

# AUSTROASIATIC STUDIES

Papers from ICAAL 4

Mon-Khmer Studies Journal  
Special Issue No. 3

Volume editors:

Sophana Srichampa

Paul Sidwell

Kenneth Gregerson



## Pacific Linguistics E-9

Pacific Linguistics is a publisher specialising in grammars and linguistic descriptions, dictionaries and other materials on languages of the Pacific, Taiwan, the Philippines, Indonesia, East Timor, southeast and south Asia, and Australia. Pacific Linguistics, established in 1963 through an initial grant from the Hunter Douglas Fund, is associated with the Research School of Pacific and Asian Studies at The Australian National University. The authors and editors of Pacific Linguistics publications are drawn from a wide range of institutions around the world. Publications are refereed by scholars with relevant expertise, who are usually not members of the editorial board.

---

FOUNDING EDITOR: Stephen A. Wurm

EDITORIAL BOARD: Malcolm Ross and I Wayan Arka (Managing Editors),  
Mark Donohue, Nicholas Evans, David Nash, Andrew Pawley,  
Paul Sidwell, Jane Simpson, and Darrell Tryon

---

Copyright in this edition is vested with the authors

First published 2011

National Library of Australia Cataloguing-in-Publication entry:

Author: International Conference on Austroasiatic Linguistics (4th., 2009 : Salaya, Thailand).  
Title: Austroasiatic studies [electronic resource]: papers from ICAA1 4, Mon-Khmer Studies Journal Special issue no. 3 / edited by Sophana Srichampa, Paul Sidwell and Kenneth Gregerson.  
ISBN: 9780858836426 (Electronic document)  
Subjects: Austroasiatic languages--Congresses.  
Other Authors/ Srichampa, Sophana.  
Contributors: Sidwell, Paul J., Gregerson, Kenneth J.  
SIL International.  
Australian National University. Research School of Pacific and Asian Studies.  
Pacific Linguistics  
Mahāwitthayālai Mahidon.

Typeset by the editors and assistants

---

Published concurrently by:



Pacific Linguistics  
School of Culture, History and Language  
College of Asia and the Pacific  
The Australian National University



Research Institute for Languages and Cultures of Asia  
Mahidol University  
Nakhon Pathom, Thailand



SIL International  
7500 West Camp Wisdom Road  
Dallas, Texas, USA

## Contents

<b>Contents</b> .....	iii
<b>Preface</b> .....	v
Vietnamese Demonstratives <i>đây, đó, kia</i> .....	1
Mayumi ADACHI	
Is Shom Pen a Distinct Branch of Austroasiatic?.....	9
Roger BLENCH and Paul SIDWELL	
Language Vitality and Mon Ethnic Group Attitudes Towards Ethnic Tourism Development in the Western Region of Thailand: A Preliminary Report .....	19
SUJARITLAK Deepadung, SUMITTRA Suraratdecha, NARONG Ardsmiti and PICHET Setaphong	
The North Bahnaric Clade: A Computational Approach.....	33
Jerold A. EDMONDSON, Kenneth GREGERSON and Paul SIDWELL	
Burmese in Mon Syntax – External Influence and Internal Development .....	48
Mathias JENNY	
An Analysis of Generic Structure Potential of Some Selected Austroasiatic Folktales .....	65
PATTAMA Patpong	
<i>Kantrum</i> : the Status of Folk Music Education within the Thai Khmer Cultural Zone .....	81
SAOWAPA Pornsiripongse and ANAN Soplek	
Demonstratives in Muöt (Nancowry) .....	89
V. R. RAJASINGH	
Katuic-Bahnaric: Austroasiatic Sub-Family or Convergence Area? .....	100
Paul SIDWELL	
Vietnamese Slogans in the 21st Century .....	112
SOPHANA Srichampa	
The Vietnamese Expression of BODY and SOUL: A Cognitive and Cultural Linguistic Study .....	127
LY Toan Thang	
Development of Northern Khmer Primer using Thai Alphabet: Opportunities and Challenges .....	134
SIRIPEN Ungsitipoonporn	

**Special Session on Iambic Effects in Austroasiatic**

Sesquisyllabicity in Khmer .....	152
John HAIMAN	
Merge Right: Iambic Effects, Grammaticalization and Cliticization in Brao Grammar .....	162
Charles KELLER, Jacqueline JORDI and Kenneth GREGERSON	
Retention And Reduction in Reduplicants of Semai .....	178
Timothy C. PHILLIPS	
Prosody and typological drift in Austroasiatic and Tibeto-Burman: Against “Sinosphere” and “Indosphere” .....	198
Mark W. POST	
A Case for Clitics in Pacoh .....	230
Richard WATSON	

## Is Shom Pen a Distinct Branch of Austroasiatic?

**Roger Blench**

Kay Williamson Educational Foundation  
(Cambridge)

**Paul Sidwell**

Center for Research in Computational  
Linguistics (Bangkok),  
Australian National University (Canberra)

### 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. One of the authors has argued in print (Blench 2007) that the Shom Pen language may well be an isolate, not a relative of Nicobarese. In this paper<sup>1</sup> we would like to put forward an alternative view that Shom Pen may be a distinct branch of Austroasiatic, but not part of Nicobarese. The bases for this are cognates with Austroasiatic that do not appear to be shared with Nicobarese.

### 2. The Shom Pen Language: previous studies

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). There is also a short ethnography by Rizvi (1990) which includes some scant linguistic data, mainly kin terms.

Although most reference books list Shom Pen as part of the Nicobarese group, evidence for this has been modest. It appears to rest on a number of isoglosses, and the assumption that being spoken in the Nicobars, it is likely to be related to the other languages of the archipelago. Rather tellingly among early sources, 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 (1886) also observes 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.

And similar references to at least two distinct groups are made by Rizvi (1990).

However, since the 1990s a couple of new sources have appeared that have improved our knowledge somewhat, enough that we can begin to offer some analysis and hypotheses worthy of further investigation:

- Chattopadhyay & Mukhopadhyay (2003), is a lexicon of about 750 words of the Shom Pen language, and includes some discussion concerning the affiliation of Shom Pen.

