

Genetic Classification of the Bahnaric Languages: A comprehensive review¹

Paul SIDWELL
Australian National University, Canberra

Summary

The Bahnaric language family is one of the more internally diverse branches of the Mon-Khmer (Austroasiatic) phylum. Among specialists a rough consensus emerged in the late 1970's that classified the Bahnaric languages into five coordinate sub-groups. The classification was established using lexical methods: lexicostatistics and 'distinctive vocabulary'.

More recent work on the reconstruction of Proto-Bahnaric and new data from the field has permitted a reappraisal and revision of the classification. A comprehensive revision of the classification is now possible based entirely upon comparative historical phonology—the method more usually preferred² by historical linguists.

This paper is in two parts. The first is a review of the relevant literature, yielding an overview of how the received classification emerged, and a critique of the methodologies. The second part discusses the new proposal, working through the detail of the phonological evidence. I propose three coordinate branches: North Bahnaric, West Bahnaric and Central Bahnaric—daughters of three Proto Bahnaric dialects defined according to overlapping independent phonological changes.

1. Introduction

In the 1960s a classification emerged which divided the Bahnaric languages into three coordinate sub-groups with geographical designations: North Bahnaric (NB), South Bahnaric (SB) and West Bahnaric (WB). Over the following decade various investigations into the classification of individual languages resulted in an expansion of this model to five coordinate branches, to accommodate newly identified sub-groups Central Bahnaric (CB) and East Bahnaric (EB, also called North East(ern) Bahnaric). Subsequent discussions in conferences and the scholarly literature concerning the classification of individual Bahnaric languages has been framed in terms of this model, i.e. within which of these sub-groups does language X belong?

To my knowledge, all the relevant discussions to date have emphasised results based upon consideration of 'distinctive vocabulary'. So far

¹I would like to thank the following people/organisations whose assistance made this research possible: Ms Pascale Jacq (ANU, Canberra), who accompanied me in fieldwork and in so many other ways has assisted me in my research efforts. Professor Bernard Comrie (Leipzig), who organised temporary employment for me with the Max Planck Institute. The Australian National University Faculty of Arts for a faculty-based small ARC grant in 2000 which assisted the present project (A\$4,800). The various academic and administrative staff of ANU who have assisted me in numerous ways. And finally Dr Gregory Anderson, University of Chicago, who provided comments on an earlier version of this paper.

as I can tell, the procedure is that, at first, self evident sub-groupings are identified by scanning the available data. A short list of basic vocabulary that appears to be restricted to each of the putative sub-groups is identified, and then used as a diagnostic device. The affiliation of an individual language—one that is not immediately self-evidently a member of any of the presumed sub-groups—is tested by counting the lexical correspondences with the lists of 'distinctive vocabulary'. The language in question is then assumed to be more closely related to the sub-group(s) with which it shares the greatest amount of 'distinctive vocabulary'. The principal practitioner of this methodology in respect of the Bahnaric languages, David Thomas, has argued that the results of this method are inherently more likely to produce a reliable genetic classification than traditional lexicostatistics or comparative phonology. I question Thomas' views on this question, and in this paper I propose a comprehensive internal classification of the Bahnaric languages on the basis of comparative historical phonology.

2. History of Bahnaric classification

2.1 *The period up to the mid-1960s*

Until the 1960's there was no coherent identification of a Bahnaric family as such—instead various low-level sub-groupings were noted in the literature, and these were listed simply as constituents of the wider Mon-Khmer family. The position of individual languages was also confused: for example, Pinnow (1959) classified Bahnaric languages into eight different groupings, with dialects of some languages assigned to different sub-groupings, and some even listed within what are now recognised as different branches of Mon-Khmer (i.e. with Katuic and Khmer).

Shorto, Jacob and Simmonds (1963) simplified and advanced Pinnow's scheme, removing some of the most striking anomalies. They placed the Bahnaric languages within the following six groups shown in Fig. 1.

Köho Group:	<i>Köho, Maa, Srê, Chrau</i>
Pnong Group:	<i>Biat, Stieng</i>
Bahnar Group:	<i>Bahnar, Sedang, Selang, Tareng</i>
Brao Group:	<i>Brao, Sué</i>
Jəru/Boloven Group:	<i>Jəru/Boloven, Niaheun</i>
Suoi Group:	<i>Alak, Kaseng, Lave, Suoi, Ca-Lo', Khoi, Kon-Tu, So, Nanyang</i>

Figure 1. Classification of Bahnaric and Katuic languages, drawn from Shorto, Jacob and Simmonds (1963)

The first two of these groups correspond to what is now generally referred to as South Bahnaric. Their Bahnar Group includes Bahnar and the North Bahnaric languages (although today specialists remain divided/confused over the status of Bahnar and Tareng). The fourth and fifth groups contain languages that are today recognised as West Bahnaric, and the last group is basically Katuic with several Bahnaric languages mistakenly included—Alak, Kaseng and Lave (Laveh), and perhaps even Suoi and So, depending upon which languages are actually indicated by these labels.

At that time there was still no clear notion of discrete West Bahnaric or Katuic groupings, and this is perhaps understandable given that very little material was available to scholars, those languages being located in relatively remote areas of Laos.

2.2 *Thomas (1966) and Thomas & Headley (1970)*

The most important advances of this time emerged from the various lexicostatistical studies of Mon-Khmer by David Thomas and Robert Headley. In respect of Bahnaric languages they distinguished two major branches—Stiengan and Bahnaran, later renaming them South Bahnaric and North Bahnaric. Thomas and Headley did not treat the West Bahnaric languages lexicostatistically for lack of data, but did correctly group them together, rectifying the scheme of Shorto, Jacob and Simmonds (1963). They also clearly established the division between Bahnaric and Katuic, which was already becoming evident to fieldworkers such as Nancy Costello and Dorothy Thomas. The lack of data also meant that some languages were, presumably on geographical grounds, assigned to the wrong families, e.g. Alak and Kaseng were mistakenly classified as Katuic rather than Bahnaric.

The classification of Thomas and Headley was the source of Diffloth's (1974:481) *Encyclopaedia Britannica* entry, which then became the standard popular source for the classification of these languages.

2.3 *Ferlus (1974)*

In France, Michel Ferlus (1974) published a paper on the classification of Bahnaric (and Katuic) languages, drawing on secondary sources (which included the studies mentioned above and many also from the substantial French colonial tradition) and his own original research. In this important paper (little cited by Anglophone scholars) Ferlus recognises the tripartite division of Bahnaric, employing the following nomenclature: *Bahnar-Sedang* (North Bahnaric), *Mnong-Maa* (South Bahnaric) and *Laven-Brao* (West Bahnaric). Very significantly for this time Ferlus identifies Alak and Kaseng as Bahnaric (specifically North Bahnaric, something he reiterated to me in person in 1997) and Tareng as Katuic. Ferlus' delimitation of West Bahnaric was based upon lexicostatistical and phonological considerations, made possible by his own fieldwork. Ferlus' results are therefore more reliable and up-to-date than those popularised at the same time by Diffloth. A fragment of the map Ferlus included with that paper is reproduced here as Fig. 2 (I include it primarily for the orientation of readers not so familiar with the geographical distribution of the languages).

