

A HYPOTHESIS ON THE ORIGIN OF OLD CHINESE PHARYNGEALIZATION

Laurent Sagart and William H. Baxter

CNRS/ INaLCO (Paris) and University of Michigan (Ann Arbor)

1. The reconstruction of OC pharyngeal consonants

Baxter and Sagart (2014) follow Norman (1994) in assigning pharyngealization to OC type-A words, and absence of pharyngealization to type-B words. As Norman showed, the main piece of evidence in support of an OC pharyngealization contrast consists of phonetic changes having affected OC onset consonants, and somewhat later, OC vowels, in Han times. Thus, strictly speaking, all the evidence relating to the phonetic nature of the A/B contrast dates to terminal Old Chinese/Han times. We have no direct evidence as to what the contrast may have been earlier.

While Norman regarded pharyngealization as a property of entire syllables, Baxter and Sagart treat it as a feature of the initial consonant of an OC word's main syllable. This is because type-A and type-B words rhyme, apparently freely, in early and late OC poetry: it would be unexpected for pharyngealized vowels to rhyme with nonpharyngealized ones. Pharyngealization had to be located outside of a main syllable's rhyme: presumably in its onset. Accordingly Baxter and Sagart (2014) reconstruct two parallel series of OC main syllable initial consonants, nonpharyngealized and pharyngealized (Table 1):

plain (type B)	p	t	ts			k	k ^w	q	q ^w	ʔ
	p ^h	t ^h	ts ^h	s		k ^h	k ^{wh}	q ^h	q ^{wh}	
	b	d	dz			g	g ^w	ɣ	ɣ ^w	
	m	n			l	r	ŋ	ŋ ^w		
	m _◦	n _◦			l _◦	r _◦	ŋ [◦]	ŋ ^{◦w}		
pharyngealized (type A)	p ^ʕ	t ^ʕ	ts ^ʕ			k ^ʕ	k ^{wʕ}	q ^ʕ	q ^{wʕ}	ʔ ^ʕ
	p ^{hʕ}	t ^{hʕ}	ts ^{hʕ}	s ^ʕ		k ^{hʕ}	k ^{whʕ}	q ^{hʕ}	q ^{whʕ}	
	b ^ʕ	d ^ʕ	dz ^ʕ			g ^ʕ	g ^{wʕ}	ɣ ^ʕ	ɣ ^{wʕ}	
	m ^ʕ	n ^ʕ			l ^ʕ	r ^ʕ	ŋ ^ʕ	ŋ ^{wʕ}		
	m _◦ ^ʕ	n _◦ ^ʕ			l _◦ ^ʕ	r _◦ ^ʕ	ŋ ^{◦ʕ}	ŋ ^{◦wʕ}		

Table 1: Nonpharyngealized and pharyngealized onset consonants in Old Chinese. Source: Baxter and Sagart (2014:69).

As Baxter and Sagart observe (2014:73-74), this is typologically unusual: languages with pharyngealized consonants generally have more nonpharyngealized consonants than pharyngealized ones. We think of typologically unusual features as short-lived, unstable ones, but not necessarily impossible ones. It is entirely possible that the onset consonant inventory in Table 1 existed only for a short period of time at the end of the OC period.

2. A model of the origin of OC pharyngealization

We outline below a hypothetical model of how such a contrast may have arisen. Briefly, we propose that the pharyngealized consonants arose out of clusters of plain consonants followed by a pharyngeal fricative segment [ʕ], that is: *CʕV- > C^ʕV-. We assume that Cʕ- type clusters existed for all OC onset consonants.¹ At a still earlier stage, we derive these *CʕV- onsets from strings in which two copies of a vowel flanked [ʕ]: e.g. Caʕa-, Ciʕi-, Cuʕu- etc., or in other words, a geminate vowel

¹ Baxter and Sagart (2014:73) observe that an alternative which “we did not retain yet do not wish to exclude” to pharyngealized consonants are Cʕ- clusters having a pharyngeal fricative as their second segment.

interrupted by a pharyngeal fricative. The *Cʕ- clusters, and ultimately the pharyngealized onset consonants we reconstruct, were formed when the vowel's first copy fell, thus:

Type A	Type B
stage 1 > stage 2 > OC	stage 1 > stage 2 > OC
Caʕa- > Cʕa- > C ^s a-	Ca- > Ca- > Ca-
Ciʕi- > Cʕi- > C ^s i-	Ci- > Ci- > Ci-
Cuʕu- > Cʕu- > C ^s u-	Cu- > Cu- > Cu-

Table 2: A hypothetical model of the formation of the OC pharyngealization contrast.

