

A note on the reconstruction of Proto West Bahnaric and investigation of early West Bahnaric-Katuic contact

Paul SIDWELL¹

Australian National University (Canberra)

Summary

The present paper reports on the result of research in progress. In 2000 I (with Pascale Jacq) presented a reconstruction of Proto West Bahnaric phonology and lexicon². That reconstruction is now under review, within the context of comprehensive reconstructions of Proto Bahnaric and Proto Katuic that I am now preparing. This review indicates that the proto-phonology we proposed in 2000 needs to be revised in some respects, particularly as some phonological mergers cannot be reconstructed without external evidence. The improved historical phonology should also facilitate a more thorough investigation of the historical effects of language contact on West Bahnaric, particularly in respect of ancient Katuic and Khmer influence.

1. West Bahnaric

1.1 Classification, geography

West Bahnaric (WB) is a genetic sub-grouping of Bahnaric languages, spoken mostly in the Attapeu, Champasak and Sekong Provinces of the Lao PDR and the Stung Treng and Ratanakiri Provinces of Cambodia. There are approximately 100,000 West Bahnaric people in total (according to 1995 census figures)—about half of whom live on the Boloven Plateau and mainly grow coffee, while the rest are lowland rice farmers who live east and south of the Plateau. WB is conventionally divided into about a dozen ethno-linguistic groups, and these can also be classified into dialect groups on linguistic grounds:

¹This paper was written in the course of my Postdoctoral Fellowship at the Australian National University. Funding for this fellowship was provided by the Australian Research Council, to whom I express my deepest gratitude.

²Jacq and I are now collaborating to produce a new monograph length treatment of the history of West Bahnaric which will supersede Jacq and Sidwell 2000. The comparative lexicon will be expanded, and the revised phonology will be presented. I am concentrating on the phonology and contact history, discussed briefly in this paper, while Jacq is researching morphology and grammar (for more details of our research see www.anu.edu.au/~u9907217). Thanks are also due to Jacq for her comments on the previous drafts of this paper.

Language	Location (province)
<i>Lavi</i>	(Sekong)
<i>Jru' (Loven), Juk, Su'</i>	(Champasak)
<i>Nhaheun</i>	(Champasak)
<i>Sapuan</i>	(Attapeu)
<i>Cheng, Oi</i>	(Attapeu)
<i>Laveh, Brao, Kavet, Krung,</i>	(Attapeu, Stung Treng, Ratanakiri)

Lexicostatistics (Jacq & Sidwell 2000:4-7) reveals that all these languages share no less than 76% of basic vocabulary, and experience in the field indicates a high degree of mutual intelligibility between the languages. The whole sub-group can be treated as having a rather simple internal genetic structure: historical phonology indicates that WB divides into two branches, one consisting of Lavi, and the other consisting of the rest of the group. The initial division is most strongly indicated by changes among the low vowels and diphthongs, and must belong to the earliest independent stage of WB. Subsequently there were various phonological changes, some involving quite drastic modifications to word structure (such as monosyllabification in Jru' and Nhaheun) but these can be characterised as language specific and therefore have no particular significant consequences for the internal genetic classification.

1.2 Historical phonology

Without modifying our 2000 proposal, the following set of PWB initial consonants is reconstructed:

voiceless stops	*p	*t	*c	*k	*ʔ
voiced stops	*b	*d	*j	*g	
plain nasals	*m	*n	*ɲ	*ŋ	
glottalised nasals	*ʔm	*ʔn	*ʔɲ	*ʔŋ	
resonants	*w	*l *r	*j		
glottalised resonants	*ʔw	*ʔl *ʔr *ʔj			
fricatives			*s		*h

These initials are reflected unchanged in Jru' and Nhaheun, while glottalisation is generally lost in all other WB languages. This is an areally typical lenition (particularly for the languages in close contact with Lao) and has no consequences for sub-grouping. Initial consonant clusters do show some interesting and occasionally complicated lenitions and assimilations, particularly Nhaheun (e.g. *tp- > pp-, *dp- > dw-, *tm- > mm-, *dm- > nw-), although again these are generally language specific and must have occurred after the separation of Lavi from PWB.

