Diversity, discontinuity and asymmetry in the typological restructuring of Mainland Southeast Asian languages.

Paul Sidwell
Australian National University (Canberra) &
Centre for Research in Computational Linguistics (Bangkok)

0. Introduction

This paper takes as its point of departure the challenge mounted by Post (2011) to the broad characterisation of MSEA languages as falling into "Indospheric" and "Sinospheric" convergence areas, advocated by the likes of Matisoff (1991), Bradley et al. (2003) and others. Post argues that the Indoshpere/Sinosphere wrongly suggests that Indic and Sinitic linguistic influences actually explain the typological divide observed in Mainland southeast Asia, yet on the ground there are languages which although broadly fitting the categories, demonstrably lack the history of pre-historical contact and dominance/subordination relationships implied by the model. While Post proposes a mechanism based on prosodic convergence that does not require extensive bilingualism or other intensive linguistic interaction (extending the insights of Donegan & Stampe 1983, 2004, etc.), his account still falls within the scope of a contact driven model of explanation. In sympathy with Post's results, I wish to add some more grist for the mill by drawing attention to some examples of restructuring that are arguably inconsistent with their local areal contexts. This is interesting because it challenges us to consider how we can know what role - if any - contact has played.

The examples that are discussed below, and many others that could be brought to bear, directly challenge strong claims that are made about Southeast Asia as a linguistics area:

> ....Mainland Southeast Asia is indeed an internally homogenous linguistics area. It is most clearly set off from the rest of Asia, though with some fluidity in the boundary for different features, with such regions as the northeast of South Asia and China often showing up as transition zones. [...] Only a few features, such as tone, show a boundary at or close to this divide.
> Comrie (2007:44)

Such claims of "homogenous" areality are based on identifying common features such as tone, and yet tone is so diverse in its manifestation and origins within the area that simply identifying two or languages as tonal may count for very little. Many languages of the area are not tonal - or otherwise lack features associated with this "homogenous" area - despite being in conditions of subordination by common dominant languages for significant periods. We cannot simply pick and choose, drawing attention to facts that we claim are consistent with our areality hypotheses and ignore those which are not, with glory going to the one who is able to pile the greatest list of consistent facts. A scientific approach must be generally explanatory, showing its worth by being able to cope with of the most inconvenient facts.

Close examination of various Austroasiatic (AA) data challenges a simple narrative of contact driven convergence; within various individual AA branches, even among close neighbours, one finds gross discrepancies in phonological restructuring, or neighbours who are restructuring in broadly parallel ways (e.g. Laven and Nyaheun towards monosyllables, Angkuic languages becoming tonal), and yet at a micro level are doing so

by different paths. These kinds of contradictions are found across AA. Even in Munda, where there would seem to be overwhelming areal pressure for stable multisyllabic words simple (C)V(C) syllables, we find the counter examples such as Gta' with strikingly SEAsian sesquisyllabic structure (discussed below). Rather strikingly, these examples involve small languages that are in clearly subordinate social positions to their larger neighbours, yet those neighbours cannot be held to model the changes at stake. Instead of making these small languages more susceptible generally to areal pressures, it may be that their size is somehow connected with the apparent independence of their histories.

The structure of the discussion that follows is as follows takes four case studies will be treated at some length, followed by more general discussion. There are potentially many more cases that could and should be examined from this perspective.

1. Two stories of atypical tonogenesis/registrogenesis

1.1 Angkuic

The Angkuic languages represent a subgrouping within the Palaungic branch of AA, and aee spoken by small communities in Yunnan and boarder areas. The group is named after the language identified as Āng-kú by Scott (1900).[1] Although Angkuic became known to scholars in the late 1800s, those early records are fairly brief and linguistically naive, and thus are not particularly useful. However, several Angkuic languages have been expertly documented and analysed in more recent times, and this work reveals a fascinating story of syllabic, segmental and tonal restructuring that runs counter to the immediate areal trends, and even within Angkuc displays some diversity. Three languages are relevant here: U and Hu, discussed by Svantesson (1988, 1989, 1991) and Muak Sa'aak, more recently described in a 2010 MA thesis by Elizabeth Hall. Both Svantesson and Hall analyse historical phonology and tonogenesis in the respective languages of their studies, while in this paper I integrate the analyses to offer a broad account of the development of Angkuic with special reference to tonogenesis.

A key insight relates to understanding the role of vowel length in the historical phonology; while the length contrast has been lost from two of the three languages - arguably under direct areal pressure - the feature is robustly if indirectly reflected in the tonal systems of all three, and contact can only be invoked as a marginal feature in tonogenetic history. This author has prepared a reconstruction of protoPalaungic phonology and lexicon, a working version of which can be accessed online at sealang.net/monkhmer. This is a work in progress, so consequently the online version lags somewhat behind the current state of my compilation and analysis, and this should be taken into account when using it. The discussion that follows in respect of Angkuic draws directly upon that working reconstruction and the works of Svantesson and Hall mentioned above.