3. Lushai and Kuki-Chin vowel length and OC pharyngealization

The Lushai (Kuki-Chin) vowel length contrast has been hypothesized to correlate with the Chinese A/B word-type contrast (Starostin 1989:327sq), as shown below:

1. Lushai long vowel : OC type A
2. Lushai short vowel : OC type B

If the PST precursor of correlation 1 above was CV_iʕV_i-, it is easy to derive the Lushai vowel length distinction simply through loss of the pharyngeal segment (Figure 1).

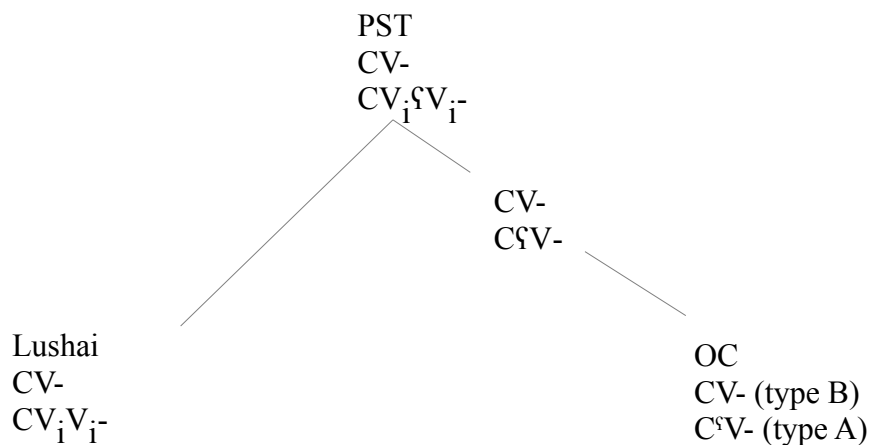


Figure 1: a hypothetical model relating OC pharyngealization and Lushai vowel length.

4. Bimoraicity in Austronesian

The reconstructible free content morphemes of PAN overwhelmingly consist of words of two syllables or more. Meaning-associated monosyllables exist too, but either reduplicated, or as formatives at the end of longer words: Austronesianists call them ‘roots’ (for collections of roots see Blust and Trussell, ongoing; Wolff 2010). For instance root *-sek ‘cram, crowd’ (Blust) occurs in PAN *seksek ‘cram in’, PMP *hasek ‘jam, cram, crowd’, *be(n)sek ‘overcrowded’, PWMP *dasek, *deseq ‘compress’; as well as in words attested in a single language, e.g. Banjarese *barasak* ‘overcrowded’: but simple monomoraic monosyllables straightforwardly derived out of root *-sek ‘cram, crowd’ cannot be found. Preservation of the root in recognizable form at the end of words in *hasek, *be(n)sek etc., as opposed to drastic phonological reduction of nonfinal elements in compounds, suggests that PAN was stress-final. In the examples above, the highly eroded lexical material preceding the root: *ha-, *be(n)-, *da-, *de-, *bara-* is typically not recognizable. Evidently roots like *-sek once existed as free morphemes, without reduplication or compounding, and have been largely driven out of the modern languages. Two languages: Bunun (Taiwan) and Cebuano (Philippines) preserve some of them: their phonetic shape in these languages indicates that as free morphemes they were realized with a geminate vowel interrupted by a glottal stop, presumably to satisfy a constraint against monomoraic free forms. For instance Cebuano *suʔík* ‘remote corner, being deep inside’ is argued by Wolff (2007) to be the regular reflex of root *-sek. Bunun retains more of these ‘stretched’ monosyllabic roots as free forms, with geminate vowel and optional glottal stop (Table 3):