---

Blench, Roger and Paul Sidwell. 2011. “Is Shom Pen a distinct Branch of Austroasiatic?” In Sophana Srichampa and Paul Sidwell (eds.) *Austroasiatic Studies: papers from ICAAL4. Mon-Khmer Studies Journal Special Issue No. 3*. Dallas, SIL International; Salaya, Mahidol University; Canberra, Pacific Linguistics. pp.9-18.  
Copyright vested in the authors.

- Elangaiyan et al. (1995) is a Shom Pen-Hindi primer with some 70 words transcribed in Devanāgarī script, and subsequently retranscribed into Roman script by van Driem (2008).

Van Driem (2008) reports that considering the data in Elangaiyan et al. (1995) and de Roepstorff (1875) he found (with assistance from Gerard Diffloth) out of a total of 222 lexemes, 109 with Nicobarese cognates, 57 with Southern Mon-Khmer<sup>7</sup> cognates, while 7 of the latter group have no apparent Nicobarese etymologies. He concludes that Shom Pen does belong within the Nicobarese group. Van Driem also discussed another source, Chattopadhyay & Mukhopadhyay (2003), but concluded that problems with the data made it too difficult to work with.

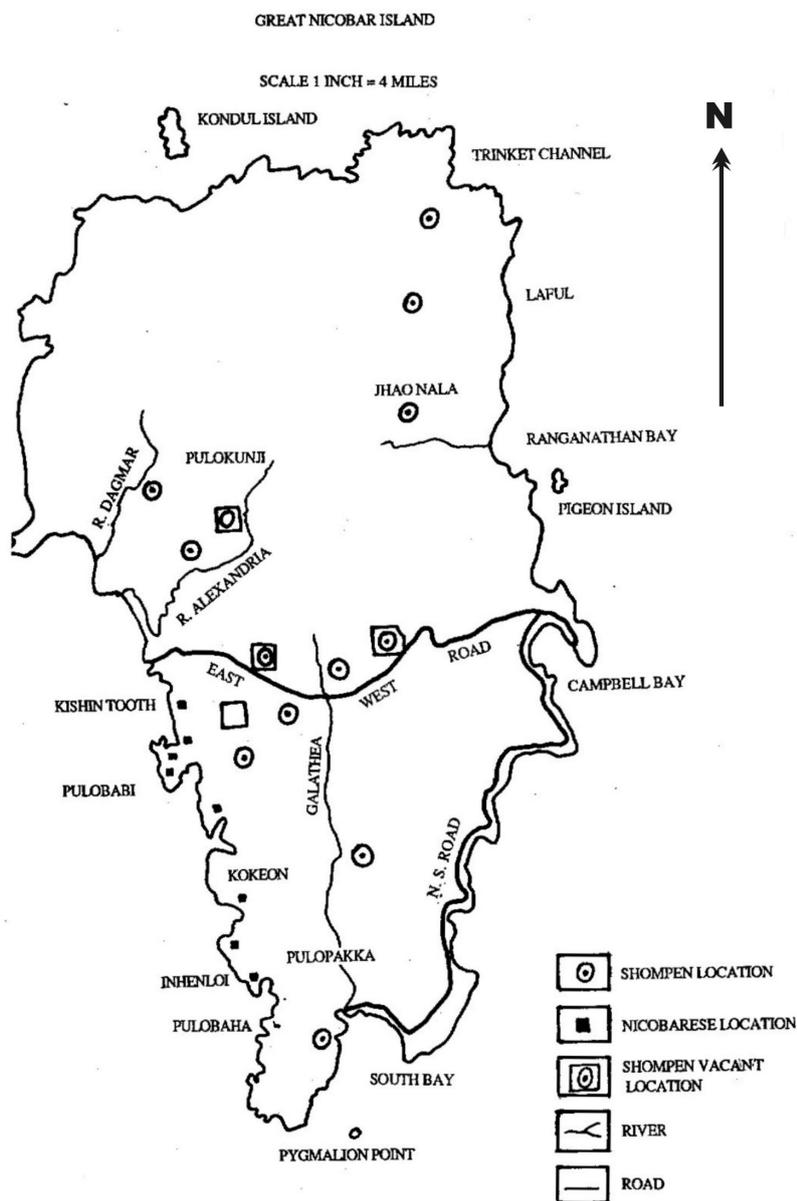


Figure 1: Map showing locations of Shom Pen Communities (Risvi 1990:8).

<sup>7</sup> The term “Southern Mon-Khmer” being Diffloth’s (2005 and *passi.*) category for Monic, Aslian and Nicobarese, which is not necessarily accepted by the present writers.

### 3. The Chattopadhyay & Mukhopadhyay (2003) data interpretation

Chattopadhyay & Mukhopadhyay (2003, henceforth C&M), makes available a significant body of new data on the Shom Pen language, enough to construct a better argument for the affiliation of Shom Pen. Unfortunately the source is not to modern standards of presentation and analysis, and in fact contains various anomalies that suggest that it originates with an earlier - perhaps colonial era - manuscript:

- the authors make no explicit claim of gathering the linguistic data in the field, only that they “collected information” from Shom Pens at the Shom Pen Hut complex
- much of the ethnographic information is lifted directly from Rizvi (1990) without attribution,
- there is no serious attempt at segmentation into words/morphemes, although many entries are transparently phrases or compounds,
- there are serious anomalies in the explanation and use of the phonetic transcription.

The last point is most telling. We are told that there are both short and long low central vowels /a, ā/, but no other central vowels (e.g. no schwa) and no length contrast elsewhere in the vocalism. Then, oddly on p.12 we are told that, “One characteristic feature of Shompen vowel system is the pronunciation of the same vowel in succession without lengthening the vowel.” This is indicated with a double vowel letter spelling, and always appears to occur within the same syllable peak. There is no indication of hiatus between such vowels, and both /h/ and /ʔ/ are discussed elsewhere unproblematically. To us this really looks like a naïve (non-linguistic) reading of a third party manuscript rather than an original piece of work.

Both internal and external evidence suggest that the a-macron is an attempt to notate a schwa or similar weak vowel, whose absence would be typologically/areally rather astounding. First of all, the /ā/ sign has a distinctive distribution:

1. in non-final syllables,
2. in final syllables apparently as on-glide/off-glide of diphthongs,
3. in Malay loan words, and close Nicobarese and Aslian cognates, /ā/ corresponds to either /ə, ɨ, u/ or unstressed /a/.