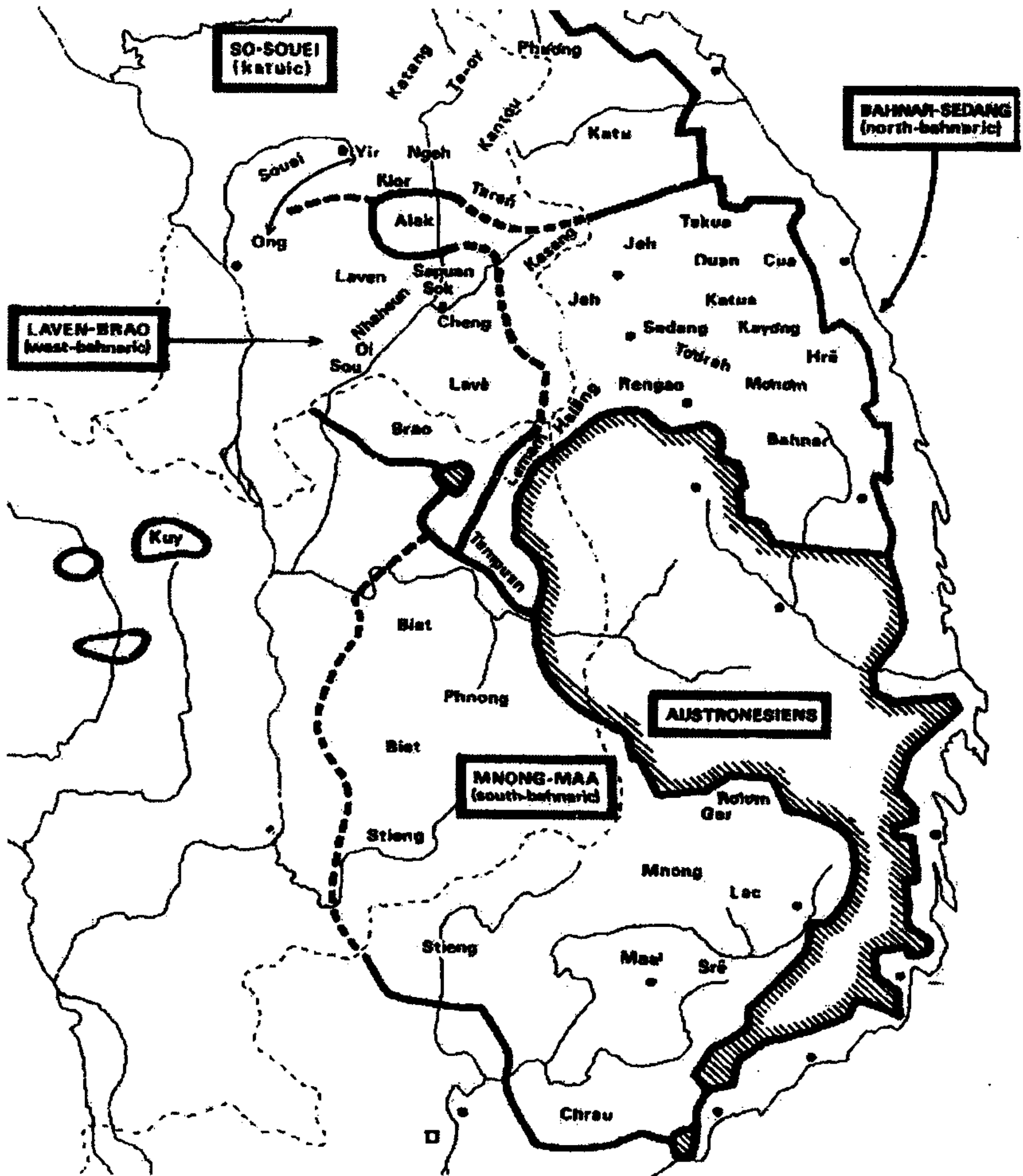


Figure 2. Part of Map from Ferlus (1974) "Carte des groupes linguistiques austroasiatiques dans le centre-indochinois"

Sidwell, P. 2002, "Genetic classification of the Bahnaric languages: A comprehensive review", in *The Mon-Khmer Studies Journal*, vol. 32, pp. 1-24. (purl.org/sealang/sidwell2002genetic.pdf)

2.4 *Smith (1972, 1979)*

Following on from the work of Thomas and Headley, Kenneth Smith also made lexicostatistical investigations of the language relationships, with special emphasis on the place of Sedang, a North Bahnaric language he researched for many years. Smith reports in his *Sedang Grammar* (1979:15) that his results essentially repeated those of Thomas and Headley, where they are comparable. One of Smith's most important achievements was a reconstruction of Proto North Bahnaric (PNB), first attempted in 1967, later revised and published in 1972. While the reconstruction was far from comprehensive, Smith's work did reveal various details of the lexical and phonological correspondences between the languages which call into question the tripartite model of Thomas & Headley.

As is generally recognised, the successful reconstruction of a proto-language is dependent (among other things) upon determining the correct model of the internal sub-groupings (if any) among the daughter languages. This is because internal sub-groupings imply the existence of lexical and phonological developments which are later than, and therefore not part of, the proto-language. Consequently an incorrect model of the internal phylogeny may result in one incorrectly assigning features to the proto-language.

Smith began with the premise (drawn from Thomas & Headley's classification) that Bahnar, which does not have phonemic registers (phonation types), sub-groups with the North Bahnaric languages, such as Sedang, which do have registers. However, Smith recognised that this premise runs into serious difficulties, as Bahnar is phonologically structured like a South Bahnaric language, raising the question of whether it was misclassified.

At this point Smith considered several possibilities: 1) that Bahnar is a high branching member of North Bahnaric, 2) that Bahnar can be treated as representative of South Bahnaric, or 3) that Bahnar itself branches directly from Proto Bahnaric. The issue is important because it directly affects whether register, which is unique to North Bahnaric, should be reconstructed as an innovation or a retention. If register is an innovation originating after the break-up of the proto-language, one must posit an intermediate node (sub-grouping) at which the feature developed, nominating which languages descend from that node. If it is a retention, there are no sub-grouping implications for North Bahnaric, which can be treated as a coordinate branch, and all other Bahnaric languages have simply lost the feature, which they may have done independently, or as descendants of one high order branch.

Smith decided to treat register as a retention from Proto Bahnaric, and left open the question of where Bahnar fits into the family tree. He returned to this question, in collaboration with colleagues Ken Gregerson and David Thomas several years later.

2.5 *Gregerson, Smith and Thomas (1976)*

Gregerson, Smith and Thomas (1976) (henceforth GS&T) is a detailed attempt to determine the proper classification of Bahnar within Bahnaric. They reviewed three types of evidence, 1) lexicostatistical, 2) distinctive vocabulary, and 3) phonological.

The lexicostatistical analyses of Thomas and Headley (1970) were reviewed, and the following two matrices shown here as Fig. 3 were discussed:

	Mnong	Chrau	Bahnar	Jeh	Sedang
Stieng			50	49	46
Mnong			53	50	45
Chrau	66		51	49	44
Köho	64	62	50	47	45

Chart 1. South Bahnaric cognate percentages

	Jeh	Halang	Sedang	Hre	Bahnar	Alak	Cua
Jeh					57	53	51
Halang	77				60	50	54
Sedang	63	65			57	52	51
Hre	59	61	68		62	53	49
Rengao	63	74	72	72	67	53	53
Bahnar						55	48
Alak							49

Chart 2. North Bahnaric cognate percentages

Figure 3. Lexicostatistical tables from Gregerson, Smith and Thomas (1976:377)

GS&T interpreted these matrices as showing that Bahnar has a consistently closer relationship with North Bahnaric (Jeh, Halang, Sedang, Hre and Rengao) than with South Bahnaric (Stieng, Mnong, Chrau, Köho). However, it should be noted that the cognate counts were conducted by straightforward inspection and subjective assessment of cognacy, in the absence of a rigorous reconstruction. It is interesting to compare the above matrices with one that Jacq and I presented in 2000, prepared on the basis of cognates identified by comparative reconstruction, and a conscientious effort to eliminate borrowing (which in retrospect was still inadequate). Our matrix covers 25 Bahnaric languages. Our interpretation at that time (now superseded) was to posit seven coordinate sub-groups. And although the cognate counts of GS&T vary considerably from those of Jacq & Sidwell, the overall patterns are similar where they are comparable—in both studies a special relationship between Bahnar and NB was indicated.

- West Bahnaric: *Brao, Cheng, Loveh, Loven, Nhaheun, Oi, Sapuon, Trieng*
- *Kaseng/Taliang*
- *Alak*
- South Bahnaric: *Mnong, Stieng, Chrau, Köho*
- *Tampuon*
- North Bahnaric: *Bahnar, Jeh, Halang, Monom, Rengao, Sedang, Didra, Hre*
- *Cua*

Figure 4. Classification of Bahnaric languages, drawn from Jacq & Sidwell (2000)

Brao	74	Cheng	75	72	Loveh	71	68	Loven	70	73	70	76	Nhaheun	70	74	71	71	72	Oi	71	73	74	72	78	69	Sapuon	56	51	59	66	61	51	55	Trieng	49	47	49	53	52	43	47	51	Taliang	51	49	51	57	52	45	51	48	76	Kasseng	48	48	48	51	51	43	48	42	45	43	Alak	46	38	43	44	42	42	42	39	38	39	49	Mnong	47	38	44	43	41	41	40	43	37	38	48	64	Stieng	42	35	38	40	40	38	38	41	36	36	51	65	64	Chrau	37	32	35	38	38	31	35	36	39	36	44	57	54	61	Koho	48	41	46	48	48	42	43	37	39	40	52	48	51	49	47	Tampuon	41	40	42	43	44	38	44	39	38	38	55	43	45	47	41	55	Bahnar	47	46	47	52	52	44	46	42	43	43	52	43	44	48	40	47	58	Jeh	49	46	46	49	51	45	49	38	44	41	54	44	47	49	43	48	63	78	Halang	41	45	43	43	43	38	44	40	42	40	51	37	37	38	35	43	64	55	Monom	45	41	43	47	45	38	44	38	37	37	49	42	43	42	37	48	60	57	Rengao	38	39	36	43	40	35	38	32	36	33	47	35	34	38	32	43	54	52	61	68	Sedang	42	41	40	45	43	40	41	38	40	38	52	40	42	44	37	44	59	58	76	81	Didra	46	42	43	45	44	41	43	39	42	39	51	39	37	41	34	44	62	55	72	67	76	74	Hre	40	40	43	49	46	41	44	41	41	35	46	39	39	39	38	43	48	49	44	46	49	46	45	Cua
------	----	-------	----	----	-------	----	----	-------	----	----	----	----	---------	----	----	----	----	----	----	----	----	----	----	----	----	--------	----	----	----	----	----	----	----	--------	----	----	----	----	----	----	----	----	---------	----	----	----	----	----	----	----	----	----	---------	----	----	----	----	----	----	----	----	----	----	------	----	----	----	----	----	----	----	----	----	----	----	-------	----	----	----	----	----	----	----	----	----	----	----	----	--------	----	----	----	----	----	----	----	----	----	----	----	----	----	-------	----	----	----	----	----	----	----	----	----	----	----	----	----	----	------	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	---------	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	--------	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	--------	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-------	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	--------	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	--------	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-------	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----

