

Three Austroasiatic branches and the ASJP

Paul Sidwell

Center for Research in Computational Linguistics, Bangkok

Visiting Research Fellow, Australian National University

paulsidwell@yahoo.com

Daft at 5/9/10

Prepared for the First Conference on ASJP and Language Prehistory (ALP-I)

Max Planck Institute for Evolutionary Anthropology (Leipzig), September 17-19, 2010

1. Introduction

Three reasonably complex 3 AA branches - Bahnaric, Katuic, and Palaungic - have been subject to comparative reconstruction to the extent that a reasonably confident account of their internal sub-branching can be discussed. Those results are compared to the most recent output of the ASJP World Language Tree of Lexical Similarity, and the similarities and differences analysed. It is evident that where the Automated Similarity Judgment method generates misleading results the causes can relate to areal factors; e.g. lexical divergence due to borrowing and/or sound changes. In some other cases languages have been restructured by sound changes that may be purely local in character or related to areal pressures. Additionally, some ASJP results are perhaps skewed by the selection of languages, which is necessarily incomplete. Finally, some observations will be offered in respect of other branches of the AA and the current ASJP tree.

2. Bahnaric

2.1 Expert classification of Bahnaric

The Bahnaric branch, with hundreds of thousands of speakers in Vietnam, Cambodia and Laos, is a grouping of tremendous internal diversity, so much so that, along with Aslian, it was not an effective consensus on its unity until around the time of the 1973 Austroasiatic conference (University of Hawaii). Even today, I am having second thoughts on the matter, seriously considering whether it might be better modelled as a convergence area than a genetic unit. It is actually difficult to compile a set of innovation that characterise Bahnaric as a single division. The historical phonology reconstructs to approximate pAustroasiatic, and in stunning contrast to neighbouring branches, such as Katuic, Vietic, Pearic etc. one cannot readily compile a significant lists of lexical innovations unique to Bahnaric. The lexical innovations that do readily identify that division are innovative forms of two Austroasiatic words, idiosyncratic reflexes of the roots for ‘bone’ and ‘fire’, and in fact the latter is not found in a couple of Bahnaric languages on the geographic periphery.¹ But these concerns aside, the orthodoxy of Bahnaric unity, established by Thomas (1966) and Thomas & Headley (1970), and endorsed by Diffloth (1974 and *passim*.) is adopted for the purposes of this paper.

Until the 1960’s there was no coherent identification of a Bahnaric branch 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. For example, Pinnow (1959) effectively listed eight different languages, which we would today recognise as Bahnaric, as members various co-ordinate level MK groupings. Some of those languages were even

¹ pAA *jʔaaŋ ‘bone’ appears to be reflected imperfectly in pBahnaric *ktsiŋ, and pAA *ʔus ‘fire’ in pBahnaric *ʔuŋ. But notably, Sre and Cua regularly reflect *ʔus rather than the distinctively Bahnaric form.

specifically grouped with others that are now recognised as belonging to different branches (e.g. with Katuic and Khmer). Shorto, Jacob and Simmonds (1963) simplified and advanced Pinnow’s scheme, removing some of the most striking anomalies, yet they still placed the Bahnaric languages into six distinct groups.

It was shortly after that various lexicostatistical studies by David Thomas and Robert Headley came to more properly characterise the linguistic situation. In respect of Bahnaric languages they distinguished two major divisions—Stiengan and Bahnaran—later renaming them South Bahnaric and North Bahnaric respectively. Thomas and Headley did not treat the West Bahnaric languages lexicostatistically for lack of data, but did anticipate the grouping, effectively rectifying the scheme of Shorto, Jacob and Simmonds (1963).

Though the 1970s, various studies (e.g.: Smith 1973, Gregerson, Smith and Thomas 1976, Thomas 1979) came to crystallize an expanded model of five coordinate divisions, accommodating newly identified divisions Central Bahnaric (CB) and East(ern) Bahnaric (the latter treated by Smith 1973) as a sister of North-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?

North Bahnaric	<i>Sedang, Hrê, 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 1: Classification of Bahnaric languages by Thomas (1979)

To my knowledge, the relevant discussions until recently have emphasised results based upon ‘distinctive vocabulary’. So far as I can tell, the procedure is that, a preliminary classification is achieved by traditional lexicostatistics, or self evident sub-groupings are identified idiosyncratically 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’.

In the 1990s, apparently beginning with a presentation to the *24th International Conference on Sino-Tibetan languages and Linguistics*, Diffloth began presenting a more elaborate nested-branching model of Bahnaric, diagrammed in the Figure XX. This classification resumes the earlier notion of a primary north-south division, the former including North and West Bahnaric, and the latter South and Central Bahnaric. Significantly East Bahnaric has disappeared, but to where is not indicated.

This scheme was later championed by L-Thungkum (1997, 2001, 2002), including an elaboration of a North-West Bahnaric as a sister of West Bahnaric. The basis of Diffloth's classification has not been made explicit, but Diffloth (personal communication) has asserted that lexical innovations are taken into account.

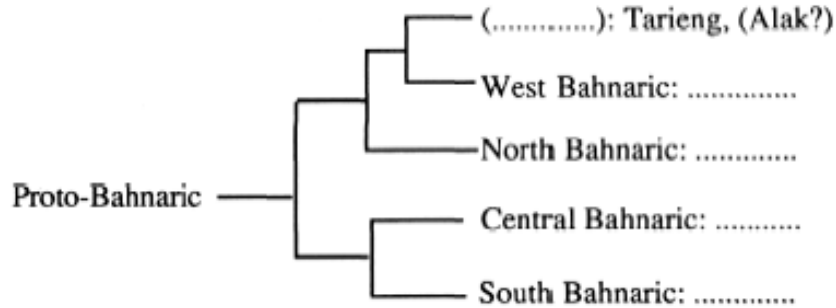


Figure 2: Classification of Bahnaric languages by Diffloth (1991)

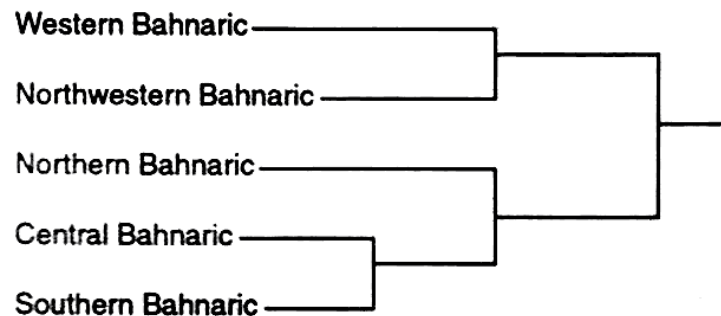


Figure 3: Classification of Bahnaric languages by Diffloth (2005)

Diffloth (2005) modified this classification (see Figure XX), so that the initial binary division is between West and Northwest Bahnaric and a diverse branch including South, Central and North Bahnaric. Again, no explicit justification has been offered.