Examples:

C&M Shom Pen	Malay	Other AA
<i>bāteāu</i> ‘catch (s.t. falling)’	<i>bantu</i>	Jahai <i>bantu?</i> ‘to catch something that is falling’
<i>kādāb</i> ‘salt’	<i>garam</i>	Jahai <i>garəm</i> ‘salt’
<i>wānāṅ</i> ‘thread’	<i>benang</i>	Jahai <i>bneṅ</i> ‘thread’
<i>nāṅ</i> ‘ear’		Nancowry <i>naṅ</i> ‘ear’
<i>ḥeāo</i> ‘to strike’		Car <i>fē:l</i> ‘strike (with hand, fist)’
<i>hāi</i> ‘nasal mucus’		Nancowry <i>hehe:</i> ‘snivel mucus from the nose’
<i>kāi</i> ‘a skin disease’		Temiar <i>maṅkai</i> ‘wound, scar, scabies’
<i>toāy</i> ‘canoe’		Nancowry <i>rue ~ due</i> ‘canoe’
<i>ḥewāi</i> ‘measure of full span’		OldKhmer, Chrau etc. <i>was</i> ‘to measure’
<i>xēāi</i> ‘blood vessel’		Nancowry <i>ihe:</i> , Kuy <i>rsai</i> ‘vein’
<i>poēā</i> ‘lift’		Chrau <i>pə:l?</i> , OldKhmer <i>pək</i> ‘to lift’

Only in some ambiguous cases does the /ā/ possibly correspond to a long /a:/ in AA etymologies. These may still be via a short vowel, and thus would not invalidate our claim, e.g. (note presyllable ā in the first example):

C&M Shom Pen	Other AA
<i>kācām</i> ‘to wait’	Khmer, Kuy etc. <i>cam</i> , but Stieng <i>ca:m</i> ‘to wait’
<i>tām</i> ‘tooth’	Bahnar <i>tə?a:m</i> ‘molar’

The double vowel notation is infrequent, but where it occurs it does appear to indicate long vowels, e.g.:

C&M Shom Pen	Other AA
<i>peii</i> ‘full’	Nancowry <i>paha:e</i> , Jahai <i>bəhi?</i> , PMK <i>*bhi:?</i> ‘sated, full’
<i>kā?eem</i> ‘bone’	Jahai <i>?eη</i> , PMK <i>*c?a:η</i> ‘bone’
<i>mooijob</i> ‘fireplace’	Car <i>muiə</i> ‘fireplace’ (note: Shompen <i>job</i> ‘fire’)

The above observations allow us to make a number of suggestions in terms of the appropriate phonetic interpretation of the C&M data. We can also suggest that in various places the digraph *in* is intended to represent the palatal nasal *ɲ*, and *i* and *u* should be treated as glides, and some other Roman to IPA equivalents. Subsequently below we will suggest a preliminary phonemic representation of forms in curly brackets { } that converts a-macron to schwa, and double vowels with their IPA length mark.

#### 4. Comparison of Chattopadhyay & Mukhopadhyay with Elangaiyan lists.

It is important, first of all, to determine if there is a reliable relationship between the two important recent sources discussed above. Both have notational problems, but is none-the-less clear that in a number of cases the same word is being represented. We have compiled the following 25 comparisons:

C&M Shom Pen	Elangaiyan Shom Pen	Nicobarese cognates
<i>giob</i> { <i>gjob</i> } ‘nail’	<i>ugiyov</i> ‘fingernail’	
<i>juāg</i> { <i>juəg</i> } ‘spider’	<i>jova:k</i> ‘spider’	
<i>kagai</i> { <i>kagaj</i> } ‘child’	<i>ok?ay</i> ‘child’	
<i>kāiāi</i> { <i>kəjəj</i> } ‘parrot’	<i>kayayov</i> ‘parrot’	
<i>kohāt</i> { <i>kohət</i> } ‘girl’	<i>ok?at</i> ‘girl’	
<i>kokugāu</i> { <i>kokugəw</i> } ‘scorpion’	<i>giya:v</i> ‘scorpion’	
<i>komeāin</i> { <i>komeəɲ</i> } ‘forehead’	<i>kuma:n</i> ‘forehead’	
<i>komeoigo</i> { <i>komeoigo</i> } ‘sand’	<i>tyōvgo</i> : ‘beach, sand’	
<i>kotoōi</i> { <i>kotōj</i> } ‘lip’	<i>to:y</i> ‘lip’	
<i>luou</i> { <i>luow</i> } ‘necklace’	<i>lovvu</i> ‘necklace’	
<i>neāu</i> { <i>neəw</i> } ‘cheek’	<i>miyov</i> ‘cheek’	
<i>meāin</i> { <i>meəɲ</i> } ‘eye’	<i>hma:n</i> ‘eye’	
<i>nujāi</i> { <i>nujəj</i> } ‘snake’	<i>hməŋəy</i> ‘snake’	
<i>omeoin</i> { <i>ome?ōj</i> } ‘banana’	<i>mə:ʔoy</i> ‘banana’	
<i>phiŋuāin</i> { <i>phiŋuəɲ</i> } ‘throat’	<i>nuva:n</i> ‘neck’	
<i>uŋāu</i> { <i>uŋəw</i> } ‘bee’	<i>phəŋə</i> ‘bee’	
<i>xiug</i> { <i>xiug</i> } ‘sun’	<i>hiv ~ hi:v</i> ‘sun’	
<i>coāi</i> { <i>coəj</i> } ‘monkey’	<i>cyo:y</i> ‘macaque’	Car <i>?əɲci</i> ‘monkey’
<i>jiāu</i> { <i>jiəw</i> } ‘crab’	<i>ləgə:v</i> ‘crab’	Car <i>?in-ji:w</i> ‘a land crab’
<i>eab</i> { <i>?eap</i> } ‘centipede’	<i>əy?ev</i> ‘centipede’	Nanc. <i>ka?eap</i> ‘centipede’
<i>guiāo</i> { <i>gujəw</i> } ‘coconut’	<i>həya:v</i> ‘coconut’	Nanc. <i>?ojaw</i> ‘coconut’
<i>lau</i> { <i>law</i> } ‘thigh’	<i>ləv</i> ‘thigh’	Nanc. <i>pulə:</i> ‘thigh’
<i>nāŋ</i> { <i>naŋ</i> }	<i>naŋ</i> ‘bamboo’	Nanc. Car <i>naŋ</i> ‘ear’
<i>niāi</i> { <i>niəj</i> } ‘house’	<i>niyo</i> ‘house’	Nanc. <i>ɲi:</i> ‘house’
<i>ŋo</i> { <i>ŋo</i> }	<i>ŋaŋvo</i> ‘bamboo’	Nanc. <i>ŋoa</i> ‘bamboo’

Given the notational issues, and the lack of perfect corresponding glosses, 25 out of 70 is a very respectable score. It also includes several isoglosses that appear to lack Nicobarese cognates, e.g.: ‘sun’,

‘bee’, ‘fingernail’, ‘child’, ‘eye’, ‘snake’. Therefore it seems clear that the same language or two closely related languages are represented in these sources.