However, it is apparent that we did make an error in relation to the reconstruction of PWB initial clusters—specifically we failed to distinguish, for each etymology, the phonemic status of vowels which may intrude into initial consonant sequences creating phonetic minor syllables. At the time of first preparing the reconstruction we applied a model of word structure that assumed that all initial clusters have an underlying minor syllable vowel. Subsequently

careful phonological analysis (e.g. Jacq 2002 in respect of Jru') shows that minor syllable vowels are epenthetic between stops (oral and/or nasal) and only potentially phonemic before resonants *r, l, j, w, h*. This has two consequences:

1. Where we reconstructed a minor syllable vowel between two stops, e.g. **təpat* 'to extinguish a fire', that vowel is notationally redundant and should be dropped or read as epenthetic.
2. Where we reconstructed a minor syllable vowel before *r, l, j, w, h*, each etymology needs to be examined to determine whether that vowel is phonemic, e.g. **təlwəŋ* 'shin' stands, but **pəliiŋ* 'sky' should be **pliŋ*. Such a revised list will appear (see f.n. 2).

The PWB finals are reconstructed as follows:

voiceless stops	<i>*p</i>	<i>*t</i>	<i>*c</i>	<i>*k</i>	<i>*ʔ</i>
plain nasals	<i>*m</i>	<i>*n</i>	<i>*ɲ</i>	<i>*ŋ</i>	
resonants	<i>*w</i>	<i>*l</i> <i>*r</i>	<i>*j</i>		
fricatives			<i>*s[ç]</i>		<i>*h</i>

The finals are completely unproblematic as they have undergone few changes.

For PWB main syllables we reconstructed the following set of vowels, without any registers or tones:

<i>*ii</i>	<i>*īī</i>	<i>*uu</i>	<i>*i</i>	<i>*ī</i>	<i>*u</i>
<i>*ee</i>	<i>*əə</i>	<i>*oo</i>	<i>*e</i>	<i>*ə</i>	<i>*o</i>
<i>*εε</i>	<i>*aa</i>	<i>*ɔɔ</i>	<i>*ε</i>	<i>*a</i>	<i>*ɔ</i>

Note that we did not reconstruct any phonemically diphthonged vowels in the PWB system, a remarkable result given that diphthongs are common in Bahnaric languages, and WB languages especially. The 'high' diphthongs *ie, iə, uo* which are found in at least Jru', Nhaheun, Sapuan and Oi are clearly diphthonged variants of *ii, īī, uu*. On the other hand the 'low' diphthongs *ia, ua* are common to all WB languages except Lavi, which has corresponding *ɔɔ, εε*. By a strict application of the comparative method we reconstructed **ɔɔ, *εε* underlying all cases of WB *ua, ia*. The correspondences supporting this reconstruction are tabled as follows:

Lawi	Lov.	Nha.	Sap.	Oi	Chg.	Lvh	Bra.	PWB	environment
<i>ɔɔ</i>	<i>ɔɔ</i>	<i>ɔɔ</i>	<i>ɔɔ</i>	<i>ɔɔ</i>	<i>ɔɔ</i>	<i>ɔɔ</i>	<i>aa</i>	<i>*ɔɔ</i>	/_C[labial, dorsal., Ø]
<i>ɔɔ</i>	<i>ua</i>	<i>ua</i>	<i>ua</i>	<i>ua</i>	<i>ua</i>	<i>ua</i>	<i>ua</i>		/_C[apical, laminal]
<i>εε</i>	<i>ee</i>	<i>εε</i>	<i>εε</i>	<i>εε</i>	<i>εε</i>	<i>εε</i>	<i>ee</i>	<i>*εε</i>	/_Ø
<i>εε</i>	<i>ia</i>	<i>ia</i>	<i>ia</i>	<i>ia</i>	<i>ia</i>	<i>ia</i>	<i>ia</i>		/_elsewhere

Without considering external comparisons (i.e. trying to avoid 'top-down' reconstruction) the logic of the comparative method demands that we group phonetically similar correspondences which are in complementary distribution, ascribing them to the same proto-phoneme. In the present case there were two apparent choices, 1) to assume that there were mergers in Lavi

