

(draft, please do not cite without author's consent)

OC *-r in early Chinese loans to Bùyāng, and related issues

Laurent Sagart, CRLAO (emeritus), Paris

Based on a proposal of Starostin's (1989), Baxter and Sagart (2014) reconstruct a single liquid coda in OC, noted *-r. This *-r evolves to MC *-n or *-j depending on dialect. In certain Tai and Mon-Khmer languages, some Chinese loanwords with the OC *-r coda end in -i, even when the MC pronunciation has -n. We pointed out that "If the original coda had been [n], there is no reason it could not have been borrowed with [n]".¹ The fate of OC *-r in Chinese loans to Bùyāng 布央, and in Càijiā 蔡家, is similar.

1. Layers of Chinese loanwords in Langjia Bùyāng

Bùyāng 布央 is a Kra-Dai language spoken in Yúnnán. It belongs to the Kra (aka 圪央 Ge-yang) branch. Its Lángjià 朗架 dialect was described by Lǐ Jǐnfāng 李锦芳 (1999). Bùyāng has borrowed extensively from Chinese since first contact between Chinese and Kra-Dai, which presumably occurred when Qín Shǐ Huáng's armies established a Chinese state in the region, around 200 BCE. A preliminary analysis of the stratification of Chinese loanwords based on the reflection of Chinese tones and segments points to at least six layers of direct or indirect Chinese loanwords.

	Yīnpíng 陰平	Yángpíng 陽平	Yīnshǎng 陰上	Yángshǎng 陽上	Yīnqù 陰去	Yángqù 陽去
Early I	11		24	24 (some 11)	11	
Early II	54	312	24		11	
Zhuàng	24	11	11	11	54/31	54/31
Late Middle Chinese	54	11	(24 ?)	?/31(QZ)	11	312
Mandarin I	33	31	54	54/24(QZ)	24	
Mandarin II	11	24			11	

Table 1. Layers of Chinese loanwords in Lángjià Bùyāng (in the Yángshǎng column, a contour followed by the mention '(QZ)' is conditioned by a voiced obstruent initial).

It is not entirely clear whether the Early I and II layers are chronologically distinct or whether they represent distinct contemporary sources. The Chinese twelve earthly branches, used as year names in Chinese imperial times, are part of the Early I layer:

¹ "Words with OC *-r that were borrowed early into other languages sometimes lack the final -n that we find in Middle Chinese, suggesting either that the donor variety of Chinese was one that had changed *-r to *-j, or else that the borrowing language had no [r]. For example, for the cyclical sign 辰 chén < *[d]ər 'fifth earthly branch', Li (1945:336) cites these forms from Tai languages: Ahom shi, Lü si¹, Dioi chi². Khmu (a minority Mon-Khmer language spoken primarily in Laos) has /sii/ (Damrong and Lindell 1994:104). If the original coda had been [n], there is no reason it could not have been borrowed with [n]." Baxter and Sagart (2014:255).

子 * $[ts]əʔ > tsiX > zǐ$ ‘1st earthly branch’	$tsuə^{24}$	rat
丑 * $[ŋ]ruʔ > trhjuwX > chǒu$ ‘2d earthly branch’	$pjau^{24}$	ox
寅 * $[ɕ](r)ər > yin > yín$ ‘3rd earthly branch’	$ŋi^{11}$	tiger
卯 * $m^sruʔ > maewX > mǎo$ ‘4th earthly branch’	mou^{24}	hare
辰 * $[d]ər > dzyin > chén$ ‘5th earthly branch’	$ɛi^{11}$	dragon
巳 * $s-[ɕ]əʔ > ziX > sì$ ‘6th earthly branch’	$ɛau^{24}$	snake
午 * $[m].q^həʔ > nguX > wǔ$ ‘7th earthly branch’	ha^{24}	horse
未 * $m[ə]t-s > mj+jH > wèi$ ‘8th earthly branch’	mut^{11}	goat
申 * $[i]n > syin > shēn$ ‘9th earthly branch’	$ɛɛn^{11}$	monkey
酉 * $N-ruʔ > yuwX > yǒu$ ‘10th earthly branch’	$ðou^{11}$	cock
戌 * $s.mi[t] > swit > xū$ ‘11th earthly branch’	$phit^{54}$	dog
亥 * $[g]^səʔ > hojX > hài$ ‘12th earthly branch’	$kuə^{11}$	pig

Table 2: Borrowed year names in Bùyāng

Bùyāng has also borrowed extensively from Zhuàng 壯, the locally dominant Kra-Dai language, and a member of the Tai subgroup. As Zhuàng loans themselves include much Chinese material, a Zhuàng layer is included in Table 1 above. Chinese loans to Zhuàng were made over as long a period as have loans to Bùyāng: material borrowed by Bùyāng from Zhuàng was potentially borrowed at the time of the earliest contact with Chinese. The Late Middle Chinese and Mandarin layers were borrowed more recently.