### 5. Austroasiatic etymologies.

Blench (2007) sets out the data from C & M in an IPA-like transcription and notes some obvious etymologies. In this paper we go further, and present the results of a fairly exhaustive comparison of the entire C&M corpus with the *Mon-Khmer Comparative Dictionary* of Shorto (2006). In summary, we identify 65 lexemes/morphemes that have reasonable Austroasiatic etymologies that go beyond Nicobarese or even Aslian, but have cognates in other branches of the phylum. These comparisons achieve several goals; 1) the Austroasiatic heritage of Shom Pen is convincingly demonstrated, and 2) we have a baseline for assembling the beginnings of an historical phonology of Shom Pen. Our compilation follows:

C&M Shom Pen	Short (2006) index and reconstruction
<i>tai</i> { <i>taj</i> } ‘hand’	66.A * <i>ti:ʔ</i> ‘hand, arm’
<i>kāināi</i> { <i>kəinəj</i> } ‘bat’	93.A * <i>kn[i]ʔ</i> ‘rat, mouse’
<i>poēā</i> { <i>poəə</i> } ‘lift’	100a.A * <i>pə:ʔ</i> ‘lift’
<i>məd-hēu</i> { <i>məd = hēu</i> } ‘weep’	1045.A * <i>mat</i> ‘eye’
<i>apāo</i> { <i>ʔapəw</i> } ‘wife’	113.A * <i>[ʔ]boʔ</i> ‘mother’
<i>niāi</i> { <i>niāj</i> } ‘hut’	152.A * <i>[j]aa[ʔ]</i> ‘house’
<i>lau</i> { <i>law</i> } ‘thigh’	223.A * <i>blu:ʔ</i> ‘thigh’
<i>laʔōōy</i> { <i>laʔ = ʔōj</i> } ‘yellow leaf’	230.A * <i>slaʔ</i> ‘leaf’
<i>okināu</i> { <i>okinəw</i> } ‘bat’	237.A * <i>klwaʔ</i> ‘bat’
<i>xēāi</i> { <i>xēəj</i> } ‘blood vessel’	249a A * <i>[r]si:ʔ</i> ‘nerve, sinew, vein’
<i>nāφe</i> { <i>nəφe</i> } ‘yawn’	251.A * <i>haʔ</i> ‘to open [mouth]’
<i>hāu</i> { <i>həw</i> } ‘wood’	254.A * <i>jho:ʔ</i> ‘wood’
<i>peii</i> { <i>pej</i> } ‘full’	259 A * <i>bhi:ʔ</i> ‘sated, full’
<i>okheāg</i> { <i>okheəg</i> } ‘vomit’	474 A * <i>hək</i> ‘to vomit’
<i>guāg</i> { <i>guəg</i> } ‘knee’	486.A * <i>kʔəŋ</i> ‘knee’
<i>kāʔeem</i> { <i>kaʔeem</i> } ‘bone’	488 A * <i>cʔaŋ</i> ‘bone’
<i>kogāg</i> { <i>kogəg</i> } ‘branch (of tree)’	496 A * <i>kaŋ</i> ‘transverse, to branch’
<i>koāuŋ</i> { <i>koəuŋ</i> } ‘leaf stalk’	506 A * <i>tkuəŋ</i> ‘stalk’
<i>khoāg</i> { <i>khoəg</i> } ‘boiling’	513.A * <i>guəŋ</i> ‘to cook in water’
<i>keuŋ</i> { <i>keuŋ</i> } ‘pillar’	518 A/B * <i>pgəŋ</i> ; * <i>pgəŋ</i> ‘beam’
<i>gināug</i> { <i>g &lt; in &gt; əug</i> } ‘beam (horizontal)’	518.A/B * <i>pgəŋ</i> ; * <i>pgəŋ</i> ‘beam’
<i>kocuoug</i> { <i>kocuoug</i> } ‘long’	537.A/B * <i>jo[ɔ]ŋ</i> ‘long, high’
( <i>ko</i> ) <i>ceog</i> { <i>ko = cəg</i> } ‘foot’	538(I).D * <i>jə:ŋ</i> ‘foot, leg’
<i>netoəŋ</i> { <i>netoəŋ</i> } ‘carry on shoulder’	548.A * <i>tu:ŋ</i> ‘to carry on pole between two’
<i>kouāu</i> { <i>kouəw</i> } ‘skin, husk’	568 A * <i>cnkuər</i> ‘integument.’
<i>nāidəŋ</i> { <i>nəj = dəŋ</i> } ‘stick’	585.A * <i>tdə:ŋ</i> ‘stalk, tail’
<i>peāg</i> { <i>peəg</i> } ‘cockroach’	630.C * <i>biəŋ</i> ‘spider’
<i>nyeŋ</i> { <i>nyeŋ</i> } ‘horn’	699.A * <i>draŋ</i> ‘horn.’
<i>lāuŋ</i> { <i>ləuŋ</i> } ‘open (bag etc.), unfold’	720.A * <i>la[:]ŋ</i> ‘to unfold’
<i>tameauŋ</i> { <i>tameauŋ</i> } ‘mouth’	911.D * <i>muəŋ[ʔ]</i> ‘mouth’
<i>cuoid</i> { <i>cuoid</i> } ‘heavy’	1146.A * <i>ʔən</i> ‘heavy’
<i>phugāi</i> { <i>phu = gəj</i> } ‘four’	1166.B * <i>puən</i> ‘four’
<i>eab</i> { <i>ʔeap</i> } ‘centipede’	1226.A * <i>kʔip</i> ; * <i>[k]ʔiip</i> ; * <i>kʔiəp etc.</i> ‘centipede’
<i>hekāb</i> { <i>hekəb</i> } ‘bite’	1231.B * <i>ka:p</i> ‘to bite’
<i>kolheəb</i> { <i>kolheəb</i> } ‘ant’	1290.A * <i>klap</i> ‘flying white ant.’