[1] The name Āng-kú does not obviously correspond to any contemporary known, although Hall compares it to Kon Keu (ISO KKN). The kú probably corresponds *ku* 'quantifier for persons' to Palaung, thus the Āng-kú likely meant something like 'the people'.

pPalaungic	gloss	Muak Sa-aak	Hu	U	Lamet	Palaung
*pəɲ	'to shoot'	**pʰɤɲ³**	**pʰɨ́ɲ**	**pʰèt**	**piɲ**	**piŋ**
*-taʔ	'tail'	**k.tʰa:²**	**θatʰáʔ**	**satʰà**	**ntaʔ**	**səta**
*kɔ:n	'child'	**kʰuan³**	**kʰɔ̀n**	**-**	**kɔn**	**kuən**
*bri:ʔ	'forest'	**pri²**	**pʁíʔ**	**qí**	**prì:ʔ**	**bri**
*plɗa:k	'palm (hand)'	**puc² ta:k¹**	**pʰltàk**	**ʔatăʕ**	**pltà:k**	**-**
*ga:ŋ	'house'	**ka:ŋ³**	**kàŋ**	**káã**	**-**	**gaŋ**
*gak	'to bite'	**kak²**	**kák**	**kàk**	**kàk**	**-**

Table: Angkuic 'Germanic' shift examples

The Angkuic languages are readily recognised as a sub-group of Palaungic by reason of their sharing a so-called Germanic sound shift (recognised by Haudricourt 1965); historical voiceless stops became aspirated, while pPalaungic plain and implosive voiced stops merged to plain voiceless stops. Not only is this shift important for identifying the Angkuic group, it also removed the voicing (phonation) contrast normally associated with high/low tonogenetic series in AA languages (see Huffman 1985 for a general theory of phonological restructuring in AA languages). We might also invoke contact with tone languages - especially Shan, Tai Lue or South-western Mandarin - but this also fails to explain the path of tonogenetic development revealed by close analysis.

Hu: high/low tones

The simplest tone system among the Angkuic languages discussed here is that of Hu. Svantesson (1991) characterises the general issues neatly:

> From a general phonological point of view, the most interesting phenomenon is the development of a two-tone system where the tones are not the reflexes of voiced/voiceless proto-initials, as is most often the case in Mon-Khmer two-tone (or two-register) languages. Instead, the tones are the reflexes of the long/short vowel opposition which existed in Proto-Palaungic (inherited from Proto-Mon-Khmer). As far as I know, no language with this kind of tonogenesis has been described before. (Svantesson 1991:67)

Hu has two phonological tones - high and low - and Svantesson (1991:75) shows the relation between tone and vowel length with the following measurements (Svantesson's mean values only shown):

		Mean durations by rime
High tone	**jám**	126 ms
	páp	102 ms
	kák	117 ms
High tone	**jàm**	200 ms
	kàp	120 ms
	ʔàk	188 ms

Table: Hu tone/length relationship

Svantesson's measurements show that vowel length distinctions have not completely leveled in Hu; while the high tone vowel in **kák** is virtually the same length as the low vowel in **kàp**, still overall there is something like an average of 40 ms difference between short and long vowels in comparable environments with low vowels longer than their high

counterparts and visa versa. In terms of how length contrasts are realised in phonologically conservative AA languages, we lack broad statistically robust datasets needed to make strong generalisations, but experience suggests that long/short is typically realised as contrasts in duration ranging roughly in the order of 150~220ms (long) versus 60~100ms (short). The Hu measurements effectively show all vowel falling within the typical range of AA long vowels, with high tone vowels in the lower end of the range while low tone vowels distribute over a wider range but one that tends toward the longer values associated with AA long vowels.

That a high/low tone contrast should emerge from length differences has been remarked upon as being very unusual; Svantesson was not aware of any parallel, and more recently Kingston (2011) in a survey of tonogenesis, could only nominate Western Lugbara (Nilo-Saharan) as another example. Svantesson tries to link Hu tones to ±ATR, given that there are various collocational restrictions between vowel height and tone, but in my reconstruction these relations are better modelled as the outcome of tonogenesis, and are not discussed further here. More important for our discussion is the fact that closer to home, both Köho/Sre (South Bahnaric; Manley 1972) and Nyaheun (West Bahnaric; Davis 1968), scholars have reported non-phonemic falling tones on long vowels. While one may assume that the falling tone is associated with the general tendency for F1 to fall somewhat over longer syllables, there is the problem how to explain the phonologization in this case. The measurements of Svantesson indicate that Hu short vowels *lengthened*, while the long vowels remained largely unchanged in quantity.

In unrestructured AA languages (in the paradigm of Huffman 1985) the long vowels are less marked phonologically than the short, and neutralisation of quantity is generally realised as lengthening; for a well understood example see the history of Sedang as analysed by Smith (1979) and elsewhere in this paper. Assuming that falling tone was associated with the unmarked long vowels, it is reasonable to suggest that the non-falling pitch contour over the historically short vowels was accentuated and phonologized as a high(er) tone. Consequently, it is tempting to propose reconstructing a proto-stage - approximating pre-Hu or pAngkuic - where the vocalism is transitioning through a restructuring of vowel length marked by the phonologisation of a high versus falling tone contrast, eventually realised as high versus low in Hu. However, we need to consider other Angkuic data before moving forward with this idea.

U: high/low/falling/rising tones

The U language - closely related to Hu - was described and analysed in detail by Svantesson (1988). U has undergone even more extensive restructuring, with much more reduction of disyllables into monosyllables and associated segmental mergers. Additionally, and more importantly here, the basic two-tone system underlying Hu split further, creating a four way high/low/rising/falling system. However, the emergence of four tones in U was not a simple matter of secondary splits of the simple high/low system; according to Svantesson's analysis there are actually nine distinct outcomes according to different rimes, which then sort out into an overall four-way tone system.