2. Treatment of OC *-r in loans to Bùyāng

Four words from the Early and Zhuàng layers which end in -n in Middle Chinese have -i in Bùyāng:

BS 2014	Bùyāng old Chinese loans	layer
辰 * $[d]ər > dzyin > chén$ ‘5th earthly branch’	$ɛi^{11}$ ‘year of the dragon’	Early I
寅 * $[ɕ](r)ər > yin > yín$ ‘3rd earthly branch’	$ŋi^{11}$ ‘year of the tiger’	Early I
蒜 * $[s]^sər-s > swanH > suàn$ ‘garlic’	$ɛu:i^{54}$ ‘garlic’	Zhuàng
犬 * $[k]^{whs}[e][n]ʔ > khwenX > quǎn$ ‘dog’	$qɔi^{24}$ ‘dog’	Early II

Table 3: Bùyāng -i for OC *-r > MC -n

The first three words in Table 3 are reconstructed with OC *-r, and the fourth, with *-[n], standing for either with *-n, *-ŋ or *-r, with *-n the preferred option.

The normal Bùyāng treatment of OC *-n appears to be -n in the same layers.

Baxter-Sagart (2014)	Bùyāng old Chinese loans	Layer
盪 *ʔun > 'won > wēn 'kind, warm (adj.)'	<i>u.n¹¹</i> 'warm leftovers'	Early I
錢 *N-ts[a][n] > dzjen > qián 'coin, money'	<i>ma⁰ ɛn¹¹</i> 'money'	Early I
填 *[d]i[n] > den > tián 'to obstruct'	<i>tɛn¹¹</i> 'to fill'	Early I
申 *i[n] > syin > shēn '9th earthly branch'	<i>ɛn¹¹</i> 'year of the monkey'	Early I
辛 *[s]i[n] > sin > xīn 'pungent; painful'	<i>ɛin^{54 2}</i>	Early II
躐 *[n]r[a][n]ʔ > nrjenX > niǎn 'trample'	<i>qa⁰ nia:n¹¹</i> 'to crush'	Early I

Table 4

An exception is 斤 *[k]ər > kj+n > jīn 'axe; catty', treated by Bùyāng as qa:n⁵⁴ 'ax', apparently in the Early II layer, despite having been reconstructed with *-r. This word cannot be part of the Late Middle Chinese layer, on account of its uvular initial. With this unexplained exception, the Bùyāng loan data are consistent with the view that the earliest Chinese donor(s) to that language distinguished between OC *-r and *-n. This argues that the coda in the OC word for 'dog': 犬 *[k]^{whs}[e][n]ʔ (BS 2014) was *-r.

3. 'dog'

One might be tempted to dismiss the comparison between 犬 quǎn 'dog' and Bùyāng *qɔi²⁴* as phonetically adventurous. Yet absence of aspiration in *qɔi²⁴* is expected in an early Chinese loanword, as no aspirated stops can be reconstructed at the earliest layer of KD; the uvular initial is regular for a Chinese type-A word borrowed before Late Middle Chinese, whether the OC initial was velar or uvular; and the tone matches Chinese shǎngshēng in a Layer-II word.

4. 'dog' in Càijiā.

We have collateral from Càijiā 蔡家, an archaic Chinese dialect of Guìzhōu, for an *-r coda in 'dog'. Càijiā was recently investigated by Lǚ Shanshan and Miguel Cortiço dos Santos. They were generous to share with me the results of their fieldwork. They recorded 'dog' as *khwi⁵⁵*. Tone 55 is the main Càijiā reflex of OC -ʔ, suggesting we are dealing with the Chinese word, rather than with a loan from a neighboring ST language. With one counterexample in the Càijiā 200-word-list at my disposal (*ŋi³³* 'person', OC *niŋ), Càijiā reflects OC *-in and *-iŋ, perhaps also OC *-en, as -iN (/iN/) or -iN (/jiN/):

2 This form occurs as part of a disyllable *ɛin²⁴ hɔ³³* 'difficult, hard' < 辛苦 in a text at the end of Li's book (Li Jinfang 1999:259). Both syllables are consistent with the Early II layer. This word was probably borrowed as a disyllabic morpheme.

