–1052. https://doi.org/10.1017/S1356186300081402

Blagden, C.O. 1911. A preliminary study of the fourth text of the Myazedi inscriptions. *The Journal of the Royal Asiatic Society of Great Britain and Ireland* 43(2). 365–388. https://doi.org/10.1017/S0035869X00041526

Blagden, C.O. 1913–1914. The *Pyu* inscriptions. *Epigraphia Indica* 12 (No. 16). 127–132. (Published as a quarterly supplement to the "Indian Antiquary".)

Blagden, C.O. 1919. The Pyu face of the Myazedi inscription at Pagan. In Taw, Sein Ko & Duroiselle, Chas. (eds.), *Epigraphia Birmanica: Being lithic and other inscriptions of Burma*, vol. 1 (Part 1, Archaeological Survey of Burma), 59–68. Rangoon: Superintendent, Government Printing.

Bucknell, Roderick S. 1994. *Sanskrit manual: A quick-reference guide to the phonology and grammar of classical Sanskrit*. Delhi: Motilal Banarsidass.

Coblin, W. South. 1994. *A compendium of phonetics in northwest Chinese*. Journal of Chinese Linguistics monograph series 7. Berkeley: Project on Linguistic Analysis, University of California.

Diffloth, Gérard. 1984. *The Dvaravati Old Mon language and Nyah Kur*. Bangkok: Chulalongkorn University Printing House.

Gage, William W. 1985. *Vietnamese in Mon-Khmer perspective*. In Ratanakul, Suriya, Thomas, David D., & Premsrirat Suwilai. (eds.), *Southeast Asian linguistics presented to André-G. Haudricourt*, 493–524. Bangkok: Mahidol University.

Griffiths, Arlo & Miyake, Marc & Wheatley, Julian K. 2017a. Corpus of Pyu inscriptions. (http://hisoma.huma-num.fr/exist/apps/pyu/index2.html) (Accessed 2017-11-07).

Griffiths, Arlo & Hudson, Bob & Miyake, Marc & Wheatley, Julian K. 2017b. Studies in Pyu epigraphy, I: State of the field, edition and analysis of the Kan Wet Khaung mound inscription, and inventory of the corpus. *Bulletin de l'École française d'Extrême-Orient* 103. 43–205. https://doi.org/10.3406/befeo.2017.6247

Hall, Christopher. 2003. *Modern German pronunciation: An introduction for speakers of English*. Manchester: Manchester University Press.

Huffman, Franklin E. 1970. *Cambodian system of writing and beginning reader with drills and glossary*. New Haven: Yale University Press.

Jacques, Guillaume. 2004. *Phonologie et morphologie du japhug (rGyalrong)*. Paris: Université Paris Diderot (Paris 7). (Doctoral dissertation.)

Jäschke, Heinrich August. 1881. *A Tibetan-English dictionary: With special reference to the prevailing dialects; To which is added an English-Tibetan vocabulary*. London: Secretary of State for India in Council.

Jones, Robert B., Jr. 1961. *Karen linguistic studies: Description, comparison, and texts* (University of California Publications in Linguistics 25). Berkeley: University of California Press.

Karlgren, Bernhard. 1957. Grammata serica recensa. *Bulletin of the Museum of Far Eastern Antiquities* 29. 1–332. (First published in 1940 in *Bulletin of the Museum of Far Eastern Antiquities* 12.)

Kato, Atsuhiko (加藤昌彦). 2005. *Pyū-go to Karen-kei shogengo no ruijisei* ピュー語とカレン系諸言語の類似性 [The similarity of Pyu and the Karenic languages]. (http://user.keio.ac.jp/~kato/Pyu_Karen.pdf) (Accessed 2015-06-04.)

Kato, Atsuhiko (加藤昌彦) & Khin Pale. 2012. The Myeik (Beik) dialect of Burmese: Sounds, conversational texts, and basic vocabulary. *Journal of Asian and African Studies* 83. 117–160. (http://repository.tufs.ac.jp/bitstream/10108/69339/1/jaas083005_ful.pdf) (Accessed 2020-01-08.)

Li, Fanwen (李範文). 1986. *Tongyin yanjiu* 同音研究 [Studies of homophones]. Yinchuan: Ningxia People's Publishing House.

Luce, Gordon Hannington. 1985. *Phases of pre-Pagán Burma: Languages and history*. Oxford: Oxford University Press.

