

Is Shom Pen a distinct branch of Austroasiatic?

[DRAFT SUBMITTED FOR A SPECIAL ISSUE OF THE JOURNAL OF MON-KHMER STUDIES]

Roger Blench
Kay Williamson Educational Foundation
8, Guest Road
Cambridge CB1 2AL
United Kingdom
Voice/ Fax. 0044-(0)1223-560687
Mobile worldwide (00-44)-(0)7967-696804
E-mail rogerblench@yahoo.co.uk
<http://www.rogerblench.info/RBOP.htm>

Cambridge, Thursday, 14 January 2010

TABLE OF CONTENTS

1. INTRODUCTION.....	1
2. THE SHOM PEN LANGUAGE.....	1
3. PROPOSALS FOR LEXICAL COGNATES.....	1
4. THE INTERPRETATION OF THESE RESULTS.....	10
REFERENCES.....	11

1. Introduction

The Shom Pen language is spoken on Great Nicobar by a few hundred hunter-gatherers. Until recently, the language was too poorly known to make any definitive statement about its affiliations, although it has usually been considered part of Nicobarese. However, the availability of new materials (Chattopadhyay & Mukhopadhyay 2003), have made possible a more extended analysis. I have argued in print (Blench 2007) that the Shom Pen language may well be an isolate, not a relative of Nicobarese. In this paper¹ I would like to put forward an alternative view that Shom Pen may be a distinct branch of Austroasiatic, but not part of Nicobarese. The basis for this are cognates with Austroasiatic that do not appear to be shared with Nicobarese. I owe these insights largely to discussions with Gerard Diffloth in January 2009 and to further comments from Paul Sidwell in October 2009. However, the actual argument given here is my own.

2. The Shom Pen Language

Until recently, the language of the Shom Pen had remained unknown apart from ca. 100 words recorded by De Roepstorff (1875), the scattered lexical items in Man (1886) and the comparative list in Man (1889). Although our knowledge of Nicobarese is imperfect there are several book-length sources for this group, for example Whitehead (1925), Das (1977) and Radakrishnan (1981). Although most reference books list Shom Pen as part of the Nicobarese group, evidence for this is slight. Apart from some numerals and body parts, Shom Pen words do not show any obvious relationship with other Nicobarese languages or other Mon-Khmer languages. The fragmentary evidence does not immediately suggest that the Shom Pen are Austroasiatic-speakers. Man (1886:436) says; ‘of words in ordinary use there are very few in the Shom Pen dialect which bear any resemblance to the equivalents in the language of the coast people’. Man’s Shom Pen data shows that numbers 1-5 are roughly cognate with Nicobarese but that above this they are quite different. Man (1886) also observes that there was substantial linguistic variation between Shom Pen settlements;

In noting down the words for common objects as spoken by these (*dakan-kat*) people I found that in most instances they differed from the equivalent used by the Shom Pen of Lafal and Ganges Harbour.

Chattopadhyay & Mukhopadhyay (2003), makes available a significant body of new data on the Shom Pen language. While not to modern standards of presentation and analysis, it is enough to construct a better argument for the affiliation of Shom Pen. Blench (2007) sets out their data in an IPA-like transcription and notes some obvious etymologies. The lexical data analysed is the same as that in my previous paper, but I cite only the words for which there appears to be a supplementary etymology. Comparisons are made using the Austroasiatic datasets in Shorto (2006) as well as additional lexical material available on the Mon-Khmer Etymological Dictionary on the SEALANG website. Muṅḍā cognates were established using David Stampe’s unpublished comparative Muṅḍā materials. Individual lexical citations are not referenced, as their sources can be easily determined from the SEALANG website. Shorto’s proposed proto-forms are cited for the convenience of the reader; this should not be taken to imply that I accept the reconstruction.