ηiN^{31}	‘year’	年 *C.n ^h i[ŋ]
$k^h iN^{55}$	‘day’	天 *t ^h i[ŋ]
ϵiN^{33}	‘new’	新 *C.si[n]
$m iN^{31}$	‘people’	民 *mi[ŋ]
$k^h iN^{33}$	‘to pull’	牽 *[k] ^h si[n]
$k iN^{33}$	‘to see’	見 *[k] ^h en-s

Table 5: *Caijia* -iN/-iN reflecting OC *-in and *-en

In general, with exceptions,³ -N is the expected Càijiā treatment of OC nasal endings *-m, *n, *-ŋ.

The treatment of OC *-r is different. Outside of a single example of a word reconstructed with OC *-r and ending in Càijiā -N: 近 *N-kər? > gj+nX > jìn ‘near’, Càijiā $t\epsilon iN^{33}$, Càijiā loses *-r, even when its MC reflex is -n: ‘dry’ 乾, Càijiā ka^{33} , OC *[k]^har; probably also in ‘liver’ 肝, Càijiā ka^{33} , OC *s.k^ha[r], and in ‘to live’ 存, Càijiā $tə^{31}$, OC *[dz]^hə[n]. Loss of OC *-r appears to be the normal Càijiā development. This implies, again, that the word for ‘dog’: *khwi*⁵⁵ developed out of an OC prototype with *-r coda.

5. Correspondence of OC *-r in Written Tibetan.

Hill (2014) examines the correspondences in Tibetan of OC *-r as reconstructed by Baxter and Sagart (2014). He finds two: one to Tibetan *-r, for which he reconstructs *-r in the parent language (I use the term ‘Proto-Sino-Tibetan’); and another to Tibetan *-l, for which he uses the index *-rl. If we now assign *-r to the Old Chinese word for ‘dog’ we seem to need a third correspondence: OC *-r to Tib. -zero, to account for Tib. ཁྱི khyi ‘dog’. Three comparisons are known to me which exhibit that correspondence:

dog	犬 *[k] ^{wh} ə[e]r? > khwenX > quǎn ‘dog’	Tib. ཁྱི khyi
water	水 *s.tur? > sywijX > shuǐ ‘water; river’	Tib. རྩུ chu
lip	脣 *sə.dur > zywin > chún ‘lip’	Tib. མཚུ mchu

Table 6. OC *-r to Tibetan zero after an OC high vowel

Hill’s evidence for the correspondence OC *-r = Tib. *-r involves only words with nonhigh vowels, *a, *e, *ə, *o on the Chinese side (the OC reconstructions and Tibetan transcriptions have been updated; the evidence is limited to words B&S reconstruct with *-r or *-r[ɹ]):

with OC *a

2. Chi. 竿 *k^har > kan > gān ‘pole, rod’ (0139k), Tib. མཁམ་ mkhar / འཁམ་ khar ‘staff, stick’
3. Chi. 難 *n^har-s > nanH > nán ‘difficulty’ (0152d), Tib. མཚན་ mnar ‘suffer, be tormented’

³ OC *-eŋ is reflected as /i/: t^hi⁵⁵ ‘hear’ < *t^hi^heŋ (tone !); ɛi³³ ‘star’ < *s-ts^heŋ; mi31 ‘name’ < *C.meŋ.

5. Chi. 燔 *[b]ar > bjon > fán ‘burn, roast’ (0195i), Tib. འབར 'bar ‘burn, blaze’ ‘shine’
 6. Chi. 獻 *ɲar-s > xjonH > xiàn ‘wise man’ (0252e), Tib. མྱར snar ‘intelligent, quick of apprehension’
 53. Chi. 鮮 *[s][a]r > sjen > xiān ‘fresh; good’ (0209a), Tib. གསར gsar ‘new’

With OC *e

54. Chi. 霰 *s[e]r-s > senH > xiàn ‘sleet’ (0156d), Tib. སེར ser ‘hail’

With OC *o

55. Chi. 裹 kwaX < *s.[k]ʰo[r]ʔ (0351d) ‘wrap (v.)’, Tib. སྐོར skor ‘go around’

With OC *ə

56. Ch. 飛 *Cə.pə[r] > pj+j > fēi ‘fly (v.)’ (dialect: *-r > *-j) (0580a) and other words meaning ‘to fly’, Tib. འཕྱར ḥphur ‘fly (v.)’

We add a few more examples with OC *a, *ə, *o:

翰 *[g]ʰar > hanH > hàn ‘white (of a horse)’ (MC implies *-s) : Tib. དཀར dkar ‘white, whitish’

播 *pʰar-s > paH > bò ‘to sow’ : Tib. འབྲོབ 'bor, pf. བྲོབ bor ‘to throw, cast, fling’

璦 *tsʰarʔ > tshaX > cuǒ ‘white and brilliant’ : Tib. འཇར 'char ‘to shine’

門 *mʰə[r] > mwon > mén ‘gate, door’ : Tib. མུར mur ‘gills’

鑽 *[ts]ʰor > tswan > zuān ‘perforate, penetrate’ : Tib. སོར sor ‘gimlet’

Only two doubtful examples of OC *-r : Tib. -r involving high vowels on the Chinese side can be cited:

糜 *mʰur > mwon > mén ‘red millet’ : Tib. དམར dmar ‘red’

君 *C.qur > kjun > jūn ‘lord; ruler’ : Tib. བཀུར bkur ‘to honour, esteem; honour, respect, homage, mark of honour’.