PAN root	monomoraic bound forms (underlined>	Bunun bimoraic ('stretched') free forms
*-taq 'earth'	PWMP *bu <u>R</u> taq 'earth, ground, land', Amis <i>ditaʔ</i> 'clay', <i>po<u>t</u>aʔ</i> 'soil, dirt', <i>so<u>t</u>aʔ</i> 'earth, dirt; mud; land'	<i>ta(?)aq</i> 'mud'
*-Ras 'nutritious substance'	PAN *be <u>R</u> as 'dehusked grain', Bunun <i>ti<u>l</u>as</i> 'crops of cereal', Cebuano <i>lu<u>g</u>ás</i> 'grain'	<i>la(?)as</i> 'food substance; red meat; flesh of a fruit'
*-baw 'above'	PAN *ba <u>b</u> aw 'above', PAN *sa <u>b</u> aw 'raise', PMP tim <u>b</u> aw 'be at the top', Cebuano <i>ig<u>b</u>áw</i> 'at the top of something'	<i>ba(?)av</i> 'high geographical location; up in the mountains'
*-Riq 'sword grass'	*e <u>R</u> iq 'Imperata cylindrica, sword grass'	<i>li(?)iq</i> 'sword grass'

Table 3: PAN monosyllabic roots, with bound forms (monomoraic) and corresponding Bunun free forms (bimoraic).

Personal pronouns also show traces of this alternation. The root *-Su '2sg' occurs unchanged as the last syllable of di- or polysyllabic forms Thao *ihu* 'thou', Atayal *isuʔ* 'you', Amis *k-iso* 'you', Kanakanabu *ii-kasu* 'thou/you'; and, as a free form, with vowel gemination in Bunun *su(?)u* 'you', Saisiyat *shoʔo* 'you', Tsou *súu* 'thou'. Similarly, root *-mu '2pl' occurs unchanged in Atayal *simu* 'you and he/she', Saaroa *ila-mu* 'you', Bunun *a-mu* 'you (pl.)', Amis *ka-mo* 'you (pl.)', Kavalan *i-mu* 'you (pl.)', but with vowel gemination in the free forms Bunun *mu(?)u* 'you (pl.)', Tsou *múu* 'you (pl.)'.

5. Bimoraicity in Austroasiatic

A constraint against monomoraic free nominals was identified in Proto-Austroasiatic by Zide (2002), who termed it 'the bimoraic constraint'. The strategies used by Austroasiatic and Austronesian to make monosyllables compatible with the bimoraic constraint are very similar. This can be illustrated by Munda forms from Anderson (2004):

	Gutob	Juang	Gorum	Sora
hand	<i>titi</i>	<i>iti</i>	<i>siʔi</i>	<i>sʔi</i>
foot	<i>susuj</i>	<i>ijiñ/ŋ</i>	<i>jiʔiŋ</i>	<i>jʔeŋ</i>

Table 4: The bimoraic constraint in Munda (source: Anderson, 2004)

In Table 4 we see monosyllabic roots for ‘hand’ and ‘foot’ made bimoraic by reduplication in Gutob, by the prothesis of *i-* in Juang and by the gemination of the vowel with insertion of a glottal stop in Gorum. The Sora forms appear to have lost a vowel before the glottal stop: this makes them similar to type A, stage 2, in Table 2.