Recently, this writer (e.g. Sidwell 2002, 2009, 2010) has been arguing for the classification of Bahnaric based on historical phonology, eschewing distinctive lexicon as far too problematic to rely on for classification. The phonological reconstruction associates distinctive innovations with four co-ordinate divisions, such that the most parsimonious account posits each emerging out of a dialect chain assumed to be equivalent to pBahnaric. This model actually correlates well with Tomas (1979), the main difference being that Sidwell treats South Bahnaric as a sub-branch of Central Bahnaric.

West Bahnaric	Central Bahnaric	North Bahnaric	East Bahnaric
Jru' (Laven), Juk, Su'	Taliang (Kasseng)	Halang, Kayong	Cua (Kor)
Nyaheun	Alak	Jeh	
Oi, The, Sok, Sapuan, Cheng	Central South	Kotau	
Brao, Laveh, Krung, Kravet	Tampuon	Tadrah, Modrah	
	Bahnar	Sedang	
	South Bahnaric	Hrê	
	Chrau	Monom (Bonâm)	
	Sre	Rengao	
	Stieng	Kaco', Ramam	
	Mnong		

Figure 4: Classification of Bahnaric languages by Sidwell (2009)

In summary, this writer takes it as significant that there is substantial coherence between Thomas' distinctive vocabulary classification, and Sidwell's historico-phonological classification. The main discrepancy relates to the status of South Bahnaric, which Thomas treated as a coordinate division, rather than a sub-division of Central Bahnaric. Actually Thomas did not test this hypothesis, but effectively continued it as a default from earlier lexicostatistical studies, so this writer is confident that his historico-phonological scheme is more or less correct.

2.2 ASJP Bahnaric classification

The ASJP Bahnaric classification (Figure XX) presents very mixed results when compared to the expert classification. The tree has a remarkable structure that does not correspond to contemporary or earlier expert classifications.

First of all, the ASJP tree does succeed in some respects. Two groupings are unambiguously detected: West Bahnaric and South Bahnaric. Cheng, Oi, Brao, Sapuan, Jru, Loven, and Nyaheun, all West Bahnaric, are correctly grouped together. Furthermore, within this division, Cheng, Oi, Brao and Sapuan are correctly grouped in opposition to Jru/Loven. The former are spoken in Sekong River valley while the latter is spoken on the Boloven Plateau. Significantly, Nyaheun is put on an outlying branch. Based on the criteria of phonetic similarity, this is understandable, since Nyaheun has undergone dramatic phonological restructuring (see Ferlus 1971, Sidwell & Jacq 2003) which radically reduces its gross similarity to other West Bahnaric languages. From a genetic viewpoint however, Nyaheun should be more properly grouped with the languages of the Sekong river valley, since its restructuring represents an extreme realization of processes already manifest in those languages, and not shared with Jru/Loven.

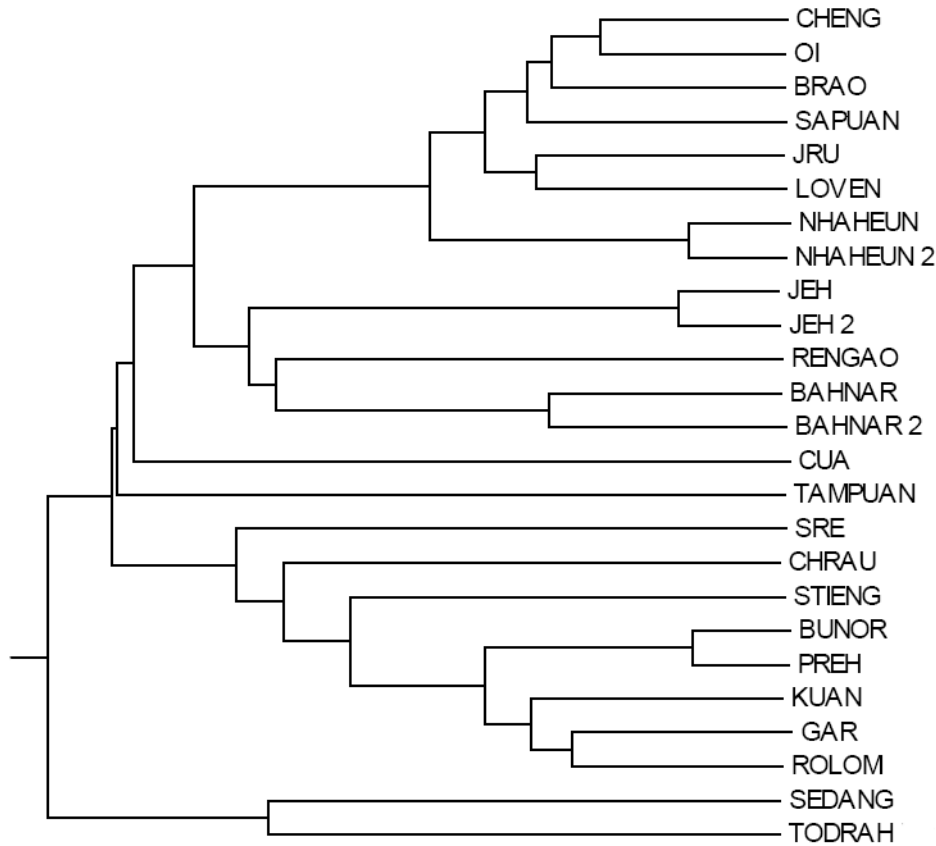


Figure 5: Classification of Bahnaric languages by ASJP (2010)

Regarding South Bahnaric, ASJP correctly groups Sre, Chrau, Bunor, Preh, Kuan, Gar, and Rolom. The last five in that list are commonly recognised as dialects of Mnong, and ASJP correctly puts them into one division. The rest of the ASJP south Bahnaric tree is consistent with what we know about the languages – the real genetic relations may be somewhat different, but it is difficult to determine with confidence because the group resembles a dialect chain rather than a clearly branching tree.