Apart from Shorto (2006), which was compiled from materials left behind by the author after his death, there are very few published comparative Austroasiatic datasets, especially those including Muṅḍā cognates. Shorto’s starred forms can be misleading, since his reflexes often come from only one or two branches of Austroasiatic and therefore may be regional or sub-branch forms. Apart from contributing to the classification of Shom Pen, this paper also puts forward extended Austroasiatic datasets as part of a broader strategy, to understand the classification of this important and still neglected phylum.

3. Proposals for lexical cognates

In the following list, the main citations are from Chattopadhyay & Mukhopadhyay (2003) while comparisons from De Roepstorff (1875) are marked R. and from Man (1886, 1889) M. Where the item is

¹ A preliminary version of this paper was presented at the 4th International Conference on Austroasiatic Linguistics, held in Bangkok in October 2009. That presentation covered the broader topic of early Austroasiatic substrates in island SE Asia, but for reasons of length, only the issue of Shom Pen is covered here. I would like to acknowledge the input of Laurie Reid, who kindly made the original of Chattopadhyay & Mukhopadhyay (2003) available, the suggestions for Austroasiatic cognates by Gerard Diffloth and Paul Sidwell, and general discussions on the topic with George van Driem.

cited in Shorto (2006) the reference number is given, though the forms are those from the sources, not from Shorto. The order is alphabetical by English gloss. The divisions of Proto-Mon-Khmer (PMK) into A, B and C are quoted from Shorto. I have also added forms from other Nicobarese languages where it is possible they are related, while noting my scepticism where relevant.

1. ‘back’ Shorto 1844.

Branch	Language	Attestation	Gloss	Comment
Shompen		gikau	back	R. <i>tamnōi</i> . M. <i>hok-ō-a</i> .
	PMK A	*krawʔ	back, behind	
Vietic	proto Vietic	*k-ro:ŋʔ	<i>dos, arête</i> , back, ridge	
Vietic	Mự̄ng	khau	back	
Khmeric	Khmer	knaɑŋ	back (person)	
Bahnaric	Laven	hkuət	back (person)	
Bahnaric	Stieng	kΛΛj	back (person)	
Khmuic	Mal	khəŋ	back side of	
Palaungic	Wa	krawng	the back	
Pearic	Chong	klɑɑŋ lùuŋ	back (person)	
Aslian	Kensiu	kiyɔʔ	back	
Aslian	Temiar	kəruʔ	back	
Munda	Sora	kəndɔŋ	back	

Commentary: Man’s form is undoubtedly the same word with a different prefix. Austroasiatic must clearly be reconstructed with a final velar nasal, but this is lost or weakened in many branches. De Roepstorff’s form might just be an error for ‘rib’ (19.).

2. ‘bat’ Shorto 237.

Phylum	Branch	Language	Gloss	Attestation	Comment
		Shompen	okināu / kāināi / oiāy	bat	
Austroasiatic	Nicobarese	Car	ʔel-ki-ləm	small bat	??
		PMK [A]	*klwaʔ	bat	
	Vietic	proto Vietic	*k-re:ŋʔ	<i>chauve-souris</i>	
	Khmeric	Khmer	crɪŋ 𑜀𑜢𑜤𑜰𑜫	large fruit-eating bat	
	Monic	Old Mon	kəlwaʔ	bat	
	Pearic	Chong, Southern	kʰre:n	bat	
	Aslian	Temiar	kΛluaŋ	bat, flying fox	< Malay
Austronesian	Malayic	Malay	kəluaŋ	bat	

Commentary: As with ‘bat’, Shom Pen appears to show a loss of a final velar nasal attested in Vietic and Pearic. However, closely related forms are also attested in Austronesian, so the regional history of this word is still uncertain. The bat in question might well be the fruit-bat or flying fox; but most definitions are too vague to be sure.