6. OC type A/B as bimoraicity

We propose that the A/B distinction, in the form that we attribute to PST (Figure 1), continues an even earlier contrast in a language *ancestral* to PST, where the very bimoraic constraint that we saw in PAA and in PAN existed. In that language type-B forms would have had two syllables, and type-A forms would have had the ‘stretched monosyllable’ CV_i?V_iC structure we saw in Bunun and in Gorum (note that this structure includes an intervocalic glottal stop, not a pharyngeal fricative). For instance the OC pair type-B 入 *n[u]p ‘enter’ and type-A 內 *n^h[u]p ‘bring or send in’ which contain the same monosyllabic root, would go back to PST type-B #nup vs. type-A #nuʔup; these in turn would go back to pre-PST #σ-nup (disyllable) vs. type-A #nuʔup (stretched monosyllable). If the bimoraic constraint is formulated so as to prohibit not just free forms of one mora, but free forms with an odd number of morae, and if the language satisfied the constraint by ‘stretching’ both monosyllables and the last syllable of trisyllabic free forms, then PST *nuʔup could go back to either of pre-PST *nuʔup or *σ-σ-nuʔup. Evolution into PST would then involve (1) reduction of non-final syllables, and (2) change of intervocalic glottal stop to pharyngeal fricative -VʔV- > -VʕV-, perhaps under interference from a neighbor language with an ATR/RTR distinction.

7. Testing the correlation between Kuki-Chin vowel length and OC type A/B

This model supposes that Starostin’s conjecture of a correlation between the Chinese word-type contrast and the Kuki-Chin vowel length contrast is valid. Below we describe our statistical

investigation of that question. The null hypothesis we wish to test is the following:

Null hypothesis: there does not exist any positive correlation between PKC long vowel and OC type A on the one hand, and between PKC short vowel and OC type B on the other hand.

We compare the Old Chinese A/B distinction in the reconstruction of Baxter-Sagart (2014) with vowel length, not in Lushai, but in Kuki-Chin, to which Lushai belongs. To that end we use VanBik's reconstruction of Proto-Kuki-Chin (VanBik, 2009). The benefits are (1) that the number of Kuki-Chin candidate forms is limited, making it more difficult to hand-pick examples; and (2) that single-language irregularities tend to be ironed out.

We scanned the Proto-Kuki-Chin material for Chinese cognates. This was done in agreement with known regularities of OC-TB comparison and Kuki-Chin phonological history. Because Kuki-Chin verbal morphology affects vowel length in ways that are only partly understood, we limited our purview to nominals: VanBik provides a list of those (2009:451-515). We excluded open-vowelled syllables, all long in PKC (VanBik, 2009:323). In addition, there is no length contrast among vowels preceded by /i/ or /u/: we excluded these forms, too. The remaining forms are all CVC or CVVC monosyllables. Also excluded from comparison are

- PKC words with long and short variants, e.g. 'elbow' *ki(i)w 3, 'egg' *du(u)y 4, *tu(u)y 4, 'yard, armspan, cord' *la(a)m 4;
- OC words with A/B variants, e.g. 入 *n[u]p 'enter' and 內 *n[u]p 'bring or send in'; 糲 *[r]at and 糲 *[r]at-s 'dehusked but not polished grain'
- OC words of uncertain type: 髟 *s(ʰ)ram 'long hair';
- probable loanwords: 'silver', PKC *ŋuun, OC 銀 *ŋrə[n];

- comparisons requiring large semantic shifts: ‘pig’, PKC *wok 3 vs. 富 *pək-s > pjuwH > fù ‘rich; wealth’.

43 comparisons in total were retained (see Appendix). A 2 x 2 table was built from the contingents in the four possible categories (long-type A; long-type B; short-type A; short-type B), as follows:

	PKC long	PKC short
Chinese type A	6	6
Chinese type B	5	26

Table 5: contingency table for Chinese word type vs. Kuki-Chin vowel length.

The P value, or significance of the deviation from the null hypothesis, was calculated from Table 5 using Fisher’s exact test (<http://www.langsrud.com/fisher.htm>). Application of the test returns the P value 0.03203186503710432, i.e. approximately 0.32. A P value under 0.05 is usually regarded as significant in scientific works. It means that the probability of obtaining the numbers in Table 5 under the null hypothesis is lower than 5%. Here, the probability is approximately 3.2%. In other words, on present evidence there probably exists a positive correlation between Kuki-Chin long-short and Old Chinese A/B, as Starostin proposed. We are aware that this result would be different if other cognates were identified. It may be necessary in future to modify our list and to calculate the P value again. The predominance of the Kuki-Chin short-OC type B category is probably robust, however.