Beyond correctly identifying West and South Bahnaric, the ASJP fails badly in respect of solidly demonstrated genetic groups. Most strikingly, North Bahnaric is not recognised at all. The North Bahnaric division is identified by scholars by the fact that all of its members show voice quality contrasts of a tense-lax type, which does not relate to historical initial consonant voicing, but instead relates to proto-vowel timbre differences (see Sidwell 1999). This remarkable shared paradigmaticity is the strongest type of sub-grouping evidence, yet ASJP results appear to be blind to it. Relevant languages in the study are Jeh, Rengao, Sedang and Todrah. The last two of these are correctly grouped, and placed on an outlier branch with respect to the rest of the Bahnaric languages. But, Rengao and Jeh are grouped closely with Bahnar, and the three of those are curiously branching from the same node as West Bahnaric. Rengao and Bahnar are close neighbours who typically intermarry, so there are a lot of loans, and we can expect this to influence automatic similarity judgment. Also, Rengao is closely related to Jeh, so their proximity in the ASJP tree is consistent with this fact. However, from a genetic point of view Rengao is probably marginally closer to Sedang and Todrah than to Jeh,

so it is somewhat surprising not to see Sedang and Todrah branching between Rengao and Jeh.

Generally any lexical method will produce confusing result for Bahnar, and this fact alone accounted for several papers in the 1970s (e.g. Gregerson, Smith and Thomas 1976, Thomas 1979). This is because of the geographical and social position of the language, especially its presence around city of Kontum. This fact probably accounts for much of the discrepancies discussed immediately above, but there are other problematic aspect to the ASJP tree. It is not at all clear why West Bahnaric should appear grouped with Bahnar, since neither contact nor lexical similarity would normally indicate such a relation. Also, it is odd that Cua branches from the first node above this Bahnar-West Bahnaric clade. Cua is lexically and phonologically very different from the rest of Bahnaric, and it is clear (see Sidwell 2009, 2010) that it is either a primary branch of Bahnaric or an aberrant sister of North Bahnaric. Above this node is Tampuon, which in genetic terms is a sister of Bahnar, but which has been under such profound Khmer influence that it has developed a very distinctive character of its own. Above this node is the branching of South Bahnaric. Although this is more consistent with early Bahnaric classifications, which were lexical and impressionistic, it is at odds with the evidence of historical phonology, which groups it more tightly with Tampuon and Bahnar, a fact that is not revealed by ASJP.

On balance, it is apparent that while ASJP does detect a couple of well attested and rather close subgroupings, overall it fails to produce anything resembling a useful genetic tree of the Bahnaric family. Some of the faulty branchings can be readily explained as reflecting real similarity due to geographical and social proximity, but other aspects appear to run counter to what I might expect according to these trends. At this stage it is not at all clear what ASJP is detecting beyond only the most obvious, and diachronically recent, genetic groupings.

3. Katuic

3.1 Expert classification of Katuic

The Katuic languages are spoken by several hundred thousand people in Thailand, Cambodia, Laos and Vietnam; and although the majority of Katuic speakers live in eastern Thailand and Cambodia, the greatest diversity of Katuic languages lies in the Salavan and Sekong provinces of Laos and adjacent border areas of Vietnam, part of a complex patchwork of small ethnic communities. This geography, combined with the difficult political history of the region, means that only recently have scholars known the extent of the family, and consequently only the most recent comparative researches have begun to adequately address the issue of classification.

The existence of the Katuic languages first came to the attention of western scholars with the various short wordlists collected by the Pavie expeditions to Indo-China in the 1800s; hence the inclusion by Schmidt (1906) of Sue and Nahhang (both Kui dialects) in his Mon-Khmer group. However, early comparative studies did not systematically classify the minority languages of Annamite Range, effectively treating them as a Mon-Khmer continuum, so until the 1960s there was no notion of distinct Katuic branch.

A turning point was the lexicostatistical study of Thomas (1966), which distinguished both Katuic and Bahnaric branches among the language of the Annamite Range. In fact SIL scholars in Vietnam had begun to recognize the distinct grouping we now call Katuic as early

as 1961². The lexicostatistics suggested a simple north/south bifurcation into sub-branches Thomas called “Brôuan” and “Katuan”.

Katuic
A. Brôuan
1. Brôu (Bru)
2. Pacoh
3. Ta-oih
B. Katuan
1. Katu
2. High Katu
3. Phương

Figure 6: *Katuic classification by Thomas (1966)*

Subsequent fieldwork quickly fleshed out the extent of the Katuic languages and in 1970 Thomas and Headley named 15 distinct languages (although their list also included, erroneously, two Bahnaric languages). Katuic was, from the late 1960s onward, subjected to perhaps more comparative-historical studies than any other Austroasiatic branch, and a remarkable number of extensive reconstructions were published and/or drafted (e.g. Thomas 1976, Diffloth 1982, Gainey 1985, Efimov 1983, Shorto ms., Peiros 1996, L-Thongkum 2001, Sidwell 2005).

Somewhat remarkably, until the most recent reconstruction was published in 2005, all but one the previous historical studies (Gainey 1985) appeared to eschew classification based on historical phonology, and instead appear to have accepted lexical and lexicostatistical classifications in the first place, and interpreted the phonological correspondences assuming the priority of lexical methods. Consider, for example, the following explicit assertion by Peiros: “The phonological correspondences do not provide any information that helps in classifying the languages, on the basis of phonological correspondences, we can treat the languages as four independent branches of the family” (1996. v). The exception, the study by Gainey, only concerned relations within the West Katuic sub-branch, so they held no implications for the Katuic branch as a whole.

Broadly speaking, the earlier lexically based classifications consistently agreed in dividing Katuic into two main branches, a West Katuic including Kui, So and Bru, and an East Katuic including Katu, Pacoh, Taioh, Ngeq and others. This is evident in Thomas’ (1966) classification (Figure XX, above) and Ferlus’ (1974) reproduced immediately below.

² For example, Thomas (1964, 162) refers to an unpublished survey and lexicostatistical study of Vietnam languages from 1961; while Thomas (1966, 195), discussing the lexicostatistically indicated division between Katuic and Bahnaric remarks: “The existence of this break was suggested previously by Phillips”. Possibly a reference to the same unpublished work he alluded to in 1964.

Miller & Miller (1996), analyzing lexical data from 50 languages, proposed seven sub-groupings without configuring them into a tree (Figure .

Katuic
North Katuic: So, Bru, Tri, Makong/Mangkong, Siliq, Katang
West Katuic: Sui/Suoi/Suai, Nheu, Kui, Kuay
Pacoh
Central Katuic: Ong, Ir
Ngeq
Katu (Laos)
Katu (Vietnam)

Figure 9: *Katuic sub-groups by Miller & Miller (1996)*

In the same year Peiros offered a lexicostatistical classification that echoed Smith's (1981) result, placing Katu in a coordinate branch above the rest of the family (Figure xx).