Sidwell, P. 2003, "A note on the reconstruction of proto West Bahnaric and Investigation of early West Bhanaric-Katuic contact", in *The Mon-Khmer Studies Journal*, vol. 33, pp. 159-166. (purl.org/sealang/sidwell2003note.pdf)

that eliminated diphthongs completely in that language (and perhaps some limited mergers in the rest of the family that created the pattern of complementary distribution evident in the correspondences), or 2) to assume that *ɔɔ and *ɛɛ are original, and diphthongised to *ua* and *ia* after the separation from Lavi. The problem boils down to determining the direction of sound change, and given that the internal data does not help us, external comparisons were sought to solve the problem.

Various comparisons with Bahnar and South Bahnaric seemed to unambiguously indicate that *ɔɔ was the source of many (if not all) of the examples of *ua* before apicals and laminals in WB, while Lavi is conservative (in this respect) e.g.

	Lawi	Jru'	Bahnar	Stieng (SB)	PB
'child'	<i>kɔɔn</i>	<i>kuan</i>	<i>kɔɔn</i>	<i>kɔɔn</i>	* <i>kɔɔn</i>
'hungry'	<i>pəŋɔɔt</i>	<i>pəŋuat</i>	<i>pəŋɔɔt</i>	<i>pəŋɔɔt</i>	* <i>pəŋɔɔt</i>
'calf/shin'	<i>pɔɔç</i>	<i>puaç</i>	<i>pɔç</i>	<i>pɔɔç</i>	* <i>pɔɔç</i>
'carry on shoulder'	<i>kɔɔj</i>	<i>kuaj</i>	<i>kɔɔj</i>	—	* <i>kɔɔj</i>

In the face of such evidence we felt compelled to reconstruct *ɔɔ underlying all cases of WB *ua*, and assuming a systemic analogy, reconstructed *ɛɛ underlying all cases of WB *ia*. However, subsequent investigations reveal that we must reconstruct a phonemic contrast between *ɔɔ and **ua* in the environment of final apicals and laminals in Proto Bahnaric, e.g.

	Lavi	Jru'	Bahnar	Stieng (SB)	PB
'four'	<i>pɔɔn</i>	<i>puan</i>	<i>pwan</i>	<i>puon</i>	* <i>puan</i>
'to bark'	<i>kaaw</i>	<i>kual</i>	<i>kwal</i>	<i>kuol</i>	* <i>kual</i>
'to buy'	<i>rɔɔt</i>	<i>ruat</i>	<i>rɔt</i>	<i>ruot</i>	* <i>ruat</i>
'elephant'	<i>rɔɔç</i>	<i>ruaç</i>	<i>rɔɔç</i>	<i>ruoç</i>	* <i>ruaç</i>
'gibbon'	-	<i>kuən</i>	<i>kwən</i>	<i>kuən</i>	* <i>kuən</i>

This data indicates that some examples of WB *ua* reflect PB **ua*, and the principles of economy and phonetic plausibility strongly suggest that these reflect an unchanged continuation of PB **ua* rather than a sequence of changes **ua* > *ɔɔ > *ua*. On the other hand it could be argued that such a sequence may be a legitimate *phonological* representation of the conditioned neutralisations that may have occurred in order to produce the above correspondences. In any case the sequence of changes that we must reconstruct in the transition from PB to PWB and WB are set out as follows:

PWB	→	Lavi
		<i>ua, ɔɔ</i> > <i>ɔɔ</i> /all places
		<i>ia, ɛɛ</i> > <i>ɛɛ</i> /all places
* <i>ua</i> / <u> </u> C[apical., laminal]		WB (-Lavi)
*ɔɔ / <u> </u> C[all places, Ø]		<i>ɔɔ</i> > <i>ua</i> / <u> </u> C[apical., dorsal]
		<i>ɔɔ</i> > <i>ɔɔ</i> /elsewhere
		<i>ua</i> > <i>ua</i> /all places

All cases of PB/PWB **ua* have simply merged to ɔɔ in Lavi, and this has parallels in at least Alak (Central Bahnaric) and Rengao (North Bahnaric) which have also merged original diphthongs with monophthongs (e.g. Alak *poon* ‘four’, Rengao *puun* ‘four’ etc.) quite independently. In WB(-Lavi) there are two different sources of *ua*, namely *ɔɔ and *ua, but no way that WB correspondences alone can determine which should be reconstructed at the PWB level for a given etymon without resorting to ‘top-down’ reconstruction to recover the PWB vocalism.