Figure 5. Lexicostatistical matrix of 25 Bahnaric languages, from Jacq & Sidwell (2000:4)

Sidwell, P. 2002, "Genetic classification of the Bahnaric languages: A comprehensive review", in *The Mon-Khmer Studies Journal*, vol. 32, pp. 1-24. (purl.org/sealang/sidwell2002genetic.pdf)

The next stage in GS&T's investigation was an attempt to reconstruct basic vocabulary that was unique to PNB and PSB respectively, and compare the results with Bahnar, "So the theory is that any sizeable amount of distinctively South Bahnaric vocabulary in Bahnar would clearly point to South Bahnaric genetic status for Bahnar." (p.381) The results are very interesting: of 24 etyma they considered to be unique to South Bahnaric, Bahnar shares only 4. Also the Bahnar pronouns and numerals are more structurally and phonetically similar to those of North Bahnaric languages. They also note some significant correspondences in distinctive vocabulary between Bahnar, Alak and Cua. Their interpretation is that:

These results could be seen as tending to put Bahnar and Alak in a Greater North Bahnaric group, contrasting slightly with central North Bahnaric. Cua also shows a strong affinity with NB, though with many idiosyncratic words, but in none of the cited items is it like SB. (p.387)

GS&T's interpretation of the evidence of distinctive vocabulary requires that the North Bahnaric register languages constitute a sub-grouping, Central North Bahnaric, within a larger northern group. While this is an immediately satisfying interpretation, it does not exhaust the possibilities. Their 'distinctive vocabulary' is rather small, and one can suggest that it reflects the results of a local sprachbund, given that the languages are effectively contiguous—alternately the 'distinctive' SB vocabulary could be a more recent innovation, occurring since the geographical separation of Bahnar and SB.

Finally GS&T examine some phonological evidence, but lacking a reconstruction of Proto Bahnaric, they recognise that the demonstration of common phonological innovations is beyond the scope of their paper, and instead they concentrate on identifying:

...phonological differences between NB and SB and then note to which B(ahnar) is aligned....The weight of the argument then lies on the number of different types of differences which point in the same direction. (p.387-8)

Strikingly GS&T find that, of five consistent phonological differences between NB and SB, "The writer knows of no consistent phonological difference between NB and SB wherein B(ahnar) follows NB". (p.398-9) In their view, the phonological differences between Bahnar and "South Bahnaric are no more than one would expect between distinct languages." (p.399) However, GS&T are reluctant to accept the phonological evidence, as it conflicts absolutely with their interpretation of the 'distinctive vocabulary'. Instead they opt for a compromise hypothesis, proposing that Bahnar is neither South Bahnaric nor North Bahnaric, but:

it represents another distinct branch of Bahnaric—Central Bahnaric....And not only Bahnar, but Alak, East Bahnaric (Cua), and West Bahnaric also share many of the phonological features of South Bahnaric and lexical features of North Bahnaric. It appears that this may argue for an attachment of Bahnar as

Central Bahnaric at a higher node on a par with the other Bahnaric branches. (p.402)

Thus GS&T conclude that the Bahnaric family has a flat structure, with five branches descended directly from Proto-Bahnaric.

2.6 Thomas (1979)

Thomas (1979) is a companion piece to GS&T. In it Thomas investigates the positions of Alak, Tampuon and West Bahnaric, using similar format and procedures to those applied by GS&T, plus some new data from a broader range of languages. This paper is a real attempt at a comprehensive and justified internal classification of the Bahnaric family.

The results are strongly consistent with those of GS&T where they are comparable:

The lexicostatistical evidence shows tight northern, southern, and western groupings, with Bahnar in the northern group, and Alak, Tampuon and Cua as isolates slightly more closely related to each other than to the northern, southern and western groups. (p.183)

However, Thomas allows himself to go further with the interpretation. While the phonological evidence presented indicates that the northern register languages are a very distinctive sub-group, in Thomas' view the same type of evidence is not particularly helpful for indicating the sub-groupings of the non-register languages. In fact, he mounts an attack against using comparative phonology for sub-grouping, stating:

.....identical phonological innovations are not at all unlikely. The possible range of sound shifts is very small so duplicate innovations should be expected. (p.181)

To illustrate his point Thomas mentions the **-k, *-ʔ > /-ʔ/* merger shared by some South Bahnaric and Chamic languages—clearly a case of an areal phonological change. However it should be equally clear that this change is a rather trivial one, and not at all the sort of conditioned split, or non-trivial sequence of changes, that would be accepted by most scholars as unambiguous evidence of genetic sub-grouping (such as the complicated development of phonemic registers in NB, which ironically Thomas does accept as indicative). On the basis of this rather weak argument the most trusted and reliable genetic sub-grouping methodology in the historical linguist's armoury is rejected, and instead, he declares that:

For measuring genetic distance the measure must be the degree of semantic shift.... (p.180).

To my mind Thomas is making a logical error, evidently confused about the basis of genetic classification. The principle at stake here is summed up by Campbell (1998:186) when he writes, "For sub-grouping, only shared innovations prove reliable". Shared innovations can be phonological or semantic, the point is that a change happens at some stage in the life of a

language, and therefore the effects of that change are reflected uniquely in the daughters of that language (excepting parallel changes that happen coincidentally by chance, borrowing or linguistic tendency). However, while both phonological and semantic changes may be invoked to support a classification, both must be considered *in the context of a reconstruction of the direction of change(s)*.

Discovering the direction of change(s) is part of a reconstruction of the history of the lexicon (which also presumes to some extent a phonological reconstruction). Effectively a comprehensive reconstruction is needed to distinguish between what is innovated and what is archaic. As Trask (1996:182) correctly reminds us, “Shared archaisms are of little or no use in establishing groupings within families.”

Even accepting that semantic/lexical comparisons can provide acceptable data for sub-grouping hypotheses, in the context of reconstruction, there remain good reasons to give priority to phonological evidence. Phonological change is systematic—affecting all words that have the relevant phonemes. On the other hand, semantic changes affect individual words—rather than the system of the language. It can be difficult to distinguish lexical changes that are internal to a language versus those that have resulted from borrowing(s) from closely related languages, even with a phonological reconstruction. In that case our lexical reconstruction may be incorrect, and dozens of words may be reconstructed for the proto-language on the basis of more recent borrowings. In contrast, a phonological change may be manifest across all types of lexicon (subject to phonotactic and other constraints), versus a loan phoneme or phonological processes restricted to borrowed vocabulary.

So Thomas appears to have done the following: 1) he confused the then poor development of Bahnaric reconstruction, and consequent inability of comparative phonology to solve the problem at hand, with a mistaken view that comparative phonology is unable to solve the problem in principal, and 2) he mistakenly asserted that evidence of semantic change in the lexicon can be used to justify a classification without requiring that the source and direction of change be demonstrated by comprehensive reconstruction.

Having considered and rejected comparative phonology, Thomas based his sub-grouping proposal on ‘distinctive vocabulary’, although the procedure he used is a little different to the one employed by GS&T. What he did was list words/generalised forms corresponding to 34 semantic categories found in the vocabularies of languages/subgroups being investigated. In effect, it was the lexicostatistics applied to a very short wordlist. On this basis he found a “striking convergence of Alak, Tampuon and Bahnar as a unified group”—agreeing in 18 out of 31 categories (forms are not listed for all languages for the 34 fields), proposing that they constitute “a very loose central group” as opposed to the other Bahnaric sub-groupings, which are “tightly knit”. The scheme (p.183) is as shown in Fig. 6.

North Bahnaric	<i>Sedang, Hré, Halang, Jeh, Rengao</i>
South Bahnaric	<i>Koho, Chrau, Mnong, Stieng</i>
West Bahnaric	<i>Loven, Nhaheun, Cheng, Oi, Laveh, Brao</i>
Central Bahnaric	<i>Bahnar, Tampuon, Alak</i>
Eastern Bahnaric	<i>Cua, Kotua</i>

Figure 6. Classification of Bahnaric languages by Thomas (1979)

2.7 *Reconsidering Gregerson, Smith and Thomas (1976) and Thomas (1979)*

The two companion papers discussed above are the two most important publications to date concerning the internal classification of the Bahnaric languages, having effectively established the framework for all subsequent discussion of the topic of which I am aware. However, as I have indicated above, I have strong reservations about their methodological approach, and am therefore sceptical of their conclusions. The authors place great weight on so-called 'distinctive vocabulary' in establishing genetic sub-groupings. GS&T examine 24 sets, while Thomas (1979) looks at a total of 34. The former also make reference to the strong agreement of the pronoun sets and numerals for Bahnar and North Bahnaric. The conclusion of both papers, that Bahnar groups with Alak and Tampuon, in opposition to NB, is reconciled with the evidence of the pronouns and numerals by positing a local sprachbund. GS&T speculate that a language contact area formed amongst those communities who found themselves on the northern side of the Chamised Central Highlands around a millennium ago. However, this proposal may be redundant, as it is also possible that the northern 'distinctive vocabulary' upon which they rely is archaic, and therefore of no significance for sub-grouping. This latter possibility is consistent with the results of my own research, which indicates that SB and WB are more lexically innovative, for example South Bahnaric has borrowed heavily Khmer (e.g. numeral *pram* 'five', pronoun *?an* 'I' etc.) while West Bahnaric has taken much from Katuic (e.g. borrowing *səŋ* 'five' and much other basic vocabulary).