Putting aside two marginal cases, the broad picture of tonal development in U, with conditioning environments specified, is as follows:

		Examples: pPalaungic → U
pAngkuic short vowels: *V̆	→ V̀ /_C[son.] → V́ elsewhere	***jam → jàm** ‘to die’ ***gak → káʕ** ‘bite’
pAngkuic long vowels: *V:	→ V̂ /C[+son.] _ → V̆ /C[any] _C[-son.] → V́ /C[-son.] _C[+son] → V́ /C[-son.] V[+high] Ø → V̀ /C[-son.] V[-high] Ø	***ʔma:r → mâ** ‘field’ ***hla:t → lăt** ‘to fear’ ***kta:m → tʰám** ‘crab’ ***ciʔ → ncʰí** ‘louse’ ***kaʔ → kʰà** ‘fish’

Table: U tonal developments

Without going into meticulous detail over each development, we can readily see that the tonogenetic history of U and Hu, when treated in isolation, suggests a straightforward starting point of the binary tonal system described for Hu, with subsequent multiple secondary changes in U. For a detailed discussion of the segmental features and their roles in these process consult Svantesson (1988).

However, this is not the end of the matter, as recently Hall’s (2010, henceforth *MS*) description and analysis of Muak Sa’aak provides evidence that a somewhat more complex history underlies tonogenesis in Angkuic.

Muak Sa’aak: three tones plus length contrast

At the time Svantesson was writing in the 1980s/early 1990s, it was not clear that any Angkuic language had retained the historical AA vowel quantity distinction, and Lamet (also *Ramet*) examples - apparently alone among Palaungic as retaining the length contrast - were cited for comparative analytical purposes. Lamet (as described variously by Lindell et al. (1978) and Narumol (1980) underwent conventional devoicing and high/low register series formation; but being spoken in Northern Laos where it was surrounded by Khmuic speakers rather than dominated by Shan (with no vowel length contrast) vowel length was never lost, at least that was the thinking. But thanks to Hall’s thesis, we now have a description of MS phonology which remarkably shows a fully elaborated vowel quantity contrast, much as we find in Lamet and typically in unrestructured AA languages. None the less, MS is unquestionably Angkuic; as it shares both the Germanic sound shift and a tonal system directly related to vowel quantity.

Close	**i**	**i**:	**ɯ**	**ɯ**:	**u**	**u**:
Close-mid	**e**	**e**:	**ɤ**	**ɤ**:	**o**	**o**:
Open		**ɛ**	**a**	**a**:		**ɔ**
Diphthongs		**ia**				**ua**

Table: Muak Sa’aak vowel phonemes according to Hall (2010)

The three MS tones are described as follows:

- Tone 1: low tone with tense phonation, not consistently creaky;
- Tone 2: so called “checked tone”, high pitch on short open syllables and with stop codas, and falling pitch on long open syllables and with nasal codas;
- Tone 3: high falling tone occurring only with sonorant codas and open syllables (“live” syllables).

Other than the restriction against stop codas with Tone 3, all three tones occur across a broad range of environments that are difficult to characterise if one tries to account for the full synchronic distribution with out considering etymology. There are also independent secondary changes that in some codas created new long syllables, such as by the loss of finals ***-ʔ***, ***-h***, and ***-s***. Additionally there is a huge cohort of Shan loans that fills what would otherwise be many systematic gaps in the distribution of rimes. It is quite striking to note that the strong influence of Shan and Tai Lue on MS - repeatedly noted and illustrated by Hall - is not associated with the loss of vowel quantity that is a feature of U and Hu - and this calls into question the basis of any contact driven explanation.

If one removes from consideration the extensive Shan loanwords, important asymmetries emerge, a point commendably recognised by Hall but not well followed through. This allows us to more effectively isolate the tonogenetic conditions, and establish that there is a clear connection between etymological vowel length and tone in MS, although it is not the only factor. The basic relations are tabled as follows:

		pPalaungic	Hu	Muak Sa'aak
Tone 1: low	V̀: _ C[-voice]	***li(:)k** ***hŋa:p** ***leh** ***-ta:k** ***ka:p** ***ʔa:k**	**lèk** **-** **líh** **ntʰàk** **kʰàp** **ʔàk**	**le:k¹** 'pig' **ŋa:p¹** 'difficult' **li:¹** 'to exit' **tʰa:k¹** 'tongue' **kʰa:p¹** 'chin' **ʔa:k¹** 'hunting bow'
Tone 2: high allotone	V́ / _ C[-voice]	***kuʔ** ***ti:ʔ** ***suk** ***ʔiət**	**-** **tʰíʔ** **θúk** **ʔɛ̀t**	**kʰu²** 'body' **tʰi²** 'hand, arm' **suk²** 'hair' **ʔɛt²** 'to sleep'
Tone 2: falling allotone				Loanwords only
Tone 3: falling	V̂ / _ C[+voice]	***jam** ***rim** ***ɓi:l** ***ja:m** ***dɨəm** ***ga:ŋ** ***ʔma:r** ***kɔ:n**	**jám** **ʁím** **(pìn U)** **jàm** **tɛ̀n** **kàŋ** **mà** **kʰɔ̀n**	**jam³** 'to die' **rim³** 'village' **pil³** 'forget' **ja:m³** 'to weep' **tian³** 'low' **ka:ŋ³** 'house' **ma:l³** 'field' **kʰuan³** 'child'

Table: Muak Sa'aak tonal developments

Commentary on tones:

Within Tone 1 historically long vowels are generally indicated, except in cases such as the 'to exit' example where apparently vowels have lengthened with the loss of weak codas such as /h/.