The situation in respect of *ia* is perhaps simpler. The distribution of *ia* in WB(-Lavi) languages is paralleled across other Bahnaric sub-groups, namely that reflexes of PB **ia* are well distributed over all types of closed rimes, while there are few if any well distributed Bahnaric etyma with *εε* in closed rimes. This may be interpreted as indicating that any relationship between *ia* and *εε* relates to the Proto Bahnaric or even pre-Proto Bahnaric level, or (perhaps less likely) there may be no particular relation between these sounds, but like *ŋ* and *h* in English, they are simply in complementary distribution. The upshot of this is that we should probably treat all cases of WB *ia* as reflecting PWB **ia*, with a merger to *εε* in Lavi.

Thus the PWB vowel inventory should stand as follows:

	Long			Short		Diphthonged	
*ii	*ii	*uu	*i	*i	*u	*ia	*ua
*ee	*əə	*oo	*e	*ə	*o		
*εε	*aa	*ɔɔ	*ε	*a	*ɔ		

2. Analysis of Borrowings into Proto West Bahnaric

Now that we have a model of WB historical phonology in which we have greater confidence, the issue of examining ancient borrowings into PWB can be investigated. This is important because it is clear that WB languages contain many borrowings, particularly from Katuic and Khmer, whereas other Bahnaric sub-groups, such as North and Central Bahnaric (the latter including South Bahnaric in my formulation (Sidwell 2002)) have been more significantly influenced by Chamic and later Vietnamese. This has many potentially significant consequences, not the least of which is the possibility of etymologising words based upon their geographical distribution, allowing us to make inferences about ancient language contact. In this paper I will not attempt to survey the broad extent of borrowings into WB, instead I will examine some apparently very ancient borrowings, and comment on their significance in the light of the improved model of historical phonology.

Such an investigation is also predicated upon having a sufficiently well developed understanding of the history of Khmer and Katuic. The former is not a problem because there are Old Khmer dictionaries available (e.g. Pou 1992), and the history of the writing system is well researched so that we have a reasonable understanding of the phonology (Jacob 1976, Pinnow 1980, Ferlus 1992 and others) and the morphological system (Jenner & Pou 1980-81). Katuic is more of a problem, as we cannot yet say that there exists a complete

Sidwell, P. 2003, "A note on the reconstruction of proto West Bahnaric and Investigation of early West Bhanaric-Katuic contact", in *The Mon-Khmer Studies Journal*, vol. 33, pp. 159-166. (purl.org/sealang/sidwell2003note.pdf)

reconstruction based upon an extensive and reliable consolidated comparative lexicon. Two extensive comparative lexicons are available (Peiros 1996, and Theraphan 2001), each of which is accompanied by a reconstruction, but it is my assessment that neither reconstruction is reliable, and I have taken it upon myself to address this problem. My present view is that the reconstruction proposed by Diffloth (1982) is basically correct, although it requires both elaboration and fine-tuning. The problem is that in that paper Diffloth effectively only dealt with the issue of long vowels, and based the analysis on 139 comparisons. I am now in the process of consolidating the comparisons of Peiros and Theraphan, and analysing them in terms of a modified Diffloth scheme, so that I have an improved provisional reconstruction of the Proto Katuic lexicon (PK).