In any case, the lexical evidence presented for Central Bahnaric is not as strong as claimed. Of the 18 isoglosses identified by Thomas (1979), only one is unique to the three languages Bahnar, Alak and Tampuon : *?akan* 'woman', which is either an archaism or a loan, as it is also found in Katuic (e.g. Bru, Kui etc. *kan* 'female'). In all other cases the etyma are shared with one or more other Bahnaric sub-group, so it is difficult for me to accept that the 'distinctive vocabulary' evidence argues for any special relationship, even if one accepts GS&T's methodological assumptions. On the contrary, the data strikes me as being consistent with Thomas' own lexicostatistics that identified Alak and Tampuon as "isolates" (p.183).

It is clear to me that the use of 'distinctive vocabulary' should not be used for genetic sub-grouping, as it is likely to seriously skew results. It is evident that once the languages are compared on the basis of the 100 wordlist the skewing effect of the very short list(s) is largely eliminated and the situation is seen in a more reasonable perspective—accepting the proviso that lexicostatistics offers only indicative, rather than determinative, results.

2.8 *Diffloth (1991)*

Theraphan (1997:114) reports that Gérard Diffloth presented a paper in 1991 to the *24th International Conference on Sino-Tibetan languages and Linguistics* in which he identified a 'new' Bahnaric language *Tarieng*, and offered the Bahnaric stammbaum shown in Fig. 7.

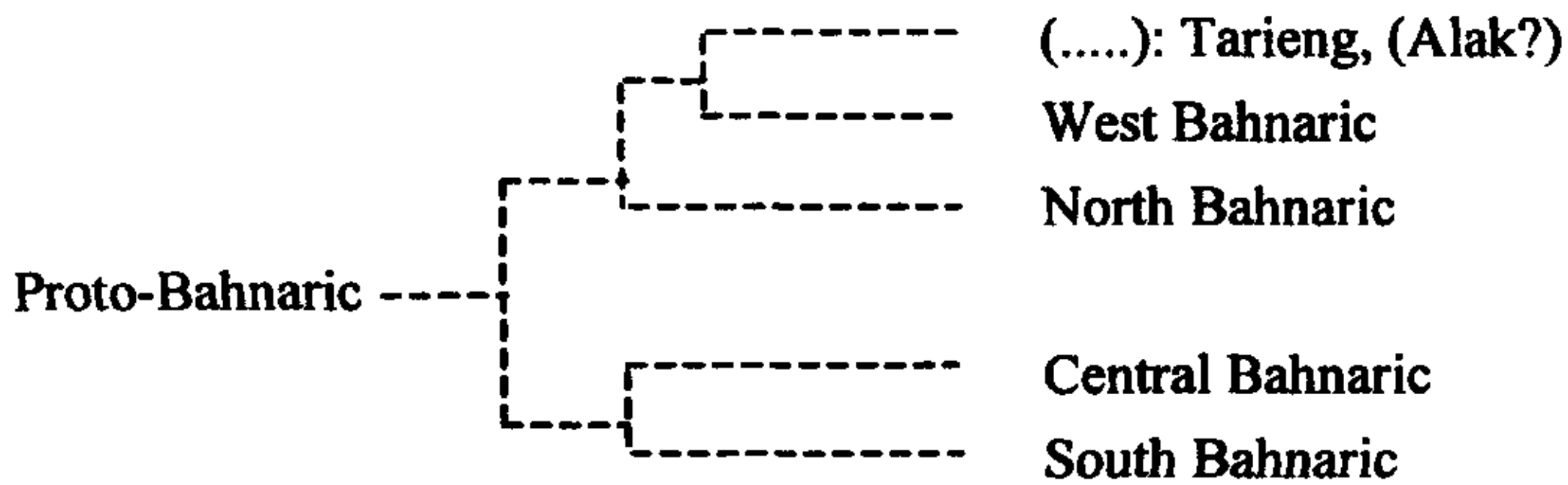


Figure 7. Classification of Bahnaric languages by Diffloth (1991)

Unfortunately I do not have Diffloth's paper², but features of the stammbaum can be compared to Thomas' (1979) scheme, assuming that Diffloth's labels refer to the same languages. An exception is that the position of East Bahnaric (Cua etc.) is not evident. Very significantly Diffloth's scheme incorporates various new sub-grouping proposals, so it deserves some consideration.

Diffloth proposes a primary bifurcation between South and Central Bahnaric versus the rest of the family—this is superficially very attractive, given the phonological evidence of close agreement between them. However, on the same grounds one might also expect West Bahnaric to be grouped with the non-northern languages—lexically and phonologically North Bahnaric and West Bahnaric are maximally different within the family, so this classification is quite anomalous. On the other hand, the grouping of Tarieng with West Bahnaric is no surprise. In Jacq & Sidwell (2000) we report the results of our lexicostatistical project which included a basic wordlist of Trieng (presumably Diffloth's Tarieng) obtained from the SIL library in Bangkok. We found that Trieng shares between 51% and 66% of basic vocabulary with West Bahnaric languages, which we interpret cautiously as indicative of high branching membership of West Bahnaric. On the other hand, if Diffloth's Tarieng is actually the language known to me as Taliang, this presents a different situation (see discussion below).

2.9 Theraphan (1997a & b, & ms.)

In three recent papers Theraphan L.Thongkum discusses the classification of several Bahnaric languages of Southern Laos whose status is undetermined and/or confused in the previous literature. The languages are Lawi, Harak (Alak), Tariang³, Kaseng, Yaeh and Suai. She investigates the issues by applying Thomas' 'distinctive vocabulary' method to her recently collected field data.

The first of these papers deals with Lawi, Harak and Tariang. In respect of Lawi, Theraphan finds and lists 13 words with West Bahnaric cognates, and on the basis of this she declares "There is no doubt that Lawi is a West Bahnaric language". While I do not doubt that her conclusion may be correct, it does not necessarily flow from consideration of this list.

²Despite e-mail correspondence with Diffloth, I have been unable to directly obtain copies of any of his unpublished materials.

³I suspect that Theraphan (or I!) may be confused about language names: her Tariang is evidently the same language as Taliang discussed in Jacq & Sidwell 2000, and which we classified as a sister dialect of Kaseng.

Etymological examination reveals that of the 13 etyma, at least seven are borrowings from Katuic! Of the rest, *tri* 'fish' is borrowed from Khmer; *tapit* 'ear' is shared with Jeh (North Bahnaric) and all of South Bahnaric. Only *pa* 'shoulder', *pihac* 'sand' and *piŋ* 'tooth' are possibly unique West Bahnaric isoglosses (so far as I can tell). Here we have a very instructive example of the limitations of the 'distinctive vocabulary' approach—54% of the short Lawi lexicon has Katuic, rather than Bahnaric, provenance—so it can hardly be offered as evidence of Bahnaric, let alone West Bahnaric affiliation, without a reconstruction to show when and how these words entered the lexicon.

The discussion of Harak and Tariang are in direct response to Diffloth's (1991) proposal, mentioned above. Harak and Tariang are compared to Lawi on the basis of a 13 word list, corresponding to the same categories used for the West Bahnaric comparison above, except that words for 'paddy' are substituted for 'shoulder'. Theraphan's interpretation of this short comparative list is inconclusive:

...Harak and Tariang have a mixture of northern, western and southern elements; therefore it is difficult to say exactly where they fit. (p.115)

She also attempts a phonological comparison with Smith's (1972) Proto North Bahnaric, but again the result is inconclusive:

The examples given in Table 6 seem to indicate that each language has its own historical phonology. (p.116)

Of course, all languages have 'their own historical phonology'—in this particular case the writer means that she sees no evidence for sub-grouping. However, the pull of the paradigm established by Thomas et al. is so strong that Theraphan is unable to resist. She suggests:

Perhaps, we need a more suitable name for a kind of loose or flexible sub-branch in which the Bahnaric languages, such as Harak, Tariang and so forth, can fit, since this type of language has a combination of western, northern and southern characteristics. (p.118)

Astonishing! Theraphan suggests that one can propose 'flexible sub-branches' into which we can classify languages that show characteristics of more than one branch (presumably archaisms or loans)! The concept is positively pre-Linnean, and runs exactly counter to the methodological principles that I understand underlie genetic classification.

In the second paper Theraphan discusses the classification of Suai of Saravan Province in Laos, for which she collected a list of 858 words during her 1995 fieldwork. 21 Suai words are given with various Bahnaric comparisons, and in all but one case they agree with West Bahnaric (in fact they agree particularly well with Laven (Jruq), in the third paper [below] she refers to Suai as Juk, which may be cognate with Jruq). The issue at stake in this paper is to distinguish this language from the various Katuic dialects/tribes that have similar names, and in that regard Theraphan succeeds.

The third paper is a more elaborated synthesis of the first two, which additionally treats some data from Kaseng and Yaeh (spoken in remote areas of Sekong Province). Her important findings are that Alak should not be grouped with Tariang, although Kaseng and Yaeh can, and neither should Alak be grouped with Bahnar within Central Bahnaric. Instead she proposes that these languages belong with an expanded North Bahnaric (having three sub-branches), according to Fig. 8.