Tone 2 is divided into two allotones by Hall, but the falling allotone is restricted to loanword vocabulary, so we can treat it as secondary. This shows both long and short vowels, but is consistently marked by stopped finals, including glottal stop which clearly persisted into pAngkuic (as shown by Hu comparsions) but was later lost in MS.

Tone 3 shows both long and short vowels, but is marked by having historically voiced codas consistently.

The above suggests the following simple schema:

	Vowel short	Vowel long
Vowel + voiceless coda	Tone 2	Tone 1
Vowel + voiced coda	Tone 3	Tone 3

Table: Muak Sa'aak historical tonal schema

Proto-Angkuic tonogenesis reconstructed

The above data and discussion raises a serious problem for the reconstruction of pAngkuic. Restricting consideration to just Hu and U, one could straightforwardly propose a direct relation between quantity and tone with the pAngkuic continued unchanged in Hu. However, the integration of the MS data is problematic: persisting with a simple high/short versus low/long historical model would force one to posit the merging of high and low tones into a common falling tone before voiced codas, which strikes me as a brutally arbitrary reconstruction.

Motivated by a desire for phonetic realism, and cognisant that we are dealing with tonal features that are directly relatable to segmental phonology, I propose that we can reconcile the different systems by reconstructing four distinct phonetic tonal contours for pAngkuic which subsequently phonologized (realised as allotones) differently in pre-Hu-U and pre-Muak. Thus the following reconstuction:

	Vowel short	Vowel long
Vowel + voiceless coda	high	mid/low
Vowel + voiced coda	high falling	mid/low falling

Table: Proto-Angkuic phonetic tones

The hypothesis is that while quantity conditioned whether the contours were generally high or mid/low, coda voicing conditioned whether contours were level or falling. It then becomes reasonably straightforward to propose that the tones in rimes with voiced codas become phonologized as a single falling tone in preMuak, while in pre-Hu-U the high and high-falling were phonologized as a high tone while the mid/low and mid/low falling were phonologized as low tones.

Ideally one would attempt to investigate this further by examining pitch traces for both Hu and MS, looking for signs of the reconstructed pAngkuic allotones. In any case, Svantesson's initial insight, that vowel length plays a special and unusual role in Angkuic tonogenesis is confirmed as valid. We are presently compelled to reconstruct a general relation between pitch and quantity in pAngkuic that was phonologized in ways that do not fit the 'typical' tonogenetic pathways associated with SEAsian languages, nor more specifically with the restructuring paradigm of Huffman (1985), or the "Laryngeally based account" of Thurgood (2007) that broadly underpin our understanding of tonogenesis and orient our approach to the reconstruction of AA vocalism. We are unable to invoke areal influences of Shan, Tai Lue etc. to explain away the changes in Angkuic, and thus must take a more sophisticated approach than vaguely invoking notions such as borrowing or metatypical restructuring to explain these developments. Programmatically speaking, internal explanations - grounded in articulatory and acoustic phonetics - with at most only partial input by models of contact driven change are required in this case.

1.2 North Bahnaric

The Bahnaric languages are spoken mainly central highlands of Vietnam and constitute perhaps the most internally diverse branch of AA. The languages are spread over three countries, where they are separated not only by typography but also the (sometimes overlapping) areal influences of Vietnamese, Khmer, Thai/Lao, Chamic and the Katuic branch of AA. Despite the diverse influences from variously restructured and/or tonal languages, most Bahnaric tongues are relatively conservative, preserving the historical AA voice distinction in stops, and vowel systems that strikingly resemble, for example, that of Old Khmer before it was restructured (see e.g. Ferlus 1992).

In this context, it is very significant that the North Bahnaric (NB) languages, a sub-group of around ten closely related speech varieties spoken in the vicinity of Kontum, evince a phonation or "register" contrast in their phonology whose origins are unusual even in Southeast Asia.[2] In short, they show tense-lax registers that appear to correlate directly with historical vowel height distinctions, rather than consonant phonation types.

The historical phonology of NB registers was discussed by Smith (1967, 1972, 1979) and further analyzed by this writer (e.g. Sidwell 1998). These studies allow us to sketch out what happened in terms of segmental changes, although the understanding of the articulatory, acoustic and perceptual mechanisms underlying the restructuring remains incomplete. Below I briefly outline our understanding of the historical vocalism, with reference to three key NB languages, contrasting this with what we know about the most important historical contact languages.