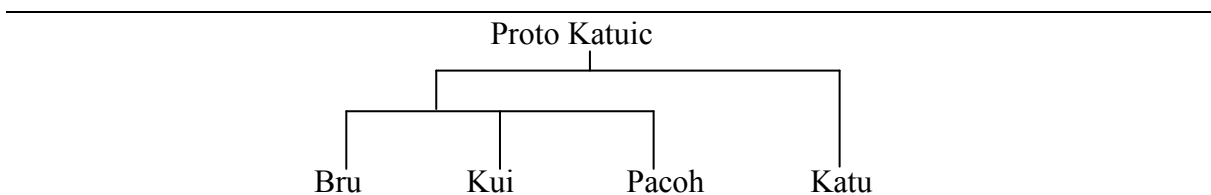


Figure 10: *Katuic classification by Peiros (1996)*

Based upon her own fieldwork in the Lao PDR, L-Thongkum (2001) presented an extensive Katuic comparative lexicon. She analyzed sub-groupings based primarily upon select lexical isoglosses, and offered the classification given here at Figure .

Katuic
West Katuic: Kui, Souei
East Katuic:
North East Katuic: Bru, So, Pacoh
Central East Katuic: Ta'Oi, Chatong, Kriang
South East Katuic: Dakkang, Triw, Kantu, Katu

Figure 11: *Katuic classification by L-Thongkum (2001)*

The curious fact here is that L-Thongkum's results are a kind of throwback to the late 1960s, except that she takes Bru and So out of West Katuic, and instead groups them with Pacoh. My own analysis of this scheme is that it actually reflects the geographical proximity of Bru and Pacoh, which otherwise show radically different historical phonologies, something glossed over by L-Thongkum.

This writer (Sidwell 2005) presented a phonological reconstruction of Proto-Katuic, including a classification based solely upon historical phonology. The results sub-group the languages into four equidistant clades as indicated at Figure . The analysis argues that Katu is

phonologically conservative, while the other three sub-branches underwent devoicing and register formation independently. The classification is laid out at Figure xx.

Katuic
West Katuic:
Kui, Souei
Bru, Sô
Ta'oih:
Ta'oih, Katang, Talan/Ong/Ir/Inh
Kriang/Ngeq
Chatong
Katu:
Kantu, Katu, Phuong, Triw, Dakkang
Pacoh:
Pacoh

Figure 12: *Katuic classification by Sidwell (2005)*

Actually, it is evident that all studies have identified more or less the same four divisions, rather the differences have been in whether these form nested branching relations, and the results have varied according to which lexical items have been considered. Therefore, on balance, I would argue that my four-equidistant branch model, derived at without reference to isoglosses, is likely to be the more reliable, and it is against this that the ASJP results should be tested.

3.2 ASJP classification of Katuic

The ASJP tree for Katuic is reproduced at Figure xx. It only considers nine sources, although they do include at least one language from each of the four commonly accepted Katuic divisions.

Broadly speaking, the ASJP results are a nice match to the earlier lexicostatistical classifications, especially for example, Thomas (1966) and the largely derivative classification by Ferlus (1974a). Bu and Kui are grouped in one branch, while Katu, Nge and Pacoh are grouped into another. This is more or less what we would expect to see produced by lexical comparison conducted with a modest wordlist, which is a rough categorization of the ASJP method. At this stage it is not especially interesting. I would expect that an analysis with a much larger set of languages would be more useful/indicative.

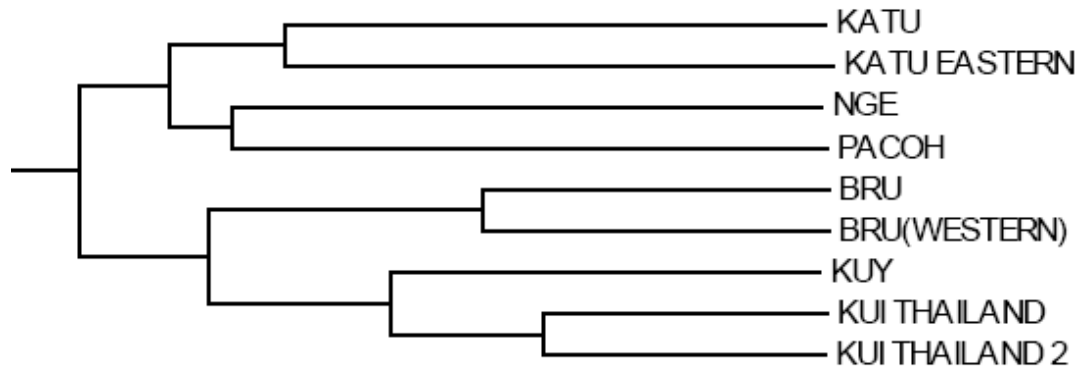


Figure 13: ASJP classification of Katuic (2010)

4. Palaungic

4.1 Expert classification of Palaungic

The Palaungic branch is widely dispersed among small communities living in the Shan State and about the borderlands of Myanmar, Laos and China. Some of the languages are well known to western scholarship thanks to descriptive grammars and dictionaries, although many of languages are only briefly studied and/or documented in Chinese publications not widely read abroad. The term *Palaung* itself is a Burmese cover-term for speakers of a group of dialects who self identify with terms such as Ta-ang, Daraang and similar.

The group has been known to Western scholarship for over a century, the first comparative study being by Schmidt (1904), who used the term Palaung–Wa (also used in the contemporaneous *Linguistic Survey of India*). Based upon comparison of numerals and other basic lexicon he grouped the languages into four divisions as listed in Figure :

I. Palaung
Palaung or Rumai of Nam Hsan
Palaung or Rumai (Shan State)
Rumai (Manton neighbourhood)
Palang Kengtung State, call themselves Darāng
II. Wa
Wa or Vü
En Tribe Kengtung State
Sön Kengtung State
Tai-loi Wa or Wa küt Kengtung State
Hsen-Hsum, call themselves Amok, Kengtung State
III. Riang or Yang Sek
IV. Danaw

Figure 14: Palaung-Wa classification by Schmidt (1904)

Half a century later Shafer attempted a modest comparative analysis of Palaungic (treating 88 etymologies), and with the basis of additional data distinguished of an Angkou (Angkuic) division (Figure xx).