3. ‘boat, canoe’ Shorto 336

Branch	Language	Attestation	Gloss	Comment
Shompen		toāy	canoe	M. <i>dō'ai, hō-a</i> .
	PMK [A])	*đuuk	boat, canoe	
Vietic	proto Vietic	*đo:k	bateau	
Khmeric	Surin Khmer	tu:ʔ	boat, ship	
Katuic	Proto-Katuic	*đuuk	boat	
Bahnaric	Brao	đuuk	boat	
Monic	Nyah Kur [Huai Khrai]	thùuk	boat	
Pearic	Chong [Samre]	tòk	boat	

Commentary: Man compares this to Nancowry *due* canoe but this is not the usual Nicobarese term for ‘boat, canoe’ and this may well have a separate cognates in Monic, *rùə* boat (Nyah Kur).

4. Another less likely external cognate is the second Shom Pen term for ‘canoe’.

Branch	Language	Attestation	Gloss	Comment
Shompen		bekuāu	canoe	R. <i>kehā. ?</i>
Vietic	Mon	klɛŋ ɔl ^ᶜ	canoe, small boat	
Khmuic	Khmu [Cuang]	clɔ:ŋ		

Commentary: Austroasiatic word formation theory suggests that the underlying root is *-Vŋ* so this may not be a good match. Shorto (2006: root 747) reconstructed **d₂luuŋ* ‘boat’ to PMK [B] which is likely to be another version of this root.

5. ‘carry on shoulder’ Shorto 548

Branch	Language	Attestation	Gloss
Shompen		netoŋ	carry on shoulder
Nicobarese	Car	to:m	to carry things on (over) one's arm
	PMK	*tuuŋ	to carry on pole between two [supports]
Bahnaric	Bahnar [Pleiku]	tuuŋ	to carry on shoulder
Vietic	Thavung	thâ:n	to carry on the head; to learn
Khmeric	Khmer	tù:l ʘŋ	to carry on head
Katuic	Pacoh	tu:n	carry on head or shoulders
Monic	Nyah Kur [Central]	tɛŋ	carry; mind, take offense
Pearic	Chong [Kompong Som]	tɿ:l taot	to carry on head
Khmuic	Khmu [Cuang]	trjo:ŋ	carry something on pole by placing on shoulders of people
Palaungic	Riang [Lang]	təŋ ¹ jəŋ ¹	to carry on pole by two men
Munda	Santali	kuʈuŋ	to carry on one's back
Munda	Sora	tiŋŋo(R)	to carry on the back

Commentary: Shorto (2006: root 565) connected this with **rt₁uuŋ* ‘bridge’ which is certainly possible, although the roots would have diverged an early phase in Austroasiatic history.

6. ‘charcoal’ Shorto 1892

Branch	Language	Attestation	Gloss
Shompen		shaak M.	charcoal
	PMK C	*kcah	charcoal
Vietic	proto Vietic	*k-caŋ	<i>charbon de bois</i>
Pearic	Chong [Ban Thung Saphan]	k ^h ajah	charcoal
Katuic	proto Katuic	*kcah	charcoal
Bahnaric	Stieng	chah	charcoal
Monic	Mon [Rao]	chəh	charcoal
Khmuic	Thin	sah	charcoal
Palaungic	Lawa	səeh	charcoal
Munda	Gutob	isai	charcoal

7. ‘cut’

Branch	Language	Attestation	Gloss
Shompen		tainhō	cut (into small pieces)
Vietic	Thavung	thalə̌:	to cut off
Khmeric	Surin Khmer	təha:ʔ	to cut into thin pieces using scissors
Pearic	Chong [Samre]	t ^h ɛtɛ	to cut
Bahnaric	Stieng	tah	to cut up
Bahnaric	Laven	təh	cut (hack)
Katuic	proto Katuic	*trɛh	cut, slash
Monic	Nyah Kur [Huai Khrai]	tháaŋ	clear (land), cut away, cut down plants
Khmuic	Khmu [Cuang]	ta:n ^h ŋɔʔ	cut up paddy rice (using a tool)
Palaungic	Palaung	təh	to cut short (grass)
Nicobar	Car	ta:-lə	to be cut open
Khasi	Khasi	t ^h oʔ	to cut trees
Munda	Kharia	~ dɛŋ	to cut

Commentary: Shorto did not identify this root but it looks reasonable to reconstruct it to proto-Austroasiatic.