Rengao

According to Gregerson (1976, 1977), Rengao has a highly symmetrical inventory of 10 long and 10 short vowels, each divided into five lax and five tense, as follows:

long lax	long tense	short lax	short tense
iː **u**ː		**i** **u**	
	ᵉi **ᵒu**		**ɪ** **ʊ**
eː **ə**ː **o**ː		**e** **ə** **o**	
	ɛː **a**ː **ɔ**ː		**ɛ** **a** **ɔ**

Figure: Rengao vowels (Gregerson 1976, 1977)

It is immediately evident that the registers correlate directly with aperture, such that tense vowels are lower, or have lower onsets compared to their apparent lax counterparts. From a purely phonological viewpoint one could treat either register or timbre as primary, and disregard the other. However, it is clear that these two features are closely linked, and it would be wrong to disregard the relationship.

Gregerson (1977) characterizes the registers as follows:

[2] Arguably the closest parallel is in Pacoh, a Katuic language spoken roughly a hundred kilometres north of the NB area. Synchronic and diachronic aspects of Pacoh registers are discussed in, e.g. Ferlus (1982), Alves (2004), Sidwell (2005). The parallel with NB registrogenesis is only partial, but shares important features especially in terms of the correlation between vowel aperture and phonation.

	First Register (tense)	**Second Register (lax)**
resonance	sharply defined oral (clear)	deep pharyngeal
pharyngeal cavity	constricted	expanded
larynx	normal to high	lowered
tongue root	retracted	advanced
tongue blade	lowered	raised

Figure: Rengao registers (Gregerson 1976)

Of the above, the height of the tongue blade is clearly important, since it correlates directly with F1, which is the most important contributor to vowel aperture. Significantly, Gregerson connects tongue blade height with the action of the tongue root and other gestures associated with manipulating pharyngeal cavity volume, and we will revisit this after discussing two other NB languages and the implications of the phonological correspondences between them.

Halang

Halang phonology is discussed by Cooper & Cooper (1966). They describe the language as having a breathy register which is only contrastive on the long vowels; the breathy short vowels are only slightly breathy such that timbre is clearly more salient. Also, their phonetic description of the allophones indicates some minor asymmetries in the phonetics of the mid and open vowels, but these are trivial. More significant is the aperture difference between the tense and lax diphthongs:

long lax		long tense		short lax		short tense	
iː	**u**ː	**i**ː	**u**ː	**ɪ**	**u**		
eː		**e**ː		**ə**			**o**
aː	**ɔ**ː	**a**ː	**ɔ**ː			**ɛ**	**a**
iə	**uə**	**ea**	**oa**				

Figure: Halang vowels (Cooper & Cooper 1966)

Although it appears that the register contrast is robust among the long monophthongs, it is also evident that among the diphthongs and short vowels there is a direct correlation between aperture and phonation that broadly parallels the situation in Rengao.

Sedang

Sedang has a strikingly different vocalism compared to other NB languages, although thanks to the comparative work of Smith (1967, 1972, 1979) we have a good understanding of how to relate Sedang phonology to the rest of NB. The vowel inventory is as follows:

monophthongs		diphthongs					
i	**u**	**iə**	**uə**	**io**	**uo**	**iɪ**	
e	**o**	**eə**	**oə**	**eo**			**oɛ**
ɛ	**a**	**ɔ**					
Vowels occur in two registers: tense/creaky & lax/modal							

Figure: Sedang vowels (Smith 1979)

Broadly speaking, Sedang dramatically restructured phonologically, with outcomes that include:

- the loss of quantity distinction in vowels, thanks to a general diphthongization of long vowels and lengthening of etymological short vowels;
- the general devoicing of voiced stops, and acquisition of new voiced stops by borrowing and cluster assimilation;
- the loss of final stops in tense syllables leading to frequent homophony;
- a general tensing of phonation such that modal (tense) syllables became creaky and breathy (lax) syllables became modal.

An effect of the above changes was that Sedang comes to look very different from the rest of the group, but the changes are so regular that it still provides a reliable witness for reconstructing historical vowel length, quality and register values. Of particular importance is the fact that, although the Sedang vowels occur generally across the registers, significant asymmetries emerge when historical mergers and loans are taken into account, and these confirm the correlations evident in the Rengao and Halang data.

Proto-North Bahanric vowel-register correlation

Fortunately we are not solely reliant on NB data in order to investigate the history of registrogenesis within the subgroup. Sidwell (1998) offered a preliminary pBahnaric reconstruction, this was modified by Sidwell (2002) and further extended and refined by Sidwell (2011), and this foundation provides both a framework for analysis the correspondences and a resource for etymological comparisons, and underlies the following discussion.

First of all, it is suggested by the lack of an effective register contrast among Rengao and Halang short vowels that this state of affairs goes back to pNB. Mostly this is confirmed by the lack of corresponding diphthongal reflexes in Sedang, although the correspondences indicate a subtly more complex story. In some environments Halang reflexes of ***i** and ***u** are lengthened, as part of a chain shift in which pNB ***iː** and ***uː** become lax diphthongs. Conversely, the pNB long high vowels lowered in Rengao in a chain shift that went the other way ultimately eliminating diphthongs in that language.

The picture that emerges from correspondences such as the above shows the etymological high vowels consistently reflected with lax phonation in NB daughter languages (breathy in Rengao and Halang, modal in Sedang). Additionally, we see restructuring of diphthongs: while the pNB diphthongs have tense reflexes (modal in Rengao and Halang, creaky in Sedang where codas have not been dropped) there are various secondary phonetic developments. In Halang, ***iə,*uə** lowered to **/ea, oa/**, while ***i**ː**, *u**: diphthonged to **/iə, uə/**, and short vowels in various environments lengthened filling the space vacated as the newly diphthonged vowels. In Rengao they are reflected as tense monophthongs, in some environments becoming short. The association with tense register suggests a connection between **iə,*uə** and the historical low vowels.