Palangique
Riang
Palaung
Nam Hsan
États Chans
Manton
Darang
Angkou
Amok
Monglwé
Angkou
Wa
Vu (Wa)
Èn
Tailoï
Wa (de Kèngtoug)
Son
Danaw

Figure 15: *Palaungic classification by Shafer (1952, 112)*

Pinnow (1959) posited a Northeast Austroasiatic sub-family comprising two branches, a *West-Untergruppe* (West Sub-group) corresponding to Palaungic and *Ost-Untergruppe* (Eastern Sub-group) corresponding to Khmuic. This West Sub-group was further divided into six clades (see Figure). The one structural difference with Shafer's scheme was the addition of Lawa as a sixth division, although later studies (e.g. Diffloth 1980) group Lawa with Wa. It is significant to the present discussion that Palaungic and Khmuic were grouped by Pinnow, and this supposed connection has been carried through subsequent scholarship, for example in Diffloth's (2005) classification (see Figure xx).

-
1. Riang (Yang Sek, Yang Wan Kun)
 2. Palaung (Palaung, Rumi) with dialects: Nam Hsan, shan States, Mantôn and Darang (Milne distinguished dialects Rumi, Tiorai, Wah, Kyusao, Kumkaw, Hupawŋ, Omachawn, Kwawnhai, Ho-mau, Paŋnim, Kwanwantok)
 3. Anku with dialects: Anku, Amok (Hsen-Hsum), Monglwé (Tailoi of Mông-Lwe)
 4. Wa with dialects: Vü (Vu, Wa-Vü, Wa), Èn, Tailoi (Tai-Loi, Wa-Küt), Wa (Kèntung), Son
 5. Danaw
 6. Lawa (La'wa) with dialects: Umpai, Bo Luang (Bo Luang), Mapɛ (Mapä), Pa-Pao
-

Figure 16: *Pinnow's West-Untergruppe (=Palaungic) (1959)*

By the mid-late 1970s comparative work on Palaungic was renewing (e.g. Mitani 1977, 1979; Diffloth 1977, 1979, 1980) was blossoming again, and a classification largely based on historical-phonology was presented by Diffloth (1977, see Fig. xx).

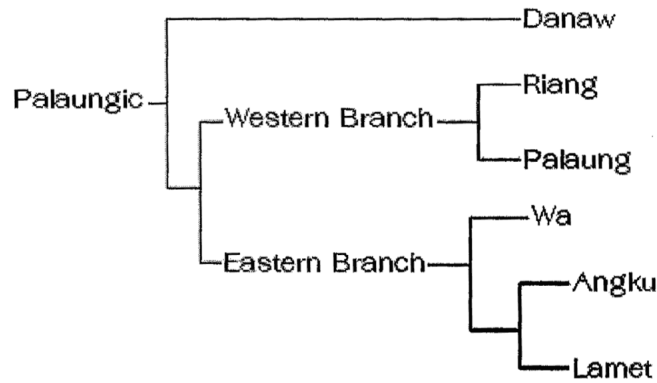


Figure 18: *Palaungic classification by Mitani (1978)*

In the late 1970s Mitani was also working on the history of Palaung, and in a series of papers and conference presentations (1977, 1978 CHECK) outlined a reconstruction of the Palaung-Riang sub-branch. In a 1978 manuscript he offered the classification reproduced here as Figure xx. Significantly, like Diffloth (1977) Danaw is the highest branching node, but below it the remainder of the languages are divided into East and West sub-branches, the former is Palaung-Riang and the latter is Wa-Angku-Lamet.

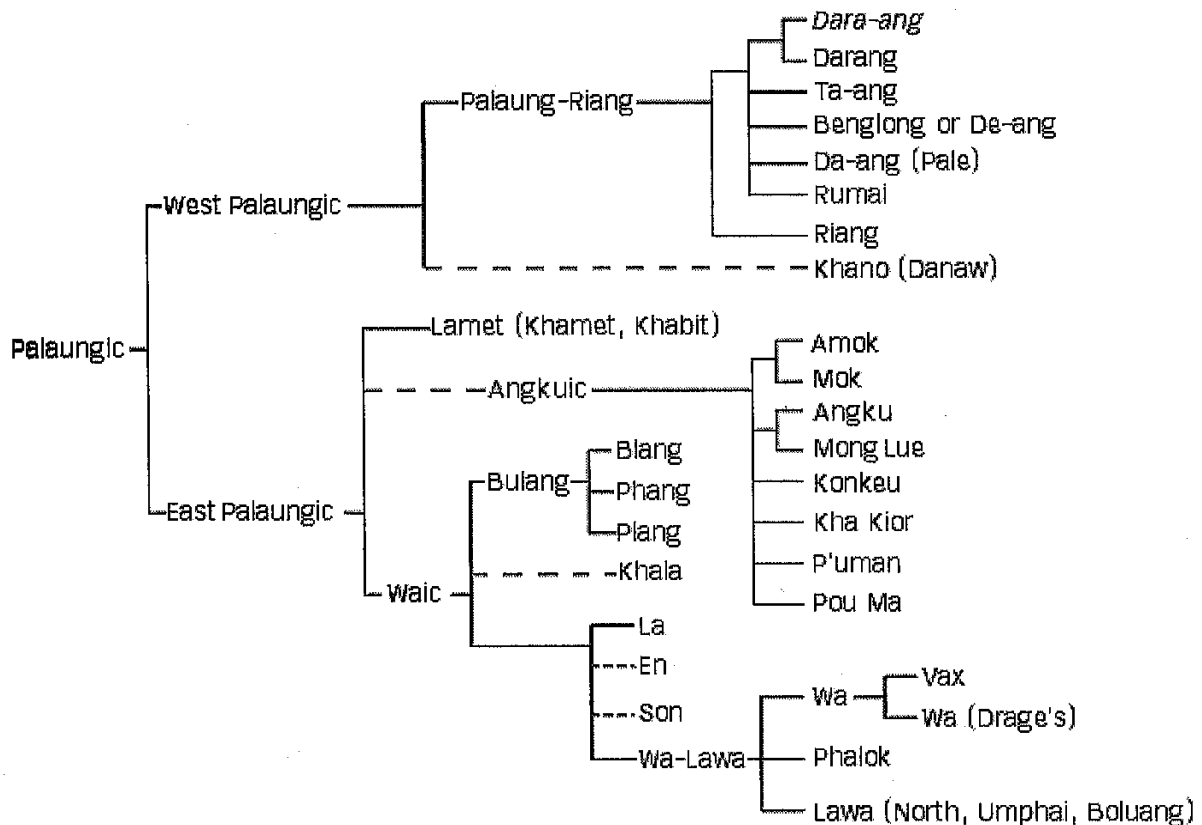


Figure 19: *Palaungic classification of Diffloth (1982), reproduced from Deepadung (2009).*

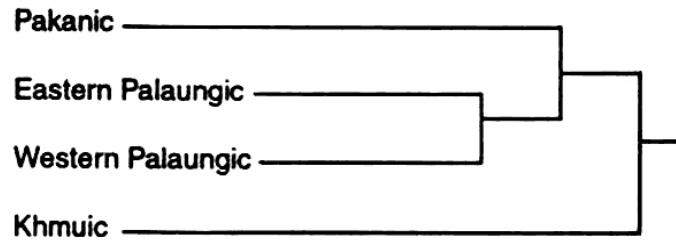


Figure 21: Palaungic classification from Diffloth (2005)

From 1982 onward Diffloth began advocating a new Palaungic classification, one that has strong elements in common with Mitani (1978), and which he has continued to present up to the present (e.g. it is clearly implied in Diffloth 2005).