8. ‘egg’ Shorto 1253

Branch	Language	Attestation	Gloss	Comment
Shompen		kate ab [ka:ʔte:əb] M.	egg	
	PMK	*[k]t,ap	egg	
Vietic	Sedang	kəta	egg, genitals	
Khmeric	Surin Khmer	kta:ʔ	cry of a hen after laying an egg	
Bahnaric	Tampuan	katap	egg	
Monic	Mon	katop	set (eggs)	
Palaungic	proto Waic	*ktəm	egg	? widespread * <i>tuy</i> root
Aslian	Temiar	tab	egg	
Munda	Gutob	utob-siŋ	egg of chicken	

Commentary: **tap* without the k-prefix is extremely widespread in Austroasiatic. A good example of a Shom Pen form absent in Nicobarese.

9. ‘dog’ < ‘bite’ Shorto 1231

Branch	Language	Attestation	Gloss	Comment
Shompen		kāb	dog	R. <i>kūp</i> .
	PMK A	*kap	to bite	
Vietic	proto Pong	*ka:t	<i>mordre</i> , to bite	
Khmeric	Khmer	kat	to bite	
Pearic	Pear of Kompong Thom	khat	bite (dog)	
Bahnaric	Alak and others	kap	to bite (dog)	
Katuic	proto-Katuic	*kap	to bite	
Khmuic	Thin	khɛt	bite (snake)	
Monic	Mon	kit 𑜀𑜢𑜤𑜂𑜫	to bite	
Monic	Nyah Kur [Klang]	káp-káp	sound of dog biting a smaller animal	
Palaungic	Lawa	kiat	to bite	
Palaungic	Riang	kak ²	to bite	
Aslian	Temiar	kab	to bite	
Nicobar	Car	kap	to bite (also passive)	
Khasi	Khasi	kahap	to attempt to bite	
Munda	Korku	kap	to bite with the incisors	

Commentary: The Shom Pen word for ‘dog’ has no direct parallel, but closely resembles the word for ‘to bite’ found virtually throughout Austroasiatic. Not elsewhere in Austroasiatic as ‘dog’ but throughout as *kap/kat for ‘to bite’ and twice associated specifically with dogs.

10. ‘eight’

Branch	Language	Attestation	Comment
Shompen		thugāy	R. <i>taw-we</i> .
Katuic	Pacoh	ti.kɔ:l	
Katuic	Kui	təkùal	
Katuic	Souei	tək'ɔl	
Khmuic	Mlabri	tii?	
Palaungic	Palaung	ta	
Munda	Sora	tamji	

Commentary: Shorto did not identify this root but it evidently has a significant time-depth within Austroasiatic. Not otherwise in Nicobarese. The Munda forms are not the most common root and the *-mji* would have to be a later suffix. However, since the *-kVl* element is apparently confined to Katuic, this is entirely possible.

11. ‘eye’ Shorto 1045

Phylum	Branch	Language	Attestation	Gloss
	Shompen		meāin	eye
Austroasiatic		PMK	mòt màt mat mət	eye
Austroasiatic	Katuic	Pacoh	mát	eye
Austroasiatic	Bahnaric	Sedang	ma	eye
Austroasiatic	Khmuic	Mal	mat	eye, face
Austroasiatic	Khasian	Khasi	k ^h mat, mat	eye
Austroasiatic	Mangic	Bolyu	lai ⁵³ mat ⁵³	eye
Austroasiatic	Monic	Proto-Mon	*mat	eye
Austroasiatic	Nicobaric	Nancowry	mat	eye
Austroasiatic	Vietic	proto Vietic	*mat	oeil, eye
Austroasiatic	Pearic	proto Pearic	*m(a)t	eye
Austroasiatic	Aslian	Kensiw	met	eye
Austroasiatic	Munda	Kharia	mə'd	eye
Austronesian		PMP	*mata	eye
Hmong-Mien		PMH	*mɤejH	eye
Sino-Tibetan		PTB	*s-may	face
Sino-Tibetan		PTB	*s-mik ~ s-myak	eye
Sino-Tibetan		Karenic	mai-	