North Bahnaric:	Northeastern: <i>Tariang, Kaseng, Yaeh, etc.</i> Northwestern: <i>Alak etc.</i> Other North Bahnaric languages (no sub-group name or language names listed)
West Bahnaric:	<i>Laven, Lavi, Juk, Brao, Su', Hyah Hueny, etc.</i>
East Bahnaric:	<i>Takau, Cua, Sedang, Hre, etc.</i>
Central Bahnaric:	<i>Tampuon, Bahnar etc.</i>
South Bahnaric:	<i>Mnong, Stieng, Sre, Chrau, etc.</i>

Figure 8. Classification of Bahnaric languages by Theraphan (ms.)

And here lies the state of Bahnaric classification at the end of the 20th century— Theraphan's model is almost exactly the same as Thomas' of 20 years earlier, except that Alak is reclassified—and it along with several 'new' languages are accommodated within an expanded North Bahnaric, recalling an earlier suggestion of GS&T. There is no change to the underlying structure of five coordinate groups in a flat tree. The scheme is established on the basis of the 'distinctive vocabulary' method rather than by comparative phonological reconstruction.

3. Bahnaric phonological reconstruction

For me the above situation is inherently unsatisfying. The reconstruction of Proto-Bahnaric has been my principal research pursuit since 1994, and continues to be so. It is clear to me that the success of this reconstruction is necessarily dependent upon having a reliable model of the internal classification of the Bahnaric family if I am to correctly determine the historical sequence(s) of any/all phonological and lexical developments. The reverse is also true—phonological reconstruction is needed to justify any/all sub-groupings. In my view phonological reconstruction and genetic classification are simply important aspects of the same thing, and as progress in one aspect feeds progress in the other a coherent model of the linguistic history emerges. I reject the 'distinctive vocabulary' approach as it has been practised in respect of Bahnaric—it cannot distinguish genetic from sprachbund effects, and cannot establish the relative diachrony of changes. I do accept that lexical techniques, such as lexicostatistics, do offer useful preliminary orientation, but their results can only be indicative, not determinative. Below I present the current finding of my on-going Proto Bahnaric project, and the consequences for the internal genetic classification of the family.

3.1 Proto Bahnaric Consonants

At this stage I am not ready to present a comprehensive reconstruction of the Proto Bahnaric consonant system—specifically I have yet to finalise the reconstruction of minorsyllable consonants and initial clusters because I am still revising the reconstruction of phonotaxis. However it is still

possible to usefully discuss some details of the history of the consonants and apply these insights to the classification problem. Basically my view is that Bahnar consonantism represents the proto-system with the least number of changes. Bahnar appears to preserve glottalisation of consonants in most/all circumstances, whereas this feature has been lost, partially or fully, from other languages of the family. As the loss of glottalisation may happen independently it appears to be irrelevant for sub-grouping. In my view there are only two consonantal changes that are important enough to raise here. The first is the merger of initial **s* and **c* to /s/ in a number of languages, including Bahnar, and the second is a *t : s* correspondence, for which I reconstruct the cluster *sʔ* (on the basis of external comparisons).

GS&T, following Smith (1972), pointed out that “B[ahnar] shares with SB and Cua the initial *s* in contrast to the NB *č* [sic.]” (p.392). Smith (1972) reconstructed PNB **ch* for this correspondence, in addition to PNB **s* and **c*. It is straightforward to show that the etymologies underlying Smith’s **c* reflect Vietnamese borrowings and local words, so we are really only dealing with a two way contrast, which can be illustrated with the following comparisons:

English	PWB	Alak	Cua	Kass.	Tamp.	PSB	Bahnar	PNB ⁴	PB ⁵
‘eat’	<i>*ca</i>	<i>ca</i>	<i>sa</i>	<i>ca</i>	<i>cha</i>	<i>*sa</i>	<i>sa</i>	<i>*ca</i>	<i>*ca</i>
‘nine’	<i>*cɿn</i>	<i>təŋcin</i>	<i>sit</i>	<i>kəcen</i>	<i>ɲchɿn</i>	<i>*sɿn</i>	<i>təsɿn</i>	<i>*təcɿn</i>	<i>*təcɿn</i>
‘horse’	<i>*seh</i>	<i>seh</i>	<i>hleh</i>	<i>seh</i>	<i>cheh</i>	<i>*ʔəsəh</i>	<i>ʔəseh</i>	<i>*ʔəseh</i>	<i>*ʔəseh</i>
‘honeybee’	<i>*sūt</i>	<i>sut</i>	<i>hlut</i>	<i>sut</i>	<i>chūt</i>	<i>*sūt</i>	<i>sut</i>	<i>*sōt</i>	<i>*sūt</i>

As there is no apparent conditioning motivation for a phonemic split, I reconstruct PB **c* for the first correspondence set, and PB **s* for the second. It is evident that there has been a merger of **c* and **s* in at least Tampuon, South Bahnaric and Bahnar. However, even though the shift of **c* > /s/ is evident in Cua, Cua preserves the phonemic contrast, with **s* reflected as /hl/. This indicates strongly that the phonetic shift of [c] > [s] was relatively late in Cua; certainly it had to occur after the shift [s] > [hl]. It also means that this change in Cua is unlikely to be related to the merger in Tampuon, South Bahnaric and Bahnar. This is very significant, because it establishes that the change of [c] > [s] (without regard to the phonemic consequences) has probably happened independently in the history of Bahnaric. It is therefore also possible, although perhaps unlikely, that the same/similar phonetic changes in Tampuon, South Bahnaric and Bahnar (which did have phonemic consequences) occurred independently of each other, and consequently this change is not a strong piece of evidence for sub-grouping.

The other correspondence is illustrated with the following comparisons:

⁴Please note that all PNB forms cited are my provisional reconstructions unless otherwise indicated.

⁵Please note that all PB forms cited are my provisional reconstructions unless otherwise indicated.

Gloss	PWB	Alak	Cua	Kass.	Tamp.	PSB	Bahnar	PNB	PB
'bone'	*gərtiŋ	kətiŋ	kətek	tiaŋ	ntəiŋ	*nətiŋ	kətiŋ	*kəsiŋ	*gəs?iŋ
'lizard'	—	—	—	—	—	*tum	—	*kəsum	*—s?um
'tree-trunk'	*təm	təm	—	təm	təm	*təm	təm	*pəsem ⁶	*s?əm
'cloud'	—	—	—	—	—	*tūk	tūk	*sūk	*s?ūk

Smith (1972) proposed **ts* for this correspondence, but that suggestion lacks any typological justification. External comparisons are available for the 'bone' etymon, and these are very instructive: Khmer *ch?iŋ*, Proto Waic **s?aŋ* (Diffloth 1980), Proto Viet-muong **šəaŋ* (Sokolovskaja 1978). The Khmer form is an excellent phonetic match with the reconstructed mainsyllable for both West Bahnaric and Proto Bahnaric. The vowels of the Proto Waic and Proto Viet-muong forms lend even further confirmation of this etymology, as I had independently reconstructed the source of PB **i* as coming from diphthongisation of an earlier **a* (compare, for example, Proto Bahnaric **prit* 'banana' with written Mon /prat/ 'banana'). The **s?* cluster would readily yield /s/ reflexes of North Bahnaric by loss of the glottal stop, which is a simple lenition that requires no particular conditioning. The /t/ reflexes elsewhere result from fusion of the apical fricative and glottal stop. This data strongly suggests a primary bifurcation between North Bahnaric and the rest of the family, although before jumping to any conclusions I will proceed below to explain my reconstruction of the Bahnaric vowels, and the implications this has for sub-grouping.

3.2 Proto Bahnaric Vowels

I now have a fairly complete model of the Proto Bahnaric vowel system and the essential details of its subsequent developments. The current proposal posits the following vowel phonemes for Proto Bahnaric as in Fig. 9.

	Long			Short			Diphthonged	
*i	*i	*u	*ĩ	*ĩ	*ũ	*ia	*ua	
*e	*ə	*o	*ě	*ǰ	*ǫ			
*ɛ	*a	*ɔ	*ě	*ǰ	*ǫ			

Figure 9. Proto Bahnaric vowels

Below I provide comparisons that are illustrative of the phonological correspondences that support the reconstruction. The list is only a fraction of the relevant comparisons that I have compiled. The approach I have taken here is to provide a selection that show the least conditioned changes. Open syllables and those closed by glottals have also been avoided in order to clearly distinguish vowel lengths. Rimes with final palatals have been avoided (with some exceptions) because of their tendency to raise vowels. Etyma that are well distributed across the family have been preferred, but some phonemes are less frequent than others, and in such cases I have had to assemble more sets with fewer members. Where possible, rimes with the same or similar finals have been compared in order to clearly demonstrate phonemic contrast. The comparisons follow with commentary:

⁶Note that in NB this word is used with meanings such as 'source, origin, primeval', while in WB both these and the 'tree trunk' meanings coexist. The semantic development is parallel to English 'root' > 'source'.