The first example is doubtful (although the comparison itself is valid) because MC mwon is equally compatible with OC *mʰər, which is a good match for Tib. dmar. We should amend our OC reconstruction from *mʰur to *mʰ[ə]r. The second example is in all likelihood erroneous (the comparison itself is erroneous) because 君 jūn ‘lord; ruler’ relates to the action of governing (尹 *m-qurʔ > ywinX > yǐn ‘govern; governor’) rather than to honor. Thus there are no secure examples of OC *-ir, *-ur : corresponding to Tib. words ending in *-r.

6. Complementary distribution

The two correspondences, Hill's OC *-r : Tib. -r and my OC *-r : Tib. zero are thus in complementary distribution with respect to the preceding vowel:

PST	preceding vowel in PST	OC	Tib.
*-r	nonhigh	*-r	-r
	high	*-r	zero

Table 7 complementary distribution between two Chinese-Tibetan sound correspondences

7. The correspondence OC *-r : Tib. zero beyond Chinese and Tibetan.

What happens to ST words with high vowels followed by the coda *-r in other ST languages ? In Table 8 we add forms from Bodo, Lushai and Karen, representing the Sal, Kuki-Chin-Naga and Karen groups of ST, respectively.

Tib.	OC (BS, revised)	Bodo (Bhat)	Lushai (Lorrain)	Proto-Karen (LuangThongkum)
ཁྱི khyi 'dog'	犬 *[k] ^{whs} [e]r? 'dog'	səy(má) 'dog'	ui 'dog'	thwi B 'dog'
ཆུ chu 'water'	水 *s.tur? 'water'	dəy 'water'	tui 'water'	thej A 'water'
མུ mchu 'lip'	脣 *sə.dur 'lip'	(gusu)təy 'lip'	—	n.a.

Table 8: OC *-r : Bodo -y : Lushai -i : Karen *-i/-j : Tib. zero

Table 8 shows that Bodo, Lushai and Karen have *-i or *-j corresponding to OC *-r, Tib. zero. We know that Tib. has lost final *-j. Thus in Tibetan, loss of ST *-r in all likelihood occurred through a *-j stage.

An informal survey of forms for the etyma in Table 8 in other ST languages shows mostly j-type endings. Consequently as I argued [here](#), the correspondence

$$\text{OC } *-r : \text{Tib. zero} : \text{other ST } -i/-j$$

probably reflects PST *-r after high vowels: 'water' and 'lip' from PST *-ur, 'dog' from PST *-ir. We would like to disambiguate the OC vowel to *[k]^{whs}ir?, but under the bottom-up reconstruction principles adopted in Baxter and Sagart 2014, we must remain with *[k]^{whs}[e]r? until Chinese-internal evidence appears that the vowel was *i.

8. Conclusion: a third candidate for a non-Chinese phonological innovation

Thus it appears that after high vowels, the PST coda *-r was maintained in Chinese but changed to *-j, merging with original *-j in all Sino-Tibetan languages outside of Chinese. This constitutes another candidate for a shared innovation of non-Chinese ST. Another candidate was proposed in Handel (2002), yet another in Sagart (2017).

References

- Baxter, William H. and Laurent Sagart. 2014. *Old Chinese: a new reconstruction*. New York: Oxford University Press.
- Handel, Zev. 2002. Rethinking the medials of Old Chinese: Where are the r's? *Cahiers de Linguistique Asie Orientale* 31, 1:3-32.
- Hill, Nathan. 2014. Cognates of Old Chinese *-n, *-r, and *-j in Tibetan and Burmese. *Cahiers de Linguistique Asie Orientale*, 43, 2:91-109.
- Lǐ Jīnfāng 李锦芳. 1999. *Buyang Yu Yanjiu* 布央語研究. Beijing: Zhongyang Minzu Daxue.
- Sagart, Laurent. 2017. A candidate for a Tibeto-Burman innovation. *Cahiers de Linguistique Asie Orientale* 46, 101-119.
- Starostin, Sergey A. 1989. *Rekonstrukcija drevnekitajskoj fonologičeskoj sistemy* [A reconstruction of the Old Chinese phonological system]. Moscow: Nauka, Glavnaya Redakcija Vostočnoj Literatury.