---

<i>toəb</i> { <i>toəb</i> } ‘kiss’	1296.A * <i>thə:p</i> ‘to sniff’
<i>tām</i> { <i>təm</i> } ‘tooth’	1303.A *[ <i>t</i> ]ʔa:m ‘(molar) tooth’
<i>kāyāb</i> { <i>kəjəb</i> } ‘cloud’	1305.A *[ <i>m</i> ]ʔə:m ‘cloud’
<i>toʔŋeāum</i> { <i>toʔ=ŋeəum</i> } ‘cheek’	1318.B * <i>dga:m</i> ‘molar tooth, jaw’
<i>kācām</i> { <i>kə=cəm</i> } ‘stay, rest’	1325.B * <i>ca:m</i> ‘to wait, watch’
<i>kāteob</i> { <i>kəteob</i> } ‘red’	1362.B * <i>dū:m</i> ‘ripe, ripe-coloured’
<i>leəb</i> { <i>leəb</i> } ‘leech’	1410.B * <i>tləəm</i> ‘land leech’
<i>hoəp</i> { <i>hə:p</i> } ‘wash, clean (with water)’	1426.A * <i>hu:m</i> ‘bathe’
<i>tiub</i> / <i>teub</i> { <i>tiub</i> } ‘blood’	1430.A * <i>ʃha:m</i> ‘blood’
<i>koi</i> { <i>koj</i> } ‘head; top’	1443.A * <i>kuj</i> ‘head’
<i>kāi</i> { <i>kəj</i> } ‘lizard (house)’	1446.A/B *[ <i>d</i> ]ku:j; *[ <i>d</i> ]kuəj ‘Calotes lizard’
<i>kākoāy</i> { <i>kə=koəj</i> } ‘sit’	1448.A * <i>ŋguj</i> ‘to sit down’
<i>thaāgge</i> { <i>thaʔəg=ge</i> } ‘thin’	1451.A *[ <i>l</i> ]rgəj ‘thin, lean’
<i>kojāi</i> { <i>kojəj</i> } ‘fly’	1534.A * <i>ruj</i> ‘fly’
<i>tīēu</i> { <i>tīēu</i> } ‘round’	1625.B * <i>dīər</i> ‘to circle round’
<i>hekāu</i> { <i>hekəʔ</i> } ‘break’	1702.A * <i>kal</i> ‘to break off’
<i>kāokāo</i> { <i>kəw=kəw</i> } ‘bark’	1709.B *[ <i>c</i> ]kuəl ‘bark.’
<i>kāo</i> { <i>kəw</i> } ‘belly’	1721.B * <i>sgəʔ</i> / <i>ʃl</i> ‘middle’
<i>deāu</i> { <i>deəw</i> } ‘heel’	1748.D * <i>kdəʔ</i> ‘heel’
<i>pī āi</i> { <i>pī=əj</i> } ‘rain’	180.A * <i>bri:ʔ</i> ‘sky, rain’
<i>oleāu</i> { <i>oleəw</i> } ‘man’	1857.A * <i>klə:w</i> ‘male sexual organs’
<i>ikheāu</i> { <i>ik=heəw</i> } ‘get up’	1869.A * <i>ha:w</i> ‘to climb, ascend’
<i>gikoāi</i> { <i>gi=koəj</i> } ‘wash; ich’	1881.B * <i>kuəs</i> ‘to scrape, scratch, shave’
<i>napīāi</i> { <i>napīəj</i> } ‘sweep (water)’	1916.D * <i>tpis</i> ‘to sweep’
( <i>ko</i> ) <i>ghiāu</i> { <i>ɣiaw</i> } ‘root’	1927.A * <i>ris</i> ‘root’
<i>φewāi</i> { <i>φe=wəj</i> } ‘measure of full span’	1951.B * <i>wəs</i> ‘to measure’
<i>kāhooi</i> { <i>kə=ho:j</i> } ‘wind’	1958.E * <i>haʔ</i> / <i>ʃh</i> ‘to blow’
<i>methou</i> { <i>mɛ=thou</i> } ‘nipple’	1999.A * <i>təh</i> ‘breast, mamma’
<i>mhou</i> { <i>mhow</i> } ‘nose’	2045.A * <i>muh</i> ‘nose’

---

The above list, at 65 items, reflects approximately 10% of the lexicon represented in C&M, given a certain amount of redundancy in the original. Given the various difficulties acknowledged above, this is probably less than is really represented in the data. On the basis of this, we can make a number of observations:

1. Many, although not all, Austroasiatic word final nasals are realized as voiced stops.
2. Prevoalcalic rhotic /r-/ is realized as the glide /j-/.
3. Word final /-h/ is realized as zero.
4. Word final /-r, -l/ are usually realized as glides /-j, -w/.
5. Austroasiatic long vowels frequently become diphthongs.
6. Regular development of voice onset timing is not evident, perhaps reflecting unreliable transcription.

Items a) and b) above may be most significant, these are discussed in the following sections.

## 6. Aslian connection?

The hardening of final nasals to oral stops in Shom Pen, already remarked upon by van Driem (2008), is especially characteristic of Northern Aslian languages, although some degree of pre-stopping of final nasals occurs in much of Aslian. In this context it is significant that we have compiled more than 30

Shom Pen-Aslian isoglosses, lacking apparent Nicobarese cognates, and these comparisons, perhaps 90% of them, only have matches, or have their best matches, with Northern Aslian languages such as Jehai and Kensiu. This is perhaps the most difficult aspect of this question for us to investigate, because we do not have extensive Aslian data for all sub-groups, so there may be some skewing of the data. Yet it is apparent that in our data matching, the etymologies with good South Aslian matches strongly tend to have general Austroasiatic etymologies. Our list of Shom Pen-Aslian isoglosses follows (Aslian items have MKLP indices included):