Commentary: ‘Eye’ is one of those perplexing regional roots that seems to show up in all the language phyla of the region in much the same form. As far as can be determined, the Shom Pen attestation does not share the final stop characteristic of Austronesian and Austroasiatic and may well be a borrowing from one of the languages with a semi-vowel.

12. ‘foot, leg’ Shorto 538

Phylum	Branch	Language	Attestation	Gloss	Comment
	Shompen		(ko) ceog	foot	R. joek. M. chuk.
Austroasiatic		PMK	*juŋ/juəŋ/jəŋ	foot, leg	
	Vietic	Thavung	ci:ŋ ²	<i>jambe, pied</i>	
	Katuic	Ngeq	j'uŋ	foot	
	Khmeric	Khmer	cəəŋ	foot	
	Pearic	Chong Kanchanaburi	c ^h uŋ	foot	
	Bahnaric	Sedang	ceəŋ	foot, leg	
	Khmuic	Mal	cəŋ	foot	
	Mangic	Bolyu	jəŋ ⁵⁵	foot	
	Monic	Mon	càŋ	foot	
	Palaungic	Wa	caŋ	foot	
	Aslian	Temiar	juk	foot	
	Munda	Santal	jaŋŋa	foot	
	Munda	Sora	jeŋ	foot	
Austronesian	Malayic	Low Malay	cakee	foot, leg	
Austronesian		Lampung	cukut	foot, leg	

Commentary: Not elsewhere in Nicobarese. The Malay forms are interesting, as they are fairly isolated within Austronesian and could well be borrowings from Aslian or whatever now-vanished language Shom Pen was related to on the mainland.

13. ‘full’ Shorto 259

Branch	Language	Attestation	Gloss
Shompen		pei	full
	PMK	*bhiiʔ	sated, full
Bahnaric	Stieng	beej	full
Khmeric	Khmer	pɨj ព័ន្ធ	to be full
Katuic	Katu [Triw]	biij	full
Monic	Mon	pɔij 𑜀𑜢𑜤𑜂𑜫	to be full, complete
Aslian	Semelai	bhih	sated, full from eating
Pearic	Chong [Samre]	pùəŋ	full
Khmuic	Mlabri	biiʔ	full (not hungry)
Palaungic	Danaw	biək ³	full
Khasian	Khasi	pi: ʔe:	full or swollen (as an overfull stomach).

14. ‘get up’ ~ ‘climb, ascend’ ? Shorto 1869

Branch	Language	Attestation	Gloss
Shompen		ikheāu	get up
	PMK	*haaw	to climb, ascend
Bahnaric	Tampuan	haaw	climb
Monic	Mon	kwàn 𑜀𑜢𑜤𑜂𑜫	to climb
Khmuic	Khmu [Cuang]	k ^h iən	climb, ascend
Palaungic	Wa	kao	to get up
Khasian	War [Amwi]	həw	go! get up
Khasian	Khasi	kiew	to go up, climb

Commentary: There may be two parallel roots in Austroasiatic with an initial velar and with a weakened initial. Wa has *kao* ‘get up’ but *hao* ‘climb’. Shom Pen would then be cognate with an older root with k- and the **haaw* root identified by Shorto for ‘climb’ a later development.