PB *ua, *ɔ, *ɔ̃

Gloss	PWB	Alak	Cua	Kass.	Tamp.	PSB	Bahnar	PNB	PB
'four'	*pɔn	pon	pɔt	puan	pwan	*puən	puan	*puan	*puan
'buy'	*rɔt	rot	rɔt	ruot	—	*ruət	rat	*ruat	*ruat
'lizard'	*kɔj	—	—	—	—	*nəkuəj	məkuej	*kuaj	*kuaj
'child'	*kɔn	kɔn	kɔt	kɔn	kɔn	*kɔn	kɔn	*kɔn	*kɔn
'hungry'	*pəŋɔt	pəŋɔt	pəŋɔt	—	—	*pəŋɔt	pəŋɔt	*pəŋɔt	*pəŋɔt
'carry'	*kɔj	—	—	—	—	—	kɔj	*kɔj	*kɔj
'white'	*ʔbək	p'ɔk	—	—	—	*bək	—	(təbək ¹⁷)	*ʔbək
'cheek'	*dəbɔk	t'əp'ɔk	dəvɔk	bɔk	—	—	—	*təbɔk	*dəʔbɔk
'take'	*cɔk	cɔk	sɔk	cɔk	j'ɔk	*sɔk	—	*cɔk	*cɔk

It is important to note the solid contrast between the low back *ɔ and the diphthong *ua. In WB and Cua this contrast was lost: in WB there was a complete merger, and reorganisation, such that both [ɔ] and [ua] occur but they are in complementary distribution (in Jacq & Sidwell 2000 we reconstructed *ɔ for this merger, which subsequently developed into a new /ɔ/ versus /ua/ contrast). In Cua *ua has simply merged with *ɔ. In Alak there was also a levelling, but in that case the diphthongs merged with the long mid vowels. The short *ɔ̃ is generally restricted to occurring before back consonants, and it is possible that comparative and/or internal reconstruction could suggest an origin for these that connects them with *ua or *ɔ.

PB *o, *ɔ̃

Gloss	PWB	Alak	Cua	Kass.	Tamp.	PSB	Bahnar	PNB	PB
'urinate'	*kənom	nnom	kanum	ʔənom	daom	*[ɲ]nom	ʔnom	*ʔnom	*ʔnom
'ear'	—	nto(?)	tol	tor	taor	*tor	—	—	*tor
'[ant]hill'	*bətul	—	ŋɔlh	—	—	*[nə]tül	pətɔl	*pətül	*bətɔl

The data supporting *o and *ɔ̃ are more substantial than presented here—these vowels are particularly well established on the basis of some widely distributed etyma with glottal finals.

PB *u, *ũ

Gloss	PWB	Alak	Cua	Kass.	Tamp.	PSB	Bahnar	PNB	PB
'tuber'	*bum	p'um	—	buom	p'um	*bum	ʔbum	*bum	*ʔbum
'axe'	*c[ə]rjuŋ	cuŋ	suak	cuaŋ	coŋ	*suŋ	suŋ	*cuŋ	*cuŋ
'fire'	*ʔũŋ	ʔuŋ	—	ʔũŋ	ʔuŋ	*ʔũŋ	ʔũŋ	*ʔũŋ	*ʔũŋ
'enter'	*mũt	mut	mũt	—	mut	*mũt	mũt	*mũt	*mũt

The reconstruction of *u and *ũ is straightforward. It is evident that some languages have lengthened reflexes of short *ũ in various environments, but this is often part of a push/pull chain which involves the diphthongisation of the long *u—most diphthonged reflexes in the above comparisons have been eliminated by the sub-group level reconstructions. A similar pattern is observed among the other high vowels presented below.

¹⁷Jeh

PB *i, *ĩ

Gloss	PWB	Alak	Cua	Kass.	Tamp.	PSB	Bahnar	PNB	PB
'bone'	*gərtiŋ	kətiŋ	kətek	tiaŋ	ntəiŋ	*nətiŋ	kətiŋ	*kəsiŋ	*gəsʔiŋ
'banana'	*pərit	prit	pəret	priat	preət	*pərit	prit	*pərit	*prit
'swidden'	*mir	—	—	miar	mer	*mir	mir	—	*mir
'thunder'	*kəriŋ	kəriŋ	—	—	—	—	grəm	*grəm	*grim
'drum'	*səgĩr	c'ak'ĩr	gəl	ʔəgər	chk'ĩr	*səgər	səgər	*səgər	*səgĩr
'stir'	*wĩr	—	—	ʔəwĩr	wĩr	*wər	—	*wər	*wĩr
'stand up'	*dĩk	t'ĩk	dĩk	—	—	*dək	—	*dək	*ʔdĩk
'heavy'	*gəjĩk	k'ac'ĩk	—	—	—	*kəjək	—	—	*gəʔjĩk
'foot'	*jĩŋ	c'ĩŋ	—	jĩŋ	c'ũŋ	*jək	—	—	*ʔjĩŋ

The reconstruction of central vowels *i and *ĩ is absolutely crucial to the classification proposals discussed below. The strongest evidence for *i lies with WB, where the vowel is frequent, and solidly in contrast with other high vowels. All other Bahnaric languages show either a corresponding [i] or diphthonged variant. However, I have found no conditioning factor which could account for split, and therefore reconstruct PB *i (at least, I am forced to reconstruct a contrast between long high unrounded vowels; although I strongly suspect that *i was realised as a diphthong [iə]). Further support for *i is the evidence for a corresponding short vowel *ĩ. Reflexes indicating a short high central vowel are found in WB, Alak, Cua, Tampuon and Kaseng—systemic considerations argue strongly for corresponding short and long vowels. The short *ĩ merged with *ɛ in SB and Bahnar, and independently in NB (the PNB *ɛI reconstruct is the antecedent of short lax register /ə/ (frequently realised [ɛ])). In Tampuon there seems to have been a conditioned split to /ũ/ before velars, otherwise /i/ is preserved. These last points are indicative of Bahnar sub-grouping with SB in opposition to the rest of the family.

PB *i, *ĩ

Gloss	PWB	Alak	Cua	Kass.	Tamp.	PSB	Bahnar	PNB	PB
'sky'	*pəliŋ	—	plek	pliŋ	pləŋ	—	plɛŋ	*pəliŋ	*pliŋ
'kill'	*gərcet	kəsit	kəset	jit	khchet	*kəsit	—	*kəcit	*gəcit
'dig'	—	cir	sil	cir	c'er	*sir	sir	*cir	*cir
'nine'	*cĩn	təŋcin	sit	kəcen	ɲchĩn	*sĩn	təsĩn	*təcĩn	*təcĩn
'full'	*bĩŋ	p'ĩŋ	vik	biŋ	—	*beŋ	ʔbeŋ	*bĩŋ	*ʔbĩŋ

The high front vowels *i and *ĩ are rather infrequent in closed syllables, and this hampers the reconstruction somewhat. None-the-less, there are several well distributed etyma which strongly support the reconstruction, particularly in open syllables (not listed here).

PB *e, *ě

Gloss	PWB	Alak	Cua	Kass.	Tamp.	PSB	Bahnar	PNB	PB
'mushr-m'	*pəset	—	pəhlet	—	—	*nəset	—	*pəsět	*pəset
'drink'	*ɲet	—	ɲet	set (?)	ʔaet (?)	*ɲet	—	—	*ɲet
'squeeze'	*det	k'at'et	—	—	—	—	—	*det	*ʔdet
'c-pede'	*gəjěp	kajip	dep	jip	—	— ⁸	kəsep	*kəjěp	*gəjěp
'muntjak'	—	—	jil	jil	c ʔl	*jʔl	jəl	*jěl	*jěl

The mid front vowels *e and *ě are even less frequent than the high front vowels, but it is clear that we can establish a contrast in the least marked of closed syllables—those with final apicals (less likely to condition changes such as fronting or rounding). Note that both mid and high front vowels are more frequent after palatal initials, and this may be historically significant, but even in these marked environments they contrast with all other vowels.

PB *ia

Gloss	PWB	Alak	Cua	Kass.	Tamp.	PSB	Bahnar	PNB	PB
'good'	*nem	lem	—	liam	rijām	*niəm	ltem	*liam	*liam
'feed/raise'	*cem	—	seap	—	chem	*siəm	hiem	*ciam	*ciam
'left'	*cəʔew	—	—	caʔiaw	chɲʔǎw	*gəʔiəw	ʔɲiew	*ʔiaw	*—ʔiaw
'tongue'	*—pet	mpet	rəpiat	piat	mpjět	*—piət	rəpiet	*rəpiat	*ləmpiat ⁹
'hail'	*pərel	—	—	prial	pril	*pəliər	prəl	*pərial	*prial
'chicken'	*ʔer	ʔer	ʔir	ʔiar	j'er	*ʔiər	ʔjěr	*ʔiar	*ʔiar
'root'	*reh	reh	reah	rias	rijāh	*nəriəh	rəh	*riah	*riah
'elbow'	*keŋ	—	—	kiaŋ	—	*kiəŋ	—	*keŋ	*kiaŋ
'hundred'	—	—	hrek	hriaŋ	rijāŋ	*həriəŋ	hrěŋ	*hərəŋ	*hriaŋ

The front diphthong *ia requires particular discussion. I have listed 9 sets of supporting comparisons to show the very regular patterning across languages before a range of finals. The distribution is roughly comparable to that for *ua presented above. However, there are few, if any, comparisons distributed across the family that support the reconstruction of *ɛ or *ě in closed rimes. There are many comparisons supporting *ɛ in open rimes, an environment in which *ia never occurs. This is clearly a complementary distribution, and strongly suggests that *ia and *ɛ could be treated as one phoneme, which is precisely the approach taken by Jacq and Sidwell (2000) for PWB. However, at the PB level there is clear evidence for a *ua / *ɔ distinction, and considerations of symmetry therefore lead me to suggest that there was a merger of *ia and *ɛ in closed syllables in Proto Bahnaric, with the latter only persisting in open rimes. Careful external comparison may further clarify this issue.