The Diffloth (1982 and onward) classification suggests a binary division between East and West Palaungic, the former accommodating Palaung-Riang and Danaw, and the latter consists of three equidistant branches Lamet, Angkuic, Wa-Lawa. The Wa-Lawa component of the tree is elaborated and justified by Diffloth (1980), but the remainder of the tree is not explained, and it is not clear to this writer how it could be reconciled with the historico-phonological model presented in 1977, rather it looks like an attempt to approximate Mitani's classification.

Recently this writer prepared a reconstruction of proto-Palaungic, based on a compilation of 1350 cognate sets, now available online at <http://sealang.net/monkhmer>. The phonological reconstruction appears to confirm that Danaw does constitute a coordinate branch, based on the independent nature of the chain shift among the palatals, largely consistent with Diffloth (1977). However, regard the developments among the initial consonants; it appears to be impossible to suggest any sequence of nested branching among the other Palaungic sub-groups. Figure xx diagrams the various developments among initials, showing that all are either variously overlapping, or isolated, suggesting that the four sub-branches – Palaung-Riang, Waic, Lamet, and Angkuic, reflect four equidistant clades. The reconstruction of vocalism also fails to suggest any nested branching; rather each group appears to have variously and independently reduced vowel length contrasts and/or developed breathy registers.

Danaw	Palaung-Riang	Waic	Lamet	Hu	U
*s > θ	*s > h		*s > θ		
*c > ts	*c > s				
*j > dʒ					θ > s
		*b, *d > b, d	p, t, k > ph, th, kh		
low series stops > -voice	low series stops > -voice				
			*b, *d > p, t		

Figure 22: Palaungic initial consonant developments (first presented by Sidwell at Southeast Asian Linguistics Society meeting, Zurich, June 2010).

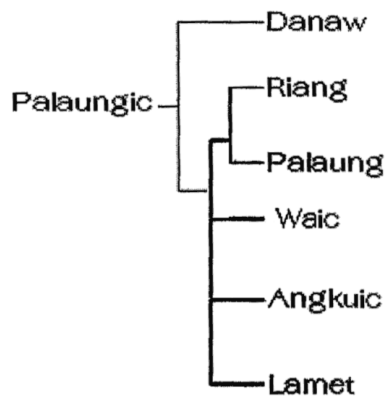


Figure 23: Palaungic classification suggested by Sidwell (ms.2010).

4.2 ASJP classification of Palaungic

The ASJP classification of Palaungic is reproduced below at Figure xx, with Khmuic languages blocked in yellow. The first thing to point out about it is that the ASJP effectively imbeds the Khmuic branch within Palaungic. Actually, ASJP fails to indentify Khmuic as a distinct clade, instead it places Khmu/Ksingmul/Bit onto a branching node immediately above the Palaungic languages minus Danaw, and then puts Mal/Mlabri on a branch one node higher again. Above them all Danaw occupies a primary branch.

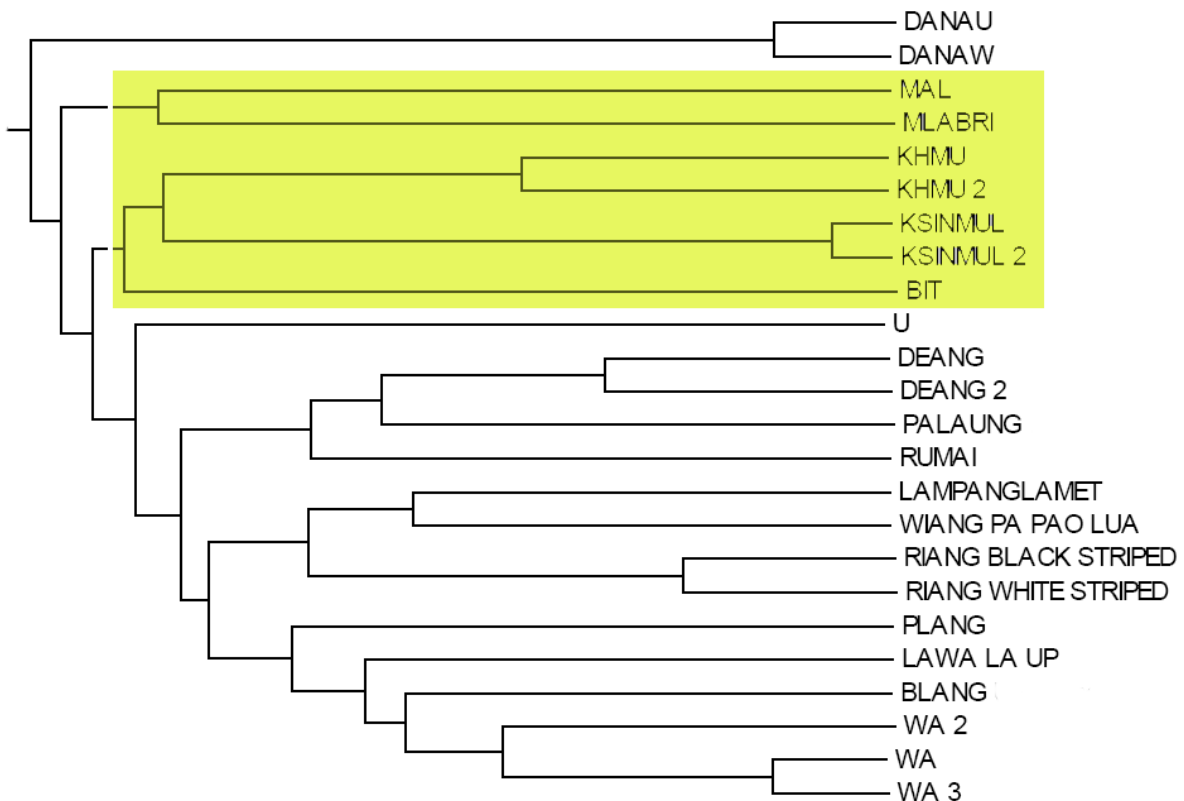


Figure 24: ASJP Palaungic classification.