15. ‘knee’ Shorto 468

Branch	Language	Attestation
	Shompen	guāg
Nicobarese	Nancowry	kɔj-kanɔŋ
	PMK A	*kʔɔŋ
Vietic	Thavung	ʔɔ:k ʔakú:n
Khmeric	Khmer	ស្រែកច្រាប ជង្គង់
Khmuic	Thin	kiʔ ɨŋjɔŋ
Khmuic	Mal	ɨkjɔŋ
Palaungic	proto Waic	*ɨjɔŋ

Commentary: Attestations of this root are highly scattered in a few branches and are often compounds so it should probably not be reconstructed to proto-Austroasiatic. Shom Pen will only be cognate if the term originally had a final velar nasal which lost the nasal element. It should be noted that words for ‘knee’ show surprising similarities in Old World languages. The table, taken from Blench (2008) shows typical attestations of the *kVN-* formula in the four language phyla of Africa.

Phylum	Family	Language	Attestation
Nilo-Saharan	Kuliak	Ik	kútúŋ
Niger-Congo	Atlantic	Bijogo C	kunu
Niger-Congo	Bantu	Zones E, F etc.	-kónò
Afroasiatic	S Cushitic	Asax	ngulu-et
Khoisan	Southern	!Xóǀ	g xúũ

16. 'leaf stalk' Shorto 506

Branch	Language	Attestation	Gloss
Shompen		koāuŋ	leaf stalk
	PMK	*tkuəŋ; *tkəəŋ	stalk
Bahnaric	Sedang	koŋ	stalk, stem
Katuic	Ngeq [Chatong]	kəəŋ	stalk, stem
Palaungic	Riang [Lang]	kuəŋ ²	stalk

Commentary: Despite only being attested in a few branches of Austroasiatic their wide distribution argues that this must be an old root.

17. 'lizard' Shorto 1446

Branch	Language	Attestation	Gloss
Shompen		kāi	lizard (house)
Nicobarese	Nancowry	huyé	big lizard
	PMK	*[d]kuuy; *[d]kuəy	<i>Calotes</i> lizard
Bahnaric	Halang	kuay	lizard
Pearic	proto Pearic	*kjoŋ	kind of lizard
Katuic	Pacoh	kəaj	lizard (finned)
Monic	Mon	həkui ɜṛṛ	Bloodsucker lizard, <i>Calotes versicolor</i>
Khmuic	Mal	ŋkhə:j	lizard (flying)
Palaungic	proto Waic	*ŋkoj	lizard

Commentary: Lizards are extremely diverse in this region and there seems to be almost no motivation for Shorto's identification of *Calotes spp.* as the initial referent. Indeed the number of vernacular terms for 'lizard' should make us wary when seeking cognates. Nonetheless the similarity of forms in widely distributed languages does suggest that there is an old Austroasiatic root.

18. 'pig'

Phylum	Language	Attestation	Gloss	Comment
	Shom Pen	Shom Pen	(na) lau	<i>R. noŋg</i> wild pig <i>aləv.</i>
Austroasiatic	Monic	Proto-Monic	*cliik	
	Monic	Mon	kloik 𑜀𑜢𑜤𑜂𑜫	
	Monic	Nyah Kur	khlijc	
	Pearic	Pear	sru:k	
	Katuic	Kuy	ʔaaliʔ	
	Katuic	Katu	alic	
	Katuic	Souei	ʔaliʔ	
	Katuic	Pacoh	ʔali:k	
	Katuic	Ngeq	ʔiik	
	Khmuic	Khabit	tə(ə)lek	
	Khmuic	Khang	ʔbu.ʔək ³³	
	Khmuic	Bumang	ək ²⁴	
	Palaungic	Palaung	leʔ	

	Palaungic	Proto-Wa	*lik
	Palaungic	Lawa	ləc
	Palaungic	Praok	lik
Austronesian	PAN		*beRek
	Formosan	Tsou	frəʔə
	Formosan	Puyuma	verek
Sino-Tibetan	Sinitic	OCM	lheʔ, lhaiʔ
	Tani	Galo	rək

Commentary: The feral pig on the Nicobars is *Sus scrofa nicobarica*, a translocated mainland species. As a consequence it is likely to be a culture word but could still be very old, as the translocation of pigs clearly was taking place in the pre-agricultural era. The longer form of this root is attested in Austronesian, and must have been shortened when it was borrowed into Old Chinese and Austroasiatic. Various Austroasiatic branches appear to have subsequently added a variety of prefixes which are not cognate with Austronesian.