The short ɛ occurs in all Bahnaric languages in closed syllables, but it is most frequent in expressive/iconic vocabulary and as an allophone of /ǎ/

⁸Although I reconstructed PSB *kərʔip 'centipede', which can be compared to Bahnar kəʔep 'centipede', I suspect that both borrowed this word from Khmer /kəʔep/.

⁹The 'tongue' word appears to be an infixed reflex of *liat 'to lick'. If so, the nominalisation may be due to an infix *-m-, and/or *-p-.

and/or /ǝ/ before palatal finals. I don't offer any comparisons here, but I suppose that it would have been similarly present in the Proto Bahnaric. Also systemic considerations arising from the fact that a long **ε* is reconstructed also suggest **ǣ*.

Another possibility is that both /*ě*/ and /*ǝ*/, variously limited in their distributions, are secondary and should not be reconstructed for PB. However, the data is not yet sufficient for me to reach any conclusions.

PB **ə*, **ǣ*

Gloss	PWB	Alak	Cua	Kass.	Tamp.	PSB	Bahnar	PNB	PB
'liver'	* <i>kələm</i>	<i>kləm</i>	<i>kləim</i>	<i>kləm</i>	<i>kləp</i>	* <i>kələm</i>	<i>kləm</i>	* <i>kələm</i>	* <i>kləm</i>
'open (v.)'	* <i>pək</i>	<i>pək</i>	<i>bəik</i>	—	—	* <i>pək</i>	—	*[<i>pək</i> ¹⁰]	* <i>pək</i>
'answer'	* <i>təl</i>	—	—	<i>təl</i>	—	—	<i>təl</i>	* <i>tel</i>	* <i>təl</i>
'lid, cover'	* <i>gədǎp</i>	<i>kərtǐp</i>	—	—	—	* <i>gərdǎp</i>	<i>kədǎp</i>	—	* <i>gədǎp</i>
'bury'	* <i>tǎp</i>	<i>tǐp</i>	—	<i>tǐp</i>	<i>tǎp</i> ¹¹	* <i>tǎp</i>	<i>hətǎp</i>	* <i>tǎp</i>	* <i>tǎp</i>
'stump'	* <i>ʔəǎl</i>	—	—	—	—	* <i>təǎl</i>	<i>dəǎl</i>	* <i>əǎl</i>	*[ʔ]əǎl
'hold'	* <i>rǎŋ</i>	—	<i>drǎk</i>	—	—	—	—	* <i>rǎŋ</i>	* <i>rǎŋ</i>
'bitter'	* <i>ʔəhǎŋ</i>	—	—	—	—	—	—	* <i>hǎŋ</i>	* <i>hǎŋ</i>

The mid central vowels are absolutely straightforward. The only significant point to note is the complete merger of /*e*/ and /*ə*/ in NB—it looks like a continuation of the tendency that began with the merger of /*i*/ and /*ɨ*/.

PB **a*, **ǎ*

Gloss	PWB	Alak	Cua	Kass.	Tamp.	PSB	Bahnar	PNB	PB
'blood'	* <i>pəham</i>	<i>p'əŋham</i>	<i>phəhap</i>	<i>pham</i>	<i>pham</i>	* <i>nəham</i>	<i>pham</i>	* <i>pəham</i>	* <i>pəham</i>
'water'	* <i>dak</i>	<i>t'ak</i>	<i>dak</i>	<i>dak</i>	<i>t'iak</i>	* <i>dak</i>	<i>ʔdak</i>	* <i>dak</i>	* <i>ʔdak</i>
'night'	* <i>mǎŋ</i>	<i>mǎŋ</i>	—	<i>mǎŋ</i>	<i>mǎŋ</i>	* <i>mǎŋ</i>	<i>mǎŋ</i>	* <i>mǎŋ</i>	* <i>mǎŋ</i>
'eye'	* <i>măt</i>	* <i>măt</i>	<i>măt</i>	<i>măt</i>	<i>măt</i>	* <i>măt</i>	<i>măt</i>	* <i>măt</i>	* <i>măt</i>

Finally the low central vowels are frequent and readily reconstructed without any problems, and no explanation is required.

3.3 Sub-grouping proposal based on phonological reconstruction

The above reconstruction of Proto Bahnaric vocalism reveals a number of changes that can be invoked to support a classification. The most important of these is the unique retention of **i* in WB, while it has merged with **ɨ* elsewhere in Bahnaric. This immediately suggests a primary bifurcation opposing WB with the rest of the family—by implication PB split into two dialects, one of which underwent the change, and subsequently diversified into what is now the rest of the family. However, this idea runs into immediate difficulty, as the equally well established split of **sʔ* into /*s*/ in NB and /*t*/ elsewhere suggests a different branching. The two sets of changes are clearly phonologically unrelated, and it is impossible to show that one preceded the other. In fact, if one were to propose such a sequence, it raises the

¹⁰Sedang

¹¹'To plant'.

logical problem that one would then be forced to posit identical parallel developments! This is possible, but quite unlikely.

To deal with this problem we need a more sophisticated approach than the simplistic idea of proto-languages being synchronically undifferentiated—more than just a simple node in a branching tree. It is commonly recognised that languages have internal dialectal variation, and we should expect that a proto-language, as a real language, may also.

My proposal is that the sound changes at issue happened within a dialectally differentiated Proto Bahnaric, while it was still a linguistic unity. There are three distinct changes, and these can be represented on a Venn diagram, approximating a dialect map, as follows:

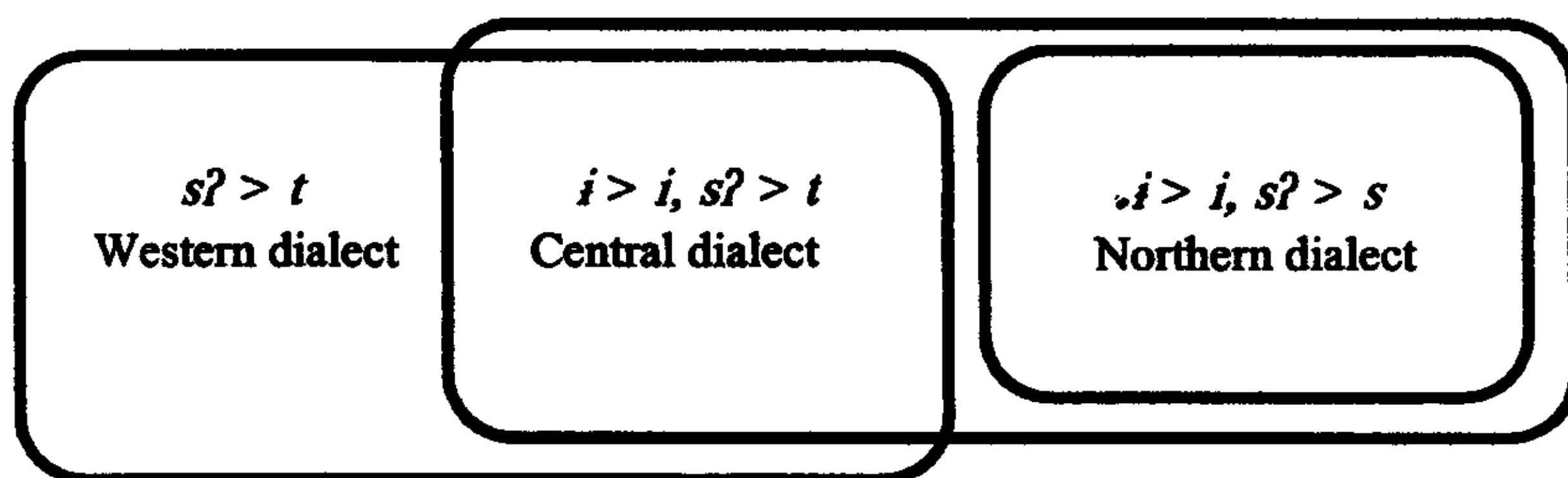


Figure 10. Proto Bahnaric dialect areas

Fig. 10 is read as indicating three distinct dialect areas, resulting from the overlapping spread of the sound changes *i > i* and *sʔ > t*. Each of these areas is the antecedent of the daughter languages that reflect those changes. Thus we have three coordinate branches: WB, NB and a third we can conveniently call CB (Central Bahnaric), and which can be thought of as an expanded version of GS&T's CB sub-group.