In some respects it is not surprising that there is such a result for Khmuic. Khmuic is readily identified on historico-phonological criteria, showing a loss of medial *h, e.g. consider the word for ‘blood’: Khmu Chuang /ma:m/, Kammu-Yuan /mà:m/, Odu /mim/, Theen /meem/, compare with Bahnar /pha:m/, Mon /chim/, Semelai /maham/. Yet the Khmuic languages share a remarkably low percentage of vocabulary in common, more proportionate to separate AA branches rather than members of one branch.⁴ Consequently, a classification based upon a small vocabulary sample, such as the ASJP, may be expected to produce ambiguous results.

Additionally, the status of at least one language, Bit, is problematic. Conventionally classified as Khmuic (e.g. Ethnologue, Peiros 2004) this is far from determinative. The language is known to scholars only from rather sketchy sources that do not offer good documentation of the basic lexicon. The data we do have gives contradictory indications, e.g. the Bit word for ‘blood’ recorded by Kingsada & **Shintani** (1999) is /s(ə)nuǎm/, which looks like a regular reflex of the Palaungic *snaam, rather than Khmuic *maam.

Specific examples such as Bit, and the generally highly heterogeneous lexicon of Khmuic languages indicates that significant borrowing is likely to have occurred. This may underlie the tradition among scholars, going back at least to Pinnow (1959), of grouping Palaungic and Khmuic into a Northern Mon-Khmer clade.

Ignoring the Khmuic languages, the ASJP Palaungic classification reduces structurally to the tree at Figure xx. Rather remarkably, this corresponds neatly to Diffloth’s 1977 classification, except that ASJP erroneously groups Riang with Lamet, when all relevant scholars agree that Riang belongs with Palaung (Da’ang etc.) in a Palaung-Riang clade.

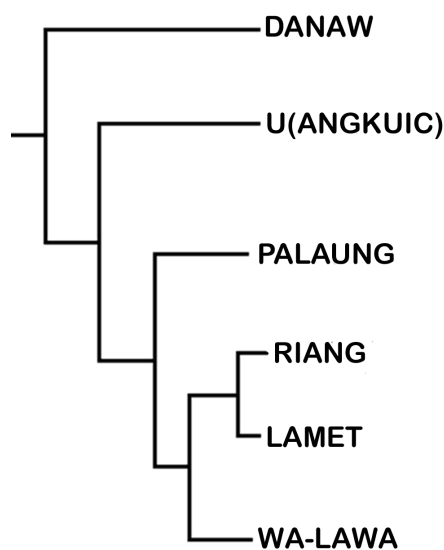


Figure 25: *ASJP Palaungic classification simplified.*

It is very interesting that the ASJP tree corresponds fairly well to the original Diffloth tree, but it is surely more significant that it generates obviously wrong result at both low and high

⁴ By my count, Mlabri and Ksingmul share only 21% on the Swadesh 100 list, while on average (calculated across 54 language) Khmuic languages share 19.25% with other AA branches.

levels within the tree. Danaw is unambiguously on a primary branching node of Palaungic, but the ASJP classification treats it as being even more distantly related to Palaungic than is the entire Khmuic group. Below this node it produces a classification of the Angkuic, Waic, Palaung and Lamet sub-branches that resembles others produced by comparative linguistics, but there is no indication of it clearly approximating the most likely classification suggested by historical phonology. At the lowest level ASJP has produced a nonsense result in respect of Riang, which is difficult to make any sense of, since there is no direct geographical or social connection between the respective speaker communities.

5. Conclusion

After carefully comparing the ASJP trees for Bahnaric, Katuic and Palaungic with expert classifications offered by comparative linguists over more or less the entire effective history of AA studies, it has become clear that the ASJP analyses provide very mixed results. At both micro and macro levels there are both striking errors and striking agreements. However, it is not at all clear that there are consistent or predictable factors underlying these results. For example, at first glance it seems reasonable to suggest that the high similarity found between Rengao and Bahnar could be explained by the known extensive contact between the speaker communities, yet no such explanation can be advanced for the high level of similarity identified for Riang and Lamet. On the macro scale the ASJP does successfully identify various AA branches, yet this is not a tremendous achievement since most are self evident, and have been recognised as such for about a century. Even so, the ASJP did not distinguish Khmuic as a unitary branch, nor did it distinguish it successfully from Palaungic.

My summary assessment is that, in respect of Austroasiatic, the ASJP generates results which, in so far as they correspond to a genetic classification, fair badly, and certainly worse than conventional lexicostatistics. It is not immediately clear to me just what the ASJP is finding, and how it might yield useful insight into the languages.