19. rib

Phylum	Branch	Language	Attestation	Gloss
	Shompen		tanuā-y	rib
Austroasiatic	Nicobarese	Nancowry	ten-yuāk	rib
	Katuic	Bru	taniar	rib
	Khmeric	Khmer	ktɔəŋ ឧត្ត	beam, cross-beam; fret; rib (of a boat, cart etc.)
	Khmeric	Khmer	trɔnuŋ ត្រនុន	central rib
	Khmuic	Mal	? niauŋ	side (of person), rib
Hmong-Mien		PMH	*tamX	rib

Commentary: Shorto has no reconstruction for ‘rib’. The Hmong-Mien reconstruction is put forward as a possible cognate.

20. ‘unfold, open’ Shorto 702

Branch	Language	Attestation	Gloss
	Shompen	lāuŋ	
Nicobarese	Nancowry	lāŋ, cuŋlāŋi	to unfold
	PMK	*la[a]ŋ	to unfold
Bahnaric	Sre	laŋ	to unfold
Vietic	proto Vietic	*m-laŋ	<i>ouvrir (l'oeil)</i> , open (eyes)
Khmeric	Khmer	tradaaŋ ត្រាដាង	to unfold
Katuic	Katu [Dakkang]	laŋ	open, unlock
Katuic	Ta'oi [Sekong]	laaj	open up, spread out, unfold, unroll
Monic	Mon	ràŋ រ៉ាង	to open, unfold, expand
Khmuic	Khmu [Cuang]	?a:ŋ	open (mouth)
Palaungic	Palaung	?aŋ	to open (the mouth)
Khasian	Khasi	ang	to open, to crack
Munda	Sora	aŋ-təm	to open the mouth

Commentary: There is clearly a highly distinctive ‘Western’ form of this root, which first reduces the initial lateral to a glottal stop and then deletes it entirely. Khuic, Palaungic, Khasi and Muṅḍā all illustrate this process.

21. ‘tooth’ Shorto 1303

Branch	Language	Attestation	Gloss
Shompen		tām	tooth
	PMK	*[t]ʔaam	(molar) tooth
Bahnaric	Bahnar	taa:m	molar
Vietic	Chút [Sách]	ʔaʔa:m ¹	<i>molaire</i> , molar
Katuic	Katu [An Diem]	taʔaam	molar

Commentary: This might well be a misleading surface similarity; the underlying root here is a compound, something like #*tagaam*, where the –g- has either weakened to a glottal, disappeared or devoiced to –k-.

22. ‘tree, wood’ Shorto 254

Branch	Language	Attestation	Gloss
Shompen		hãu	
	PMK A,B,C	*jhooʔ; *jhəəʔ; *jheeʔ	tree, wood
Nicobarese	Nancowry	ci:ə	tree, wood
Khmeric	Old Khmer	jhe	tree, wood
Monic	Mon	chuʔ əʔ	tree, log, wood, timber
Palaungic	Lawa [Umphai]	cheʔ	wood, firewood

Commentary: Shorto’s proposed reconstruction looks rather optimistic based on actual attestations. The Nancowry cognate looks more probable than the Shom Pen form.