	North Bahnaric				CentralB	WestB	
gloss	Rengao	Halang	Sedang	pNB	Bahnar	Laven	pBahnaric
‘mushroom’	**bətsi̤t**	**pəsi̤ːt**	**kəset**	***psit**	-	**pseːt**	***pseːt**
‘sky’	**pli̤ŋ**	**pli̤ːŋ**	**pleŋ**	***pliŋ**	**pleŋ**	-	***pleːŋ**
‘banana’	**pre̤ːt**	**pri̤ət**	**priət**	***priːt**	**priːt**	**priət**	***priːt**
‘to dig’	-	**ci̤ər**	**ciəl**	***ciːr**	**siːr**	-	***ciːr**

'tongue'	**rəpɛt**	**rəpeat**	**rəpe**	***lpiət**	**rəpjɛt**	**hpiat**	***lpiət**
'chicken'	**ʔiːr**	**ʔear**	**ʔi̱**	***ʔiər**	**ʔjɛr**	**ʔiar**	***ʔiər**
'to push'	**drụt**	**drụt**	**trɔt**	***drut**	**drut**	**-**	***drut**
'fire'	**ʔụn**	**ʔụn**	**ʔɔn**	***ʔuɲ**	**ʔuɲ**	**ʔuɲ**	***ʔuɲ**
'axe'	**cọːŋ**	**cụəŋ**	**cuoŋ**	***cuːŋ**	**suːŋ**	**cuːŋ**	***cuːŋ**
'smoke'	**ɲọːj**	**ʔɲụəj**	**ŋoj**	***ʔɲuːj**	**ʔɲuːj**	**ʔɲuːj**	***ʔɲuːj**
'four'	**puːn**	**poan**	**pu̱n**	***puən**	**pwan**	**puan**	***puən**
'to cut'	**-**	**poat**	**poe**	***puət**	**pwat**	**puat**	***puət**

Table: Proto North Bahnaric high vowels and diphthongs

At the other end of the vowel chart, the etymological low vowels have consistently tense reflexes in NB languages. Some examples are tables below, although note that finding suitable examples of ***ɛ**ː is difficult, so I restrict the treatment of long vowels here to ***a**ː and ***ɔ**ː:

	North Bahnaric				CentralB	WestB	
gloss	Rengao	Halang	Sedang	pNB	Bahnar	Laven	pBahnaric
'squirrel'	**prɔːk**	**proak**	**proə**	***prɔːk**	**prɔːk**	**prɔːk**	***prɔːk**
'tree'	**lɔːŋ**	**ʔloaŋ**	**lo̱əŋ**	***ʔlɔːŋ**	**ʔlɔːŋ**	**ʔlɔːŋ**	***ʔlɔːŋ**
'child'	**kɔːn**	**koan**	**ku̱ən**	***kɔːn**	**kɔːn**	**kuan**	***kɔːn**
'hungry'	**məŋɔːt**	**məʔŋoat**	**məŋuə**	***pŋɔːt**	**pəŋɔːt**	**pŋuat**	***pŋɔːt**
'navel'	**klɔk**	**klɔk**	**klɔ**	***klɔk**	**klɔk**	**klɔk**	***klɔk**
'back'	**rɔŋ**	**rɔŋ**	**ro̱ŋ**	***rɔŋ**	**hərɔŋ**	**-**	***(k)rɔŋ**
'water'	**daːk**	**daːk**	**teə**	*daːk	**ɗaːk**	**daːk**	***ɗaːk**
'eagle'	**klaːŋ**	**klaːŋ**	**kəklo̱əŋ**	***klaːŋ**	**klaːŋ**	**klaːŋ**	***klaːŋ**
'blood'	**məhaːm**	**məhaːm**	**məho̱əm**	***phaːm**	**phaːm**	**phaːm**	***bhaːm**
'to slap'	**taːp**	**taːp**	**teə**	***taːp**	**taːp**	**-**	***taːp**
'to hunt'	**daŋ**	**daŋ**	**to̱ŋ**	***daŋ**	**-**	**daŋ**	***daŋ**
'trap'	**dak**	**dak**	**ta**	***dak**	**ɗak**	**dak**	***ɗak**
'woodpecker'	**təlɛh**	**təlɛh**	**təlo̱ej***	***təlɛh**	**təlɛh**	**-**	***təlɛh**
'to think'	**kəcɛŋ**	**kəcɛŋ**	**təco̱ŋ**	***təcɛŋ**	**təcɛŋ**	-	-

Table: Proto North Bahnaric low vowels

* 'hook used to pick fruit from tops of trees'

It is evident that the historical low vowels are only reflected as tense among the NB languages; the more problematic correspondences are those reflecting historical mid vowels. In this regard it is especially important to consider the fate of the pB mid central vowels ***ə**ː and ***ə**; these are frequent in pB, and typically reflected as mid central vowels in Bahnaric languages, but strikingly there is no **/ə**ː**/** vowel in NB languages, and typically short **/ə/** vowel is either absent or is functionally the lax equivalent of tense **/a/**.