8. Explanation of mismatches

Eleven comparisons in Table 5—almost one in four—behave contrary to prediction. They are italicized in the Appendix. The preferred explanation for these mismatches is that we are dealing not with exact cognates but with free forms having odd vs. even numbers of syllables, both ending in

the same monosyllabic root. For instance PKC *raʔ ‘fruit’ and OC 糲 *[r]at-s ‘dehusked but not polished grain’ would both go back to the monomoraic final syllable of a pre-PST disyllable #σ-ras, while OC *[r]ʔat would go back to the pre-ST stretched monosyllable #raʔas. Evidently this explanation has the potential of accounting for A/B variants in Chinese and for long-short variants in Kuki-Chin.

Abbreviations

AA: Austroasiatic

AN: Austronesian

OC: Old Chinese

PAA: Proto-Austroasiatic

PAN: Proto-Austronesian

PKC: Proto-Kuki-Chin

PMP: Proto-Malayo-Polynesian

PST: Proto-Sino-Tibetan

PWMP: Proto-Western-Malayo-Polynesian.

References

- Anderson, Gregory S. 2004. Advances in Proto-Munda reconstruction, *Mon-Khmer Studies* 34:159-184.
- Baxter, William H. and Laurent Sagart. 2014. *Old Chinese: a new reconstruction*. New York: Oxford University Press.
- Blust and Trussell, ongoing. Online at <http://www.trussell2.com/acd/> (accessed Oct. 26, 2015).
- Norman, Jerry [Luó Jiérui 罗杰瑞]. 1994. Pharyngealization in Early Chinese. *Journal of the American Oriental Society* 114:397-408.

- Starostin, Sergei A. (1989) *Rekonstrukcija drevnekitajskoj fonologičeskoj sistemy* [A reconstruction of the Old Chinese phonological system]. Moscow: Nauka, Glavnaya Redakcija Vostočnoj Literatury.
- VanBik, Kenneth. 2009. *A reconstructed ancestor of the Kuki-Chin languages*. STEDT monograph 8. Berkeley, University of California.
- Wolff, John U. 2007. "Petrified prefixes in the Austronesian languages". *Studies in Philippine Languages and Cultures* 1,15:25-31.
- Wolff, John U. 2010. *Proto-Austronesian phonology with glossary*. 2 vols. Ithaca: Cornell Southeast Asia program publications.
- Zide, Norman H. 2002. "Issues in Proto-Munda and Proto-Austroasiatic nominal derivation: The Bimoraic constraint." In Marlys A. Macken [ed.] *Papers from the 10th Annual Meeting of the Southeast Asian Linguistics Society*. Tempe, Arizona State University, South East Asian Studies Program, Monograph Series Press, 55-74.

APPENDIX

	PKC gloss	PKC	Old Chinese > Middle Chinese > Modern Standard Chinese	
1.	waist	taay 1	帶 *C.tʰa[t]-s > tajH > dài 'girdle, strap' (Lushai <i>tai</i> 'waist', <i>tai-von</i> 'wear in the belt')	IA
2.	crab	?aay 1	蟹 *m-kʰre? > heaX > xiè 'crab'	IA
3.	fat, grease	thaaw 3 / 4	臊 *[m.s]ʰaw > saw > sāo 'fat of swine or dog'	IA
4.	lip, beak, mouth	hmuur 3	門 *mʰə[r] > mwon > mén 'gate, door'	IA
5.	chaff	waay 4	簸 *pʰaj? > paX > bǒ 'winnow (v.)'	IA
6.	layer	pleep 2	牒 *lʰep > dep > dié 'tablet'	IA
1.	monkey	ngaaw 4	禺 *ŋ(r)o-s > ngjuH > yù 'monkey'	IB
2.	child	naaw 4	孺 *no-s > nyuH > rù 'child'	IB
3.	appearance, face, shape	hmeel 4	面 *C.me[n]-s > mjienH > miàn 'face'	IB
4.	snake	ruul 4	虺 *[r]u[j]? > xjw+jX > huǐ 'snake-brood' (<i>W dialect</i> : *r _o - > MC x-)	IB
5.	bat	baak 1	蝠 *[p]ək > pjuwk > fú 'bat'	IB
1.	earth	lay 1	地 *[l]ʰej-s > dijH > dì 'earth, ground'	sA
2.	bridge	lay 3 hlay 3	梯 *lʰ[ə]j > thej > tī 'stairs'	sA
3.	stump, base	bul 4	本 *C.pʰə[n]? > pwonX > běn 'tree trunk'	sA