Thomas (1979) presented a classification of Bahnaric languages based upon so-called ‘distinctive vocabulary’—etyma that appear to be restricted to, and thus indicative of, particular sub-groupings. While Thomas’ method is not strictly etymological, and I have been rather critical of it as a method of classification (Sidwell 2002), application of etymological analysis to distinctive vocabulary can produce some interesting results. Among the WB distinctive vocabulary presented by Thomas (and also discussed by Theraphan 1997) are the following which have good isoglosses with Katuic (reconstructions are my own):

PWB	PKatuic	
*ptɔɔr	*bntɔɔr	‘star’
*dkɔɔj	*tkɔɔj	‘horn’
*sɔɔj	*sɔɔj	‘tail’
*səəŋ	*səəŋ	‘five’
*kmɔɔ	*kmɔɔ	‘year’
*piir	*piar	‘flower’

The first three of these have the problematic *ɔɔ vowel in closed rimes. As they lack cognates elsewhere in Bahnaric the internal data does not allow us to decide whether the proto-vowel was *ɔɔ or *ua, but the Katuic comparisons are clearly indicative of *ɔɔ. The close agreement among the other segments strongly suggests borrowing directly from PK into PWB. This is also supported by the fact that in contemporary Katuic languages located close to the WB area these words have undergone various sound changes that indicate that recent borrowing is unlikely, e.g.: In the word for ‘star’ the initial labial has become a nasal, e.g. Bru. *məntɔɔr*, Sô. *mətɔɔr*, while in WB the *n* which conditioned this nasalisation has been lost, which is a normal regular development in PWB, leaving the labial stop intact. Also all WB languages except for Lavi have *ua* as the mainsyllable vowel, which presumably diphthongised as part of the conditioned change that occurred after the split from Lavi, while all contemporary Katuic languages have a monophthong in these etyma. This latter point also applies to ‘horn’ and ‘tail’.

Attempts to further etymologise these words reveal their likely sources. Jenner & Pou (1980-81) give a Khmer word *bantóor* ‘to discharge, emit; lightning bolt’ which is a possible source for the ‘star’ word in WB and

Katuic. The phonological match is excellent, and the semantic link between 'star' and the notion of emitting light is plausible. They derive it from a root *tóor* meaning 'to lean heavily, tip, tilt, list' which by extension comes to mean 'emit'. One may hazard that such a development is unlikely to occur independently, and thus Khmer may be considered the most likely ultimate source for this 'star' word, perhaps mediated via Katuic into PWB. The word for 'horn' also has a Khmer parallel—Jenner & Pou (1980-81) list Khmer *koj* 'straight, upright; rhinoceros horn' and a derived form *ckoj* 'to grow straight up or out'. However, the same word is attested in Vietic and Ferlus (1991) reconstructs Proto Viet Muong **tkɔj* 'horn', so it appears to have a good East Mon-Khmer etymology. The rest of Bahnaric has reflexes of another root, **ʔəkɛ* 'horn', which is also attested in Katuic. On balance I suggest that PWB **tkɔj* 'horn' was borrowed from Katuic into PWB. The 'tail' word appears to be restricted to Katuic—Peiros (1996:66) attempts to link it to Khmer *kansaj* 'back', but Jenner & Pou show that this is derived from a root *-saj* 'line, thread', and this is more properly compared to PWB **kɛɛ* 'string, rope' which has an excellent MK etymology. Ferlus (1991) reconstructs PVM **tɔj* 'tail', but this may be better compared to Khmer *kantuj* 'tail' < *-tuj* 'to jut, protrude', so Katuic/WB **sɔj* 'tail' appears to be isolated, and borrowing from Katuic into PWB is most likely.

The root **səəŋ* 'five' is also restricted to Katuic and WB, at least in the immediate geographical area—Mon has *masoon* 'five' and Diffloth (1976) reconstructs a Proto South Aslian **sɔŋ* 'five', but there is really no more than superficial similarity suggesting any connection between these and the Katuic/WB etymon. Other Eastern Mon-Khmer languages suggest a different root for 'five', cf. Khmer *pram*, Bahnar *pdam*, Vietnamese *năm*, etc. Thus I suggest that PWB borrowed **səəŋ* from PK, which had innovated the word. Similarly **kmɔɔ* 'year' is restricted to Katuic and WB, as other Eastern Mon-Khmer languages reflect another root, cf. Khmer *chnam*, Bahnar *snam*, Vietnamese *năm*, and we must assume borrowing from PK into PWB.