It is apparent that historical ***ə**ː is reflected as a front vowel - mostly **/e**ː **~ ɛ**ː**/** - where it has variously merged with etymological ***e**ː or been involved in a chain shift in which it displaced etymological ***e**ː (merging to **/i/** as in examples 'mushroom' and 'sky' above), in

any case the reflexes are consistently lax, while the reflexes of ***o**: and ***o** are tense. Examples:

	North Bahnaric				CentralB	WestB	
gloss	Rengao	Halang	Sedang	pNB	Bahnar	Laven	pBahnaric
'leech'	**plɐ̣ːm**	**plɐ̣ːm**	**pliəm**	***plɐ̣ːm**	**pləːm**	**plʌːm**	***pləːm**
'beginning, stump'	**tsɐ̣ːm**	**sɐ̣ːm**	**pəsiəm**	***tsɐ̣ːm**	**təːm**	**tʌːm**	***təːm**
'answer'	**tɐ̣ːl**	**tɐ̣ːl**	**tiəl**	***tɐ̣ːl**	**təːl**	**tʌːl**	***təːl**
'crowded'	-	**gədrə̣m**	**kram**	***krə̣m**	**kədrəm**	-	***krəm**
'thick'	**həbə̤l**	**həbə̤l**	**həbɔ**	***həbə̤l**	**həɓəl**	**kbəl**	***-ɓəl**
'to blow'	**hluːm**	**hluːm**	**hlu̱m**	***hloːm**	**hloːm**	**kloːm**	***k(h)loːm**
'hot'	**tuʔ**	**tuʔ**	**to**	***toʔ**	**tɔʔ**	**ʔtoʔ**	***toʔ**
'to spit'	**cuh**	**kəcuh**	**kəcow**	***kcoh**	**kəsɔh**	**kcoh**	***kcoh**

Table: Proto North Bahnaric mid vowels

While the examples given above illustrate only part of a larger more complicated history, they are strongly indicative of the historical processes that governed the evolution of NB phonology. Broadly we see that, regardless of vowel length, and without regard to the phonation of onset or coda consonants, vowels with historical quality **/i, u, e, ə/** are reflected as breathy register vowels, and historical **/o, ɔ, ɛ, a, iə, uə/** have tense reflexes. This correlation between aperture and phonation is almost perfect; there is a small number of ambiguous etymologies that may show historical **/o/** vowels with lax outcomes, but this trivial compared to the otherwise overwhelming tendency evident within the history of NB, and we are forced to ask not just how, but why, did this come about?

Proto-North Bahanric registrogenesis in context

Much of our present understanding of tonogenesis and registrogenesis is well summed up in the recent works by Thurgood (2002, 2007), Brunelle (2005) etc. Thurgood tables the phonetic correlates of register as follows:

	Tense Register	**Unmarked**	**Breathy Register**
original initials:	proto-voiceless		proto-voiced
voice quality:	tense (creaky)	modal (clear)	breathy
vowel quality:	lower (open); more fronted vowels; tendency to diphthongization; often shorter		higher (closed); more backed vowels; tendency to centralization; often longer
pitch distinctions:	higher pitch		lower pitch
state of larynx:	larynx tense and/or raised (reduced supraglottal cavity)		larynx lax and/or lowered (increased supraglottal cavity

Figure: "The three most common register complexes" Thurgood (2007:274)

And recently Mortensson explains:

> It is well know that there is an association between phonation type and vowel quality. Specifically, breathy phonation requires a lowering of the larynx, which has the effect of lowering formants, especially F1 (Fagan 1988; Ladefoged and Maddieson 1996; Gordon and Ladefoged 2001; Brunelle 2005). This effect may give rise to either low-level allophonic splits or to more phonologically significant developments.
> (Mortenson 2006:13)

Mortenson is certainly correct that formant lowering associated with breathy phonation can give rise to significant developments, and this is described neatly by, e.g. Huffman (1985). But what is at stake here, apparently, is that there was no trigger of the conventionally understood type (e.g. voicing of onsets with subsequent devoicing) which would have given rise to breathy phonation on syllable nuclei. The distribution of registers in NB is utterly blind to onset voicing, or any other consonantal features, with the exception of the well understood much later secondary developments within Sedang. What seems to have occurred in the history of NB is that phonation differences inherently associated with aperture differences became phonologized. This is unusual, as Kingston (2011:2319) remarks (including register within a broad understanding of tone), "Tone splits from vowel height or ATR contrasts are decidedly rare."

Brunelle, discussing the correlations between register and vowel height remarks:

> A study of vowel quality in another Mon-Khmer language, Wa, shows no clear difference between registers (Watkins 2002). This is not surprising, as vowel quality is the most variable correlate of register, varying from systems where there are no differences between the two registers to systems where vowel quality becomes the only register cue. For example, register has lost its phonemic status in Standard Khmer (Huffman, F. 1978), after conditioning a two-way split of the vowel system, leaving the vowel space unusually overcrowded. However, when register does condition vowel quality differences, the high register always has lower vowels than the low register.
> (Brunelle 2005:165-6)

Elsewhere in his thesis Brunelle discusses the correlations between vowel timbre and register in Cham, and a significant relation between F1 and phonation:

> Overall, the first formant of the high register (h) has higher frequency, which means that high register vowels tend to be more open. This is expected because the lengthening of the vocal tract due to the lowering of the larynx during the production of the low register results in lower formant frequencies, especially for F1.
> (Brunelle 2005:187)

However, the general theory of registrogenesis, with which Brunelle is in concord, assumes laryngeal lowering is connected with voicing, whereas in NB words with voiceless onsets were just as likely to become breathy as those with voiced onsets. In order to reconcile these facts with our general theory of registrogenesis, we are forced to hypothesise that pre- or proto-NB speakers were articulating their high and mid-high vowels with a markedly lowered larynx, conditioning breathy phonation, which then became analysed as a salient component of the bundle of features treated as +high.