C&M Shom Pen	Aslian isoglosses
<i>ecoāu</i> { <i>?ecoəw</i> } ‘blind’	Jehai <i>cū?</i> ‘to be blind’ (Bur2005:C:292)
<i>kaiug kao</i> { <i>kaiug kaw</i> } ‘carry’	Jehai <i>gəl</i> ‘to carry on one’s back or shoulder’ (Bur2005:C:354)
<i>gigāb</i> { <i>gi = gəb</i> } ‘waist’	Jehai <i>gel</i> ‘waist’ (Bur2005:C:359)
<i>geāu</i> { <i>geəw</i> } ‘belching’	Jehai <i>gos</i> ‘belch’ (Bur2005:C:390)
<i>okəy</i> { <i>okəj</i> } ‘slow’	Jehai <i>hakij</i> ‘to be slow’ (Bur2005:C:420)
<i>kaiugnhā</i> { <i>kaiug = hā</i> } ‘road’	Jehai <i>har</i> ‘small path, trail’ (Bur2005:C:440)
<i>tāogheu</i> { <i>təw = ghew</i> } ‘take’	Jehai <i>jəw</i> ‘to take’ (Bur2005:C:549)
<i>anikā</i> { <i>?anikə</i> } ‘right (side)’	Jehai <i>loka?</i> ‘right’ (Pha2006:C:475-2)
<i>kaugāi</i> { <i>kaugəj</i> } ‘feed’	Jehai <i>prgej</i> ‘to feed’ (Bur2005:C:1206)
<i>gōāŋ</i> { <i>gōəŋ</i> } ‘frog (non edible)’	Jehai <i>riŋkoŋ</i> ‘frog (type of)’ (Bur2005:C:1275); Kensiu <i>?ikegŋ</i> ‘green frog’ (Bis1994:C:1578)
<i>opugāi</i> { <i>opu = gəj</i> } ‘mosquito’	Jehai <i>?agas</i> ‘mosquito’ (Pha2006:C:395-2)
<i>job</i> { <i>ʃob</i> } ‘fire’	Kensiu <i>cəbm ?əs</i> ‘to start a fire, to ignite a fire’ (Bis1994:C:165)
<i>cuāg</i> { <i>cuəg</i> } ‘stream’	Kensiu <i>ʃənoŋŋ</i> ‘stream, creek’ (Bis1994:C:1447)
<i>kāitohe</i> { <i>kəit = ohe</i> } ‘cut (wood etc.)’	Kensiu <i>kec</i> ‘to cut’ (Pha2006:C:687-1)
<i>ʃi?uŋāi</i> { <i>ʃi?uŋ = əj</i> } ‘Adam’s apple’	Kensiu <i>lahoŋŋ</i> ‘larynx; esophagus’ (Bis1994:C:691)
<i>tauŋhe</i> { <i>tauŋhe</i> } ‘kick’	Kensiu <i>tənaŋŋ</i> ‘to kick’ (Bis1994:C:1323)
<i>hāub</i> { <i>həub</i> } ‘close, shut’	Semai <i>dahup</i> ‘to shut, to close (a door, a mouth, etc.)’ (Mea1987:C:4503-7)
<i>kāu</i> { <i>kəw</i> } ‘bend’	Semai <i>glaʊ</i> ‘crooked, curved, bent’ (Mea1987:C:1051)
<i>kāpuāu</i> { <i>kəpuəw</i> } ‘soft’	Semai <i>ʃabu?</i> ‘soft, pleasant to touch’ (Mea1987:C:3003)
<i>kwāo</i> { <i>kwəw</i> } ‘frog (edible)’	Semai <i>kaiʊ?</i> ‘edible toad, giant frog’ (Mea1987:C:1315)
<i>gigou</i> { <i>gi = gow</i> } ‘pour; strain off’	Semai <i>kəh</i> ‘to pour water into (something)’ (Mea1987:C:4314)
<i>heiei</i> { <i>hejej</i> } ‘smell’	Semelai <i>hajr</i> ‘smell of burning flesh or bones, or postnatal blood’ (Kru2004:C:1031)
<i>giāo</i> { <i>ʃəw</i> } ‘snake’	Semelai <i>tʃə</i> ‘snake (noun)’ (Kru2004:C:1039); Jehai <i>taʃu?</i> ‘snake (generic)’ (Bur2005:C:1471)
<i>ugābeāu</i> { <i>ugə = beəw</i> } ‘child (male)’	Temiar <i>bə</i> ‘be male (adjective)’ (Mea1998:C:230)
<i>gipē</i> { <i>gi = pē</i> } ‘forest’	Temiar <i>be?</i> ‘forest, jungle (noun)’ (Mea1998:C:291)
<i>diāi</i> { <i>diəj</i> } ‘hammer’	Temiar <i>dʌdal</i> ‘to hit, to hammer, to pound’ (Mea1998:C:707)
<i>hēin</i> { <i>hēj</i> } ‘move’	Temiar <i>həit</i> ‘to go, to move along’ (Mea1998:C:1071)
<i>guice</i> { <i>guice</i> } ‘many’	Temiar <i>ʃəje?</i> ‘many, much, abundant’ (Mea1998:C:1118)
<i>dəo</i> { <i>dəw</i> } ‘seat of canoe’	Temiar <i>kende?</i> ‘seat, chair (noun)’ (Mea1998:C:1403)
<i>kāi</i> { <i>kəj</i> } ‘skin disease sp.’	Temiar <i>maŋkai</i> ‘wound, scar, scabies, injury (noun)’ (Mea1998:C:1689)
<i>hāneŋeāg</i> { <i>həne = ŋeəg</i> } ‘slap’	Temiar <i>pa:g/</i> ‘to slap, to spank (transitive verb)’ (Mea1998:C:1956)

## 7. Nicobarese isoglosses

Van Driem and Diffloth also observed that, for the Shom Pen data they closely examined, they could find Nicobarese cognates for about half the items, and this is broadly in line with our findings for the C&M data. This by itself does not mean that Shom Pen is a Nicobarese language, since one could model various scenarios in which an unrelated language become relexified by contact with neighbouring tongues. We can also hypothesise that normal difficulties in elicitation could see many Nicobarese words offered and recorded among Shom Pen lexicon. Below we present 47 examples of apparent Shom Pen-Nicobarese isoglosses lacking wider external cognates (Nicobarese items have MKLP indices included):