The last example, the word for 'flower', is also apparently restricted to Katuic and WB, but presents phonological complications, as we would expect PWB to borrow the word as **piar* rather than **piir* (~*piər*). However, it is evident that in various etyma PK **ia* and PWB **ii* share a common origin from a raised/diphthonged ***aa* vowel, cf. 'banana' PK **priet*, PWB **priit*, Old Mon **praat*. In this case we have phonological evidence that this word for 'flower' goes back to a hypothetical Proto East Mon-Khmer ***paar*, and is an archaism rather than evidence of contact (nor evidence for sub-grouping).

3. Conclusion

Close etymological investigation coupled with an improved understanding of historical phonology confirms that there were close contacts between PWB and PK speakers, such that various basic vocabulary items were borrowed. While a much greater body of lexicon remains to be investigated, it seems likely that PK was consistently the donor language, and this may indicate something about the relative societal relations between these two groups. Work on improving the phonological and lexical reconstruction of Katuic and

Sidwell, P. 2003, "A note on the reconstruction of proto West Bahnaric and Investigation of early West Bhanaric-Katuic contact", in *The Mon-Khmer Studies Journal*, vol. 33, pp. 159-166. (purl.org/sealang/sidwell2003note.pdf)

Bahnaric linguistic history is not only progressing, but shows promise of usefully informing research on the historical contacts and relations of Eastern Mon-Khmer peoples in prehistory.

REFERENCES

- Diffloth, Gérard. 1976. "Mon-Khmer numerals in Aslian languages." *Linguistics* 174:31-38.
- Diffloth, Gérard. 1982. "Registres, dévoisement, timbres vocaliques: leur histoire en Katouique." *MKS* 11:47-82.
- Ferlus, Michel. 1991. "Vocalisme du Proto-Viet-Muong." Paper circulated at *the Twenty-fourth ICS-TL&L*. Chiang Mai University, October 10-11, 1991.
- Ferlus, Michel. 1992. "Essai de phonétique historique du khmer (Du milieu du premier millénaire de notre ère à l'époque actuelle)." *MKS* XXI:57-89.
- Jacob, Judith M. 1976. "Some problems arising from the orthography of consonants on the Khmer inscriptions." In Jenner et al., eds., pp 625-58.
- Jacq, Pascale and Paul Sidwell. 2000. *A Comparative West Bahnaric Dictionary*. München, Lincom Europa.
- Jacq, Pascale. 2002. "Phonetic realisations of /ʔC/ and /hC/ word-initial sequences in Jruq (Loven)." *MKS* 32:25-53.
- Jenner, Philip N., Laurence Thompson and Stanley Starosta (eds.). 1976. *Austroasiatic Studies*, Volume 1. Honolulu: University of Hawaii Press.
- Jenner, Philip N. and Saveros Pou. 1980-81. "A lexicon of Khmer morphology." *MKS* 9&10.
- Theraphan LuangThongkum. 1997. "The place of Lawi, Harak, and Tariang within Bahnaric." *MKS* 27:109-117.
- Theraphan LuangThongkum. 2001. *Languages of the Tribes in Xekong Province Southern Laos*. The Thailand Reseach Fund.
- Peiros, Ilia. 1996. *Katuic Comparative Dictionary*. Pacific Linguistics Series C-132. Canberra, Australian National University.
- Pinnow, Heinz-Jürgen. 1980. "Reflections on the history of the Khmer phonemic system." *MKS* VIII:103-30.
- Pou, Saveros. 1992. *Dictionnaire vieu khmer-francais-anglais: an Old Khmer-French-English Dictionary*. Paris: Centre de documentation et de recherche sur la civilisation khmre.
- Sidwell, Paul. 2002. "Genetic Classification of the Bahnaric languages: a comprehensive review." *MKS* 32:1-24.
- Thomas David. 1979. "The Place of Alak, Tampuon, and West Bahnaric." *MKS* 8:171-186.

Received: 10 May 2002

Department of Linguistics
 School of Pacific & Asian Studies
 Australian National University
 ACT, 0200,
 Australia
 <paul.sidwell@anu.edu.au>