23. ‘to vomit’ Shorto 474

Branch	Language	Attestation	Gloss
Shompen		okheāg	to vomit
Nicobarese	Car	ku:-ʔəl	to vomit (tr.)
	PMK	*həək	to vomit
Vietic	Thavung	ʔakò:n	to vomit, to throw down into a hole or on the ground violently
Katuic	Ngeq	kʔətʔaʔ	vomit (v.)
Khmeric	Khmer	kʔuət ក្បឺត	to vomit
Bahnaric	Sedang	kəʔũə	to vomit
Khmuic	Mal	khiaʔ	to spit up, vomit
Monic	Nyah Kur [N.]	kətáaʔ	vomit (of a person)
Aslian	Temiar	kəʔ	to vomit
Aslian	Kensiw	kəʔ	to vomit

Commentary: The final nasal in Vietic may be the relic of a velar nasal, which would have reduced to a velar in Shom Pen.

4. The interpretation of these results

The uncertainties in our knowledge of Shom Pen and the doubtful quality of the transcriptions make it unlikely a definitive conclusion can be reached until better material is available. However, the comparative data set out here makes a reasonable case for lexical cognates with mainland branches of Austroasiatic that are not otherwise attested in Nicobarese. Cognates do not in themselves make for genetic affiliation; either Shom Pen is a branch of Austroasiatic or it came into contact with Austroasiatic languages at some point during the sojourn of its ancestors on the mainland. Indeed, both are possible; there are items in the basic lexicon such body parts and numbers which may point to genetic affiliation, while words such as ‘pig’ are probably culture words.

This suggests that the Shom Pen represent a pre-Nicobarese migration to Great Nicobar, which is in itself not unlikely; foragers reached many islands in the western Andaman Sea in pre-Austronesian times. The ‘Barrier Islands’ west of Sumatra have abundant archaeology attesting to this (e.g. Forestier 2009; Bonatz 2009) and the Andaman Islands must surely have been reached long ago. Regrettably, restrictions on archaeology in the Andamans and Nicobars have made it impossible to establish this with certainty.

References

- Blench, R.M. 2007. The language of the Shom Pen: a language isolate in the Nicobar Islands. *Mother Tongue XII*: 179-202.
- Blench, R.M. 2008. The problem of pan-African roots. In *Hot Pursuit of Language in Prehistory*. J. Bengtson ed. 189-209. Amsterdam: John Benjamins.
- Bonatz, Dominik, John Miksic, J. David Neidel, Mai Lin Tjoa-Bonatz eds. 2009. *From Distant Tales: Archaeology and Ethnohistory in the Highlands of Sumatra*. Newcastle-upon-Tyne: Cambridge Scholars Publishing.
- Chattopadhyay, Subhash Chandra and Asok Kumar Mukhopadhyay 2003. *The Language of the Shompen of Great Nicobar : a preliminary appraisal*. Kolkata: Anthropological Survey of India.
- Das, A.R. 1977. *A study on the Nicobarese language*. Calcutta: Anthropological Survey of India.
- De Roëpstorff, 1875. [2nd ed.] *Vocabulary of dialects spoken in the Nicobar and Andaman islands*. Calcutta: [no publisher].
- Forestier, Hubert et al. eds. 2009. *Mentawai: l’île des homes fleurs*. Paris: IRD.
- Man, E.H. 1886. A Brief Account of the Nicobar Islanders, with Special Reference to the Inland Tribe of Great Nicobar. *The Journal of the Anthropological Institute of Great Britain and Ireland*, 15:428-451.
- Man, E.H. 1889. *A dictionary of the Central Nicobarese language*. London: W.H. Allen.
- Radhakrishnan, R. 1981. *The Nancowry word: phonology, affixal morphology and roots of a Nicobarese language*. Edmonton, Canada: Linguistic Research Inc.
- Rizvi, S.N.H. 1990. *The Shompen*. Calcutta: Seagull Books.
- Shorto, H. 2006. *A Mon-Khmer comparative dictionary*. Paul Sidwell, Doug Cooper and Christian Bauer (eds.) PL 579. Canberra: ANU.
- Whitehead, G. 1925. *Dictionary of the Car-Nicobarese language*. Rangoon: American Baptist Mission Press.