C&M Shom Pen	Nicobarese isoglosses
<i>hehen</i> { <i>he</i> = <i>hen</i> } ‘annoyed (to be)’	Car <i>lin-hen</i> ‘to be annoyed with’ (Whi1925:C:3430)
<i>bāḥōāg</i> { <i>bə</i> = <i>ḥōəg</i> } ‘blowing nasal mucus’	Car <i>puŋ</i> ‘slime’ (Whi1925:C:4613)
<i>okād</i> { <i>okəd</i> } ‘blunt’	Car <i>tətkuat</i> ‘blunt’ (Das1977:C:1852)
<i>tha</i> { <i>tha</i> } ‘catch (s.t. falling)’	Car <i>təe</i> ‘catch’ (something thrown)’ (Das1977:C:1614)
<i>heiaghe</i> { <i>hei</i> = <i>aghe</i> } ‘cough’	Car <i>ʔehē</i> ‘cough’ (Das1977:C:2100)
<i>jiāu</i> { <i>jiəw</i> } ‘crab’	Car <i>ʔin-ji:w</i> ‘a big land crab’ (Whi1925:C:2052)
<i>kagau</i> { <i>ka</i> = <i>gaw</i> } ‘cut (wood etc.)’	Car <i>kuəl</i> ‘to chop; cut down’ (Whi1925:C:3106)
<i>cuou</i> { <i>cuow</i> } ‘dark, night’	Car <i>cuəl ha-təm</i> ‘night, midnight’ (Whi1925:C:551)
<i>okheāŋ</i> { <i>ok</i> = <i>heəŋ</i> } ‘deaf’	Car <i>tat haŋ</i> ‘deaf person’ (Das1977:C:1741)
<i>hekkao</i> { <i>hek</i> = <i>kaw</i> } ‘defecate’	Car <i>ʔət kə:</i> ‘defecate’ (Das1977:C:2143)
<i>kokeə</i> { <i>kokew</i> } ‘father’s b., mother’s b.’	Car <i>jəŋ kiko:</i> ‘uncle’ (Das1977:C:476)
<i>mooijob</i> { <i>mo:j</i> = <i>job</i> } ‘fireplace’	Car <i>muwiə</i> ‘fireplace’ (Das1977:C:1150)
( <i>nē</i> ) <i>kugāu</i> { <i>kugəw</i> } ‘front legs (animal)’	Car <i>kəl</i> ‘hand, foreleg of animal’ (Das1977:C:634)
<i>heheu</i> { <i>he</i> = <i>hew</i> } ‘jerk’	Car <i>ho:k</i> ‘to draw; pull’ (Whi1925:C:1690)
<i>tēou</i> { <i>tēow</i> } ‘jerk’	Car <i>tuuk</i> ‘pull, jerk, pull out, weed’ (Das1977:C:1904)
<i>gihou</i> { <i>gihow</i> } ‘land’	Car <i>taku scho</i> ‘land cleared for ploughing’ (Das1977:C:1669)
<i>ije?ā</i> { <i>ʔij</i> = <i>e?ə</i> } ‘little (a)’	Car <i>həj</i> ‘thin’ (of dimension, object, density)’ (Das1977:C:387)
<i>houou</i> { <i>houou</i> } ‘moon’	Car <i>soho</i> ‘full moon’ (Das1977:C:1580)
<i>ḥeāu ghāu</i> { <i>ḥeəw ghəw</i> } ‘run’	Car <i>fəl</i> ‘run’ (Das1977:C:110)
<i>ĩāo</i> { <i>ʔjəw</i> } ‘sneeze’	Car <i>je:sə</i> ‘to sneeze’ (onomatopoeic)’ (Whi1925:C:6630)
<i>ḥeāo</i> { <i>ḥeəw</i> } ‘strike’	Car <i>fəl</i> ‘strike’ (with hand, fist)’ (Das1977:C:130)
<i>gao xeunāo</i> { <i>gao xeu</i> = <i>ŋāo</i> } ‘swallow’	Car <i>ŋə:k</i> ‘to swallow’ (Whi1925:C:4247)
<i>ghuou</i> { <i>g</i> = <i>huow</i> } ‘tear (as of grass)’	Car <i>hə:ʔ</i> ‘to make trips for broom by tearing coconut leaf’ (Whi1925:C:1777)
<i>ije?ā</i> { <i>ʔij</i> = <i>e?ə</i> } ‘thin (as thread)’	Car <i>həj</i> ‘thin’ (of dimension, object, density)’ (Das1977:C:387)
( <i>nē</i> ) <i>tāi</i> { <i>təj</i> } ‘tusk of boar’	Car <i>kanel haeti</i> ‘tusk’ (Das1977:C:566)
<i>muou</i> { <i>muow</i> } ‘collect, pile up’	Car, Nancowry <i>hamu:l</i> ‘collect’ (Das1977:C:255)
<i>haguinaŋ</i> { <i>haguj</i> = <i>naŋ</i> } ‘hole of ear’	Car <i>naŋ</i> ‘ear’ (Das1977:C:1184); Nancowry <i>naŋ</i> ‘ear’ (Man1889:C:3831)
<i>ca</i> { <i>ca</i> } ‘my’	Car <i>cu-ʔəʔ</i> (Whi1925:C:542); Nancowry <i>cua</i> ‘chüa I, my.’ (Man1889:C:361)
<i>nho</i> { <i>n</i> = <i>ho</i> } ‘bark’	Nancowry <i>ʔohə</i> ‘bark of dog.’ (Man1889:C:4034)
<i>puggāi</i> { <i>pug</i> = <i>gəj</i> } ‘carry on back’	Nancowry <i>ʔokaj-haŋa</i> ‘bear, carry, take away.’ (Man1889:C:4047)
<i>thā</i> ( <i>ki tai</i> ) { <i>thə</i> } ‘clap’	Nancowry <i>təʔ</i> ‘to slap, hit’ (Sho2006:C:2010-8)