References

- Cabaton, Antoine. 1905. Dix dialectes indochinois recueillis par Prosper Oden'hal. Etude linguistique par Antoine Cabatone. *Journal Asiatique*, Dixième série, tome V:265-344.
- Deepadung, Sujaritlak. 2009. Ethnicity and the Dara-ang (Palaung) in Thailand. *Journal of Language and Culture* (Research Institute for Languages and Cultures of Asia, Mahidol University) 28.1:7-30.
- Diffloth, Gérard 1977. Mon-Khmer Initial Palatals and “substratumized” Austro-Thai. *The Mon-Khmer Studies Journal*, 6: 39-57.
- Diffloth, Gérard. 1974. Austro-Asiatic Languages. *Encyclopaedia Britannica*. Chicago/London/Toronto/Geneva, Encyclopaedia Britannica Inc. pp 480-484.
- Diffloth, Gérard. 1979. Aslian languages and Southeast Asian prehistory. *Federation Museums Journal*. 24ns:3-16.
- Diffloth, Gérard. 1980. *The Wa Languages*. Linguistics of the Tibeto-Burman Area. Vol. 5.2. Berkeley: University of California.
- Diffloth, Gérard. 1982. Registres, dévoisement, timbres vocaliques: leur histoire en Katouïque. *Mon-Khmer Studies* 11:47-82.
- Diffloth, Gerard. 1982. Subclassification of Palaungic and notes on "P'uman". Paper presented at the 15th Sino-Tibetan Conference on Languages and Linguistics. Beijing.
- Diffloth, Gérard. 1991. Palaungic vowels in Mon-Khmer perspective. In Davison ed. *Austroasiatic Languages, essays in honour of H. L. Shorto*. SOAS, University of London.
- Diffloth, Gérard. 2005. The contribution of linguistic palaeontology to the homeland of Austro-asiatic. In: Sagart, Laurent, Roger Blench and Alicia Sanchez-Mazas (eds.). *The Peopling of East Asia: Putting Together Archaeology, Linguistics and Genetics*. Routledge/Curzon. 79-82.
- Efimov, Aleksandr. 1983. *Problemy fonologicheskoy rekonstrukcii proto-katuicheskogo jazyka*. Kandidat Dissertation, Institute of Far Eastern Studies Moscow.
- Ferlus, Michel. 1971. Simplification des groupes consonantiques dans deux dialectes austroasiens du sud-Laos. *Bulletin de la Société de Linguistique de Paris*, 66: 389-403.
- Ferlus, Michel. 1974. Delimitation des groupes linguistiques austroasiatiques dans le centre indochinois. *Asie du Sud-Est et Monde Insulindien* 5.1:15-23.
- Gainey, Jerry. 1985. *A comparative study of Kui, Bruu and So phonology from a genetic point of view*. Bangkok, Thailand: Chulalongkorn University thesis.
- Gregerson, Kenneth J., Kenneth D. Smith, and David D. Thomas. 1976. The place of Bahnar within Bahnaric. In Philip N. Jenner, Laurence C. Thompson and Stanley Starosta (eds.), *Austroasiatic studies, part I*. Oceanic Linguistics Special Publication, 13. Honolulu: University of Hawaii Press. pp.371-406.
- Kingsada & Shintani. 1999. Basic vocabularies of the languages spoken in Phongxaly, Lao P. D. R. Tokyo : Institute for the study of languages and cultures of Asia and Africa.
- Miller, John and Carolyn. 1996. Lexical comparison of Katuic Mon-Khmer languages with special focus on So-Bru groups in northeast Thailand. *Mon-Khmer Studies* 26:255–290.
- Mitani, Yasuyuki. 1977. Palaung dialects: a preliminary comparison. *Tōnan Ajia Kentyū (South East Asian Studies)*, 15.2: 193-212.
- Mitani, Yasuyuki. 1978. *Problems in the classification of Palaungic*. Paper presented at the 2nd International Conference on Austroasiatic Linguistics, 19-21 December 1978, Mysore, India.
- Mitani, Yasuyuki. 1979. Vowel Correspondences between Riang and Palaung. *Studies in Thai and Mon-Khmer Phonetics and Phonology in Honour of Eugénie J.A. Henderson*. Chulalongkorn University Press, Bangkok, pp142-50.

- Peiros, Ilia J. 2004. *Genetičeskaja klassifikacija avstroaziatskix jazykov*. Moskva: Rossijskij gosudarstvennyj gumanitarnyj universitet (doktorskaja dissertacija).
- Peiros, Ilia J. 1996. *Katuic comparative dictionary*. Canberra, Pacific Linguistics.
- Pinnow, Heinz-Jürgen. 1959. *Versuch Einer Historischen Lautlehre Der Kharia-Sprache*. Wiesbaden, Otto Harrassowitz.
- Schmidt, Pater Wilhelm. 1904. Grundzüge einer Lautlehre der Khasi-Sprache in ihren Beziehungen zu derjenigen der Mon-Khmer-Sprachen. Mit einem Anhang: die Palaung-Wa-, und Riang-Sprachen des mittleren Salwin. *Abh. Bayerischen Akademie der Wissenschaft*, 1.22.3:677-810.
- Schmidt, Pater Wilhelm. 1906. Die Mon-Khmer-Völker, ein Bindeglied zwischen Völkern Zentralasiens und Austronesiens. *Archiv für Anthropologie, Braunschweig*, 5:59-109.
- Shafer, Robert. 1952. Études sur l'austroasiens. *Bulletin de la Société de Linguistique* 4 (8):111-58.
- Shorto, H. L. nd. untitled manuscript, various lexicostatistical statistical analyses, held at the National Library of Australia .
- Shorto, Harry L. 2006. *A Mon-Khmer Comparative Dictionary*: Canberra Pacific Linguistics 579.
- 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 & Pascale Jacq. 2003. *A Handbook of Comparative Bahnaric: Volume 1, West Bahnaric*. Canberra, Pacific Linguistics 551.
- Sidwell, Paul. 1999. *A Reconstruction of Proto-Bahnaric*. PhD thesis, University of Melbourne.
- Sidwell, Paul. 2005. *The Katuic Languages: classification, reconstruction and comparative lexicon*. Munich, Lincom Europa.
- Sidwell, Paul. 2009. *Classifying the Austroasiatic languages: history and state of the art*. Munich, Lincom Europa.
- Sidwell, Paul. 2010. How many branches in a tree? Cua and East (North) Bahnaric. In Bethwyn Evans (ed). *Discovering history through language. Papers in honour of Malcolm Ross*. Canberra: Pacific Linguistics.
- Smith, Kenneth, D. 1973a. Eastern North Bahnaric: Cua and Kotua. *Mon-Khmer Studies*, 4: 113-118.
- Smith, Kenneth, D. 1981. A Lexico-statistical study of 45 Mon-Khmer languages. In Andre Gonzalez and David Thomas (eds), *Linguistics Across Continents*. Manila SIL, pp 180-205.
- Theraphan Luang-Thongkum. 1997. *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.
- Theraphan Luang-Thongkum. 2001. *Languages of the Tribes in Xekong Province Southern Laos*. The Thailand Reseach Fund.
- Theraphan, Luang-Thongkum.. 2002. A brief look at thirteen Mon-Khmer languages of Xekong province, southern Laos. In *Collected papers on southeast Asian and Pacific languages*, edited by R. Bauer: Canberra Pacific Linguistics.
- Thomas David. 1964. A survey of Austro-asiatic and Mon-Khmer comparative studies. *Mon-Khmer Studies* 1:49-163.
- Thomas David. 1966. *Mon-Khmer subgroupings in Vietnam*. In Norman Herbert Zide, ed. *Studies in Comparative Austroasiatic Linguistics*. Indo-Iranian Monographs, V. The Hague, pp 194-202.
- Thomas David. 1979. The Place of Alak, Tampuon, and West Bahnaric. *Mon-Khmer Studies* 8:171-186.
- Thomas, David & Robert Headley. 1970. More on Mon-Khmer subgroupings. *Lingua* 25:398-418.
- Thomas, Dorothy M. 1976. A phonological reconstruction of Proto-East-Katuic. SIL (Univ of ND session) Work papers 20, supplement 4.