4.	ditch, hole, pit, well	khur 4 khor 4	款 * ^h [k] ^h [o]r? > khwanX > kuǎn ‘empty, hollow; hole’	sA
5.	ruler, chief	mang (no assigned tone)	猛 *m ^h raŋ? > maengX > měng ‘fierce’ (also ‘eldest’)	sA
6.	dog	ɣuy 1	犬 *[k] ^{wh} [e][n]? > khwenX > quǎn ‘dog’ (probably *[k] ^{wh} [e]r?)	sA
1.	water	tuy 1	水 *s.tur? > sywijX > shuǐ ‘water; river’ (E dialect: *-r > *-j)	sB
2.	rope, cord	ruy 1 hruy 1	纍 *[r]uj? > lwijX > lěi ‘a creeping plant’	sB
3.	tree	thing 1	薪 *[s]i[n] > sin > xīn ‘firewood’	sB
4.	louse	hrik 3	蝨 *srik > srit > shī ‘louse’	sB
5.	house	ɣim 1	窞 *q(r)[ə]m-s > ‘imH > yin ‘subterranean room’ (Shuo Wen 地下室也)	sB
6.	dream	mang 1	夢 *C.məŋ-s > mjuwngH > mèng ‘dream’	sB
7.	hair (body)	mul 1 hmul 1	眉 *mr[ə][r] > mij > méi ‘eyebrow’	sB
8.	maggot, worm	lung 1	蟲 *C.lruŋ > drjuwng > chóng ‘insect’	sB
9.	thigh	phay 3	腓 *[b]əj > bj+j > fēi ‘calf of the leg’	sB
10.	tail	may 1	尾 *[m]əj? > mj+jX > wěi ‘tail’	sB
11.	pillow	kham khum (no assigned tone)	枕 *[t.k][ə]m? > tsyimX > zhěn ‘pillow’	sB
12.	bank, shore	kam 4	岑 *N-q ^h (r)[ə]m? > ngimX > cén ‘high river bank’	sB
13.	mouth	kam 4	垠 *[k] ^h (r)[ə]m? > khimX > qīn ‘pit’	sB
14.	land, country, forest	ram 4	林 *[r]əm > lim > lín ‘forest’	sB
15.	mat	pher 4	編 *pe[r] > pjien > biān ‘plait, weave’	sB
16.	three	thum 4	三 *s.rum > sam > sān ‘three’ (*s.r- > s-; infl. by 四 *s-?)	sB
17.	bear	wom 4	熊 *C.[g] ^w (r)əm > hjuwng > xióng ‘bear (n.)’	sB
18.	inside, within, household	tshung 4	中 *truŋ > trjuwng > zhōng ‘center’ (on the initial correspondence see Benedict 1972)	sB
19.	name	ming 4 hming 4 min 4 hmin 4	名 *C.meŋ > mjieng > míng ‘name’	sB
20.	door, gate	kot 3	闕 *k ^{wh} at > khjwot > què ‘tower over gate’	sB
21.	bone	ru? 3	律 *[r]ut > lwit > lù ‘law, rule (n.)’ (semantics via ‘long tubular bone’ > ‘pitch-pipe’ see Sagart paper)	sB
22.	armpit	yak 3	亦 *[g](r)Ak > yek > yì ‘armpit’	sB
23.	deer (male), antler	yuk 3	鹿 *mə-r ^o k > luwk > lù ‘deer’	sB
24.	two	ni? 3 hni? 3	二 *ni[j]-s > nyijH > èr ‘two’	sB

25.	six	ruk 3	六 *k.ruk > ljuwk > liù 'six'	sB
26.	eye	mik 3	目 *C.m(r)[u]k > mjuwk > mù 'eye'	sB