---

<i>hecāu</i> { <i>he = cəw</i> } ‘climb’	Nancowry <i>cuak-lare</i> ‘ascend a ladder or stairs.’ (Man1889:C:382)
<i>lubiāu</i> { <i>lub = iaw</i> } ‘coconut kernel’	Nancowry <i>cīa-?ojaw</i> ‘coconut tree’ (Man1889:C:270)
<i>jāg</i> { <i>ʃəg</i> } ‘coconut kernel’	Nancowry <i>juan-?ojaw</i> ‘coconut-fruit.’ (Man1889:C:5941)
<i>kaiug</i> { <i>kajug</i> } ‘go’	Nancowry <i>kajiŋa</i> ‘go away, leave, depart.’ (Man1889:C:2420)
<i>juou</i> { <i>juow</i> } ‘hair of head’	Nancowry <i>jok</i> ‘hair’ (Man1889:C:5866)
<i>xiug</i> { <i>xiug</i> } ‘sun’	Nancowry <i>heŋ</i> ‘sun, day.’ (Man1889:C:1683)
<i>ēcag</i> { <i>ēcag</i> } ‘lame’	Nancowry <i>?oŋ-cuaŋ</i> ‘cripple’ (noun).’ (Man1889:C:4261)
<i>puggāi</i> { <i>pug = gəj</i> } ‘marriage’	Nancowry <i>?okaj-haŋa</i> ‘bear, carry, take away.’ (Man1889:C:4047)
<i>eguφeg</i> { <i>?egu = φeg</i> } ‘mouth’	Nancowry <i>?oal-faŋ</i> ‘mouth.’ (Man1889:C:3968)
<i>hāi</i> { <i>hāj</i> } ‘nasal mucus’	Nancowry <i>hehe</i> ‘snivel mucus from the nose’ (Man1889:C:1625)
<i>oŋān</i> { <i>?oŋā</i> } ‘roast’	Nancowry <i>ŋəf-ŋa</i> ‘scald’ (verb), burn, singe.’ (Man1889:C:3902)
<i>eiōu</i> { <i>?ei = ?ōw</i> } ‘sky’	Nancowry <i>cu-?oal</i> ‘clear sky.’ (Man1889:C:459)
( <i>nā</i> ) <i>ŋoāin</i> { <i>ŋoāj</i> } ‘stone’	Nancowry <i>maŋ-ŋe</i> ‘stone, rock, rock, sandstone.’ (Man1889:C:3582)
<i>xiug</i> { <i>xiug</i> } ‘sun’	Nancowry <i>sup-heŋ</i> ‘sunset.’ (Man1889:C:5111)
<i>helein</i> { <i>helein</i> } ‘wrestling’	Nancowry <i>halina</i> ‘wrestle in anger or in earnest.’ (Man1889:C:1190)

---

It is difficult to make strong claims based on the above data, since at this preliminary stage it is not clear to what extent loans, cognates, and chance resemblances are contributing. But at least it is clear that some loans are represented. A clear case in point is the word *naŋ* ‘ear’, which is reflected without change in all three languages. It is especially telling that *naŋ* preserves the final nasal, rather than the velar stop /-g/ which is expected by regular correspondence. Other examples that look like loans include ‘deaf’, ‘wrestling’. On the other hand, a form such as. Shom Pen *xiug* {*xiug*} ‘sun’ does appear to correspond to Nancowry *heŋ* ‘sun, day’, with regular hardening of the final. Others in this class include: ‘mucus’, ‘coconut kernel’, ‘lame’, ‘go’. Multiple examples that seem to indicate a regular loss of etymological final oral stops in Shom Pen, examples include: ‘jerk’, ‘swallow’, ‘tear (as of grass)’, ‘clap’, ‘climb’, ‘hair’.

## 8. Conclusion

The data above present various ambiguities that make interpretation difficult. However, we can make a number of observations:

- Shom Pen shows numerous lexical items common to Nicobarese and Aslian languages (jointly and severally).
- Comparative analysis indicates a chain shift in the history of Shom Pen in which final stops were lost, and then reintroduced to the language by hardening of final nasals.
- Superficially final nasal hardening resembles Aslian, but in Aslian languages the result was a merger of final nasals and stops, which is structurally different.
- There are Nicobarese loans in Shom Pen, but these do not appear to be enough to explain the extent of common vocabulary.

Consequently, our answer to the question “Is Shom Pen a distinct branch of Austroasiatic?” posed by Blench at the ICAAL4 meeting, is, maybe. On the present evidence Shom Pen looks to some extent like to be an

innovative Nicobarese language, but that on its own would not account for the extent of apparent Aslian cognates, especially North and Central Aslian cognates. A reasonable suggestion is that Shom Pen, Nicobarese, and Aslian form three branches of a Southern Mon-Khmer family. If such a hypothesis is adopted, then Shom Pen is a distinct - although not coordinate - branch of Austroasiatic.

### References

- Bishop, Nancy M. & Mary M. Peterson. 1994. *Kensiw rhyming list*. Bangkok: Thammasat University.
- Blench, R.M. 2007. The language of the Shom Pen: a language isolate in the Nicobar Islands. *Mother Tongue XII*: 179-202.
- Burenhult, N. 2005. *A grammar of Jahai*. Pacific linguistics, 566. Canberra: Pacific Linguistics. <http://en.wikipedia.org/wiki/Special:BookSources/0858835541>
- 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. [2<sup>nd</sup> ed.] *Vocabulary of dialects spoken in the Nicobar and Andaman islands*. Calcutta: [no publisher].
- Elangaiyan, Rathinasabapathy, et al. 1995. *Shompen-Hindi Bilingual Primer Śompen Bhāratī 1*. Port Blair and Mysore: Tribal Welfare Department of the Andaman and Nicobar Administration, Ādivāsī Kalyāṇa Vibhāg Aṇḍamān tathā Nikobār Praśāsan, and Central Institute of Indian Languages [23 unnumbered and 192 numbered pages].
- Kruspe, Nicole D. 2004. *A Grammar of Semelai*. Cambridge: Cambridge University Press (Cambridge Grammatical Descriptions).
- 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.
- Means, N., and P. Means (1986). *Sengoi-English, English-Sengoi dictionary*. University of Toronto and York University: The Joint Centre on Modern East Asia.
- Means, Nathalie, Gordon P. Means; Abuk Alang 1998. *Temiar-English, English-Temiar dictionary*. Saint Paul, Minn. : Hamline University Press
- 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.
- van Driem, George. 2008. The Shompen of Great Nicobar Island: New linguistic and genetic data, and the Austroasiatic homeland revisited. *Mother Tongue* 13:227-247.
- Whitehead, G. 1925. *Dictionary of the Car-Nicobarese language*. Rangoon: American Baptist Mission Press.