The closest we have got so far to an explanation, one that has not been generally well received (e.g. see Brunelle 2005 for a critique), is that of Gregerson (1976). In Gregerson's scheme registers are modelled as primarily a function of tongue root advancement/retraction such that lax = [+ATR], tense = [-ATR], and the movement of the tongue root is assumed to have a direct association with the position of the tongue blade. For example, it is supposed that advancing the tongue root to enlarge the pharyngeal cavity in association with production of breathy voice inherently acts to raise the tongue blade and thus lower F1. Gregerson's theory gives primacy to the role of pharyngeal resonance in registers, and this runs directly counter to the primacy the standard model gives to the state of the larynx, but it does not discount an important role for ±ATR at the margins.

For the time being we are left to speculate that perhaps ancient NB speakers associated subtle acoustic correlates of an enlarged pharyngeal cavity, arising from the +ATR feature of high vowels. Speakers augmented these features by further enlarging with cavity with lowering of the larynx, yielding a more breathy phonation. In due course the phonation quality feature came to dominate the perceptual character of high and mid-high vowels and the register contrast was phonologized, facilitating subsequent and diverse vowel restructuring within the group.

Given the evident rarity of these NB developments (regardless of the correctness of our proposed explanations), we may reasonably speculate about the possible role of language contact. For example, we may wonder if NB speakers were in contact with register languages, and somehow mis-analysed what they were hearing and trying to pronounce, and carried those new habits over into their everyday speech. However, there is a strong difficulty with this idea, namely that we know that the development of registers in important neighbouring languages, principally Chamic and Khmer, occurred historically rather late, probably well after the bulk of their known influence on NB occurred. The great areal de-voicing swept into Indo-China rather late, from the 1700s (e.g. Ferlus 2011) whereas Khmer and Chamic power was dramatically declining in the immediate region from the 1300s through 1500s, robbing us of a potentially convenient if expedient hypothesis.

2. Two stories of atypical syllable restructuring

Laven and Nyaheun: restructuring toward monosylables

Laven and Nyaheun are two closely related West-Bahnaric languages spoken on the Boloven Plateau in southern Laos; the 1995 census counted around 40,000 Laven and 4,000 Nayheun. Laven is described in some detail in Jacq's (2001) MA thesis, the historical phonology of Laven and Nyaheun is analysed by Ferlus (1974) and Sidwell & Jacq (2003), a Nyaheun lexicon is published (Ferlus 1998) and grammar is briefly discussed by Davis (1973) and I have personally done field work on both languages. Although Laven and Nyaheun are closely related, they do not specially subgroup, and phonological and syntactic differences are so extensive that there are strong barriers to mutual intelligibility. My own field observation is that mutual comprehension between the speakers is a function of multilingualism more than from linguistic similarity. Additionally, my understanding is that there was a previous tendency for Nyaheun speakers to use Laven as in inter-language, while these days I observe Lao now dominantly employed in that purpose.

Both Laven and Nyaheun have restructured sesquisyllabic words into monosyllables, completely in the case of Nyaheun, and in a subset of sesquisyllabic lexicon in Laven. This quite unlike anything that has occurred in Lao or any other known contact language of the immediate area. In fact, Lao and Khmer robustly maintain sequisyllabic words, as do the Katuic languages of southern Laos which are the main other contact languages. Of course, Vietnamese and some Chamic languages have restructured sesquisyllables into monosyllables (see, for example Thurgood 1999, Brunelle 2005) but neither of these were geographically or temporally aligned to exert influence on the languages of the Boloven. What happened on the plateau in this respect was apparently quite local and specific.

Broadly, Nyaheun speakers restructured the entire lexicon into monosyllables, predominently by two processes:

- Assimilation, changing clusters into geminates, and
- Lention of prevocalic segments creating onsets with rising or plateaued sonority.

Laven, on the other hand, effected a partial restructuring:

- Lenition of rhotic and sibilant clusters into simple or complex segments with begin with devoiced transitions (CC, CrC > hC~Ç̥),
- Loss of prenasalization yielding preglottalized onsets,
- Reduction in the prominence of minosyllable vowels.

The first block of examples below highlights the creation of geminated onsets in Nyaheun, which includes both oral stops and nasals. For Laven we see reduction where initials in clusters are historically glottal, palatal or stop+rhotic. To illustrate the historical processes comparisons are also offered for Brao (another West Bahnaric language), Proto-West Bahnaric (Sidwell & Jacq 2003), Khmer and Stieng (South Bahnaric):

gloss	Laven	Nyaheun	Brao	pWBahnaric	other
'right