Maddieson, Ian & Precoda, Kristin. 1984. *The UCLA Phonological Segment Inventory Database (UPSID)*. (http://web.phonetik.uni-frankfurt.de/upsid_info.html) (Accessed 2017-10-04.) https://doi.org/10.1017/CBO9780511753459.012

Matisoff, James A. 1973. Tonogenesis in Southeast Asia. In Hyman, Larry M. (ed.), *Consonant types and tone* (Southern California Occasional Papers in Linguistics No. 1), 71–95. Los Angeles: Linguistics Program, University of Southern California.

Matisoff, James A. (ed.). 1987–Ongoing. *The Sino-Tibetan etymological dictionary and thesaurus*. Berkeley: University of California. (http://stedt.berkeley.edu/~stedt-cgi/rootcanal.pl) (Accessed 2017-11-07.)

Miyake, Marc Hideo. 2012. Complexity from compression: A sketch of pre-Tangut. In Popova, Irina (Попова, Ирина) (ed.), *Tanguty v central'noj azii: sbornik statej v čest' 80-letija prof. E. I. Kyčanova* (*Тангуты в Центральной Азии: Сборник статей в честь 80-летия проф. Е.И.Кычанова*) [Tanguts in Central Asia: A collection of articles marking the 80th anniversary of Prof. E. I. Kychanov], 244–261. Moscow: Oriental Literature.

Miyake, Marc Hideo. 2017a. A sketch of Pyu synchronic phonology. (Paper presented at the 27th Annual Meeting of the Southeast Asian Linguistics Society, Padang, 11–13 May 2017.)

Miyake, Marc Hideo. 2017b. Pyu unique non-Indic akṣaras (version 171109). (Data set). https://doi.org/10.5281/zenodo.1044075 (Accessed 2017-11-09.)
Miyake, Marc Hideo. 2018. Studies in Pyu phonology, II: rhymes. *Bulletin of Chinese Linguistics* 11(1–2). 34–76. https://doi.org/10.1163/2405478X-01101008
Pittayaporn, Pittayawat. 2009. *The phonology of proto-Tai*. Ithaca: Cornell University. (Doctoral dissertation.)
Pulleyblank, Edwin G. 1991. *Lexicon of reconstructed pronunciation: In Early Middle Chinese, Late Middle Chinese, and Early Mandarin*. Vancouver: UBC Press.
Sagart, Laurent. 1999. *The roots of old Chinese*. Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.184
Shafer, Robert. 1943. Further analysis of the Pyu inscriptions. *Harvard Journal of Asiatic Studies* 7(4). 313–366. https://doi.org/10.2307/2717831
Schütz, Albert J. 1994. *The voices of Eden: A history of Hawaiian language studies*. Honolulu: University of Hawaii Press.
Short, David. 1993. Slovak. In Comrie, Bernard & Corbett, Greville G. (eds.), *The Slavonic languages*, 533–592. London: Routledge.
Shorto, Harry L. 1971. *A dictionary of the Mon inscriptions: From the sixth to the sixteenth centuries, incorporating materials collected by the late C. O. Blagden* (London Oriental Series 24). London: Oxford University Press.
Starostin, Sergej Anatol'evič (Старостин, Сергей Анатольевич). 1989. *Rekonstrukcija drevnekitajskoj fonologičeskoj sistemy* (*Реконструкция древнекитайской фонологической системы*) [Reconstruction of the Old Chinese phonological system]. Moscow: Nauka.
Svantesson, Jan-Olof. 1983. *Kammu phonology and morphology* (Travaux de l'Institut de Linguistique de Lund 18). Lund: Liber. (Doctoral dissertation.)
Tha Myat. 1963. *Pyū phat cā: Pyū akkharā samuiṅḥ* [Pyu reader: A history of the Pyu alphabet]. Yangon: U Hla Din, National Printing Works.
Thomas, David. 1992. On sesquisyllabic structure. *The Mon-Khmer Studies Journal* 21. 206–210.
Yabu Shirō. 2014. Professor Nishida Tatsuo and the study of Tibeto-Burman languages. *Memoirs of the Research Department of the Tōyō Bunko* 72: 179–205.

Author's address

Marc Miyake
amritavira@gmail.com

Publication history

Date received: 13 September 2017
Date accepted: 16 April 2018