CB itself has somewhat complicated internal sub-grouping, which can be determined by further mapping of historical phonological changes I have reconstructed. This paper does not permit a full discussion of these—that requires a full book-length treatment. However, I present in Fig. 11 a condensed version of the new stammbaum, incorporating sufficient phonological detail to justify the sub-groupings:

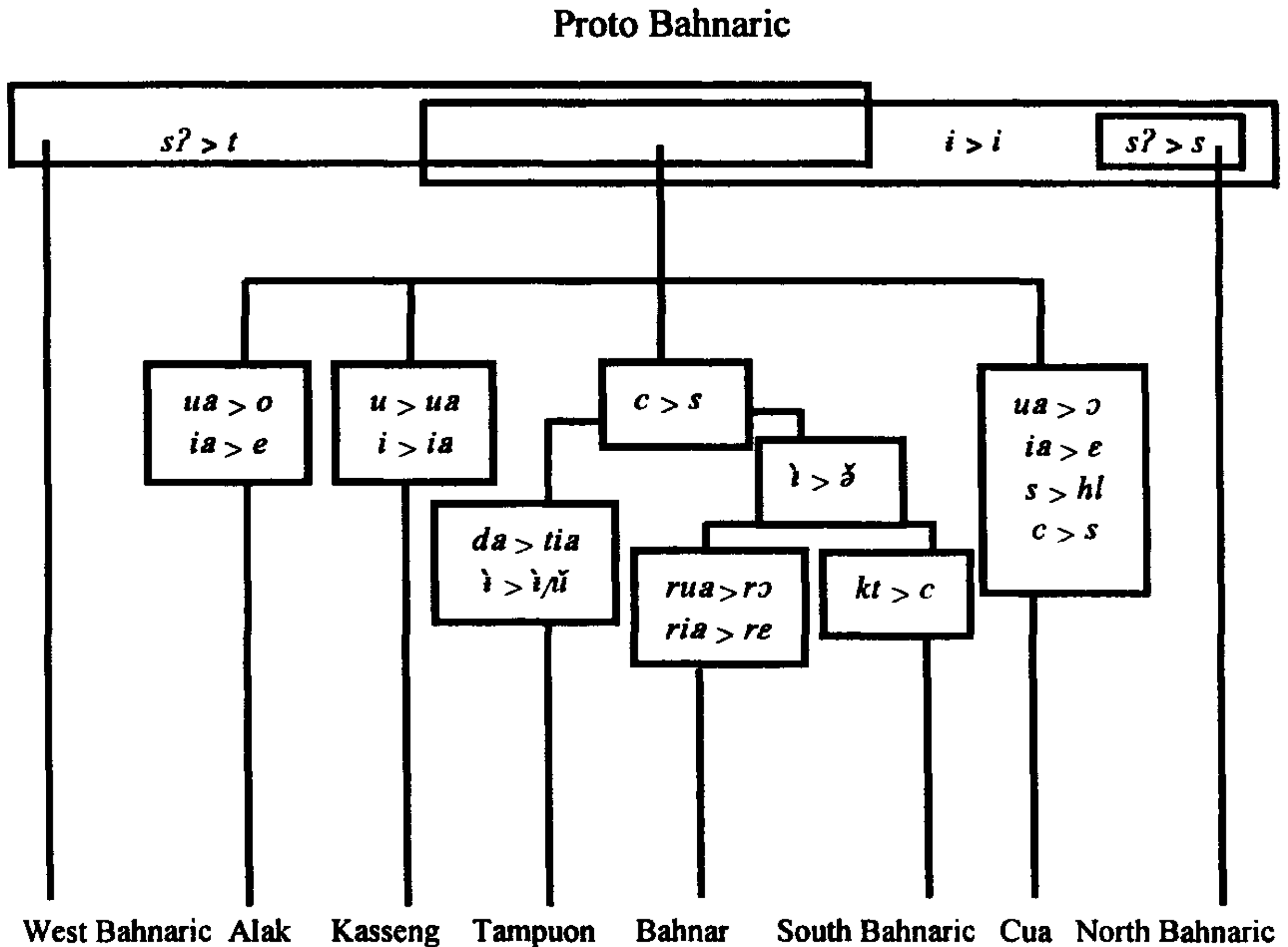


Figure 11. Revised Bahnaric Classification

The classification in Fig. 11 represents the state of my current working model. It is based entirely upon comparative phonological reconstruction. It is my strong conclusion that both 'distinctive vocabulary' and lexicostatistics are not reliable indicators of linguistic phylogeny, and should only be used as preliminary indicators in any comparative historical research. It is in the course of comprehensive comparative reconstruction that a justified internal classification can emerge.

The above model has various consequences for the reconstruction of Bahnaric (pre)history. Among them are those which flow from considerations of dialect geography. Considering the geographical distribution of the Bahnaric languages (refer to map, Fig. 2), it is evident that two of the three primary groups (WB and NB) are located within the northern range of the Bahnaric family, as are a majority of the CB languages. The southern range, largely separated from the north by Chamic languages, is as large or even a little larger than the northern range, yet it is occupied by a very small, genetically close sub-grouping, namely South Bahnaric. As I suggested in Sidwell (2000:3), SB appears to be "around a thousand years old", certainly it is not a very ancient grouping. Similarly, it has been suggested that the Chamic occupation/Chamisation of the highlands is also about a thousand years old (e.g. Thurgood 2000). These factors are strongly suggestive of a scenario whereby the Bahnaric communities were geographically localised approximately within the present northern range during the first Millennium, while the present Chamic and SB areas were occupied by an unknown people. Perhaps Khmeric or some other ethnic community associated with the Fu-Nan kingdom. These people withdrew westward from that area, as first Fu-Nan and then Chen La (Land) waned in favour of Angkor, facilitating the spread of Bahnaric and Chamic settlers over a vastly expanded range. Angkor

itself collapsed in the 1300s, facilitating the southern expansion of Thai/Lao peoples, and perhaps the spread into Cambodia of Katuic and West Bahnaric. I offer these speculations to stimulate further discussion, particularly with specialists beyond the narrow realm of comparative linguistics.

REFERENCES

- Campbell, Lyle. 1998. *Historical Linguistics: An Introduction*. Edinburgh, Edinburgh University Press.
- Diffloth, G. 1974. "Austro-Asiatic Languages." *EB* 15th edition 2:484.
- Diffloth, G. 1980. "The Wa Languages." *LTBA* 5.2 Berkeley: University of California.
- Ferlus, Michel. 1974. "Delimitation des groupes linguistiques austroasiatiques dans le centre indochinois." *ASEMI* 5.1:15-23.
- Gregerson, Kenneth, Kenneth Smith & David Thomas. 1976. *The Place of Bahnar within Bahnaric*. In Jenner et al. (1976a), pp 371-406.
- Jacq, Pascale and Paul Sidwell. 2000. *A Comparative West Bahnaric Dictionary*. München, Lincom Europa.
- Jenner, Philip N., Laurence Thompson and Stanley Starosta (eds.). 1976a. *AS* 1. Honolulu: University of Hawaii Press.
- Jenner, Philip N., Laurence Thompson and Stanley Starosta (eds.). 1976b. *AS* 2. Honolulu: University of Hawaii Press.
- Pinnow, Heinz-Jürgen. 1959. *Versuch Einer Historischen Lautlehre Der Kharia-Sprache*. Wiesbaden, Otto Harrassowitz.
- Shorto, Harry L., Judith M. Jacob, and E.H.S. Simmonds. 1963. *Bibliographies of Mon-Khmer and Tai linguistics* (London Oriental bibliographies 12). London, Oxford University Press.
- Sidwell, Paul. 2000. *Proto South Bahnaric*. Canberra, Pacific Linguistics.
- Smith, Kenneth, D. 1967. *A Phonological Reconstruction of Proto Central North Bahnaric*. University of North Dakota Summer Institute of Linguistics Workpapers.
- Smith, Kenneth. 1972. *A Phonological Reconstruction of Proto-North-Bahnaric*. Summer Institute of Linguistics, Language Data Series, Dallas.
- Smith, Kenneth. 1979. *Sedang Grammar*. Pacific Linguistics. Series B-No. 50.
- Sokolovskaja, N.K. 1978. "Materialy k sravnitel'no-ëtimologicheskomu slovar'u v'etmyongskix jazykov." *Issledovanija po fonologii i grammatike vostochnyx jazykov*, Moskva, Nauka, pp126-80.
- Thomas David. 1979. "The Place of Alak, Tampuon, and West Bahnaric." *MKS* 8:171-186.
- Thomas, David & Robert Headley. 1970. "More on Mon-Khmer subgroupings." *Lingua* 25:398-418.
- Thomas, David. 1966. Mon-Khmer subgroupings in Vietnam. In Zide ed., 1966, pp 194-202.
- Therapan L. Thongkum. 1997a. "The place of Lawi, Harak, and Tariang within Bahnaric." *MKS* 27: 109-117.
- Therapan L. Thongkum. 1997b. "Suai, a newly discovered West Bahnaric language of Southern Laos." Paper presented at the *Thirtieth International Conference on Sino-Tibetan Languages and Linguistics*, Beijing, August 24th-28th, 1997. Published in Chulalongkorn University 80th Anniversary Research Conference Proceedings. Chulalongkorn University, Oct. 15-17, 1997.

- Therapan L. Thongkum. ms. "A brief look at the 13 Mon-Khmer languages of Xe Kong province of southern Laos". In Robert Bauer ed., *Collected Papers on Southeast Asian and Pacific languages*. Canberra, Pacific Linguistics.
- Thurgood, Graham. 2000. *From Ancient Cham to Modern Dialects: 2000 Years of Language Contact and Change*. University of Hawaii Press.
- Trask, R. Larry. 1996. *Historical Linguistics*. London, Arnold.
- Zide, Norman H. (ed.) 1966. *Studies in Comparative Austroasiatic Linguistics*. London/The Hague/Paris, Mouton.

Received: 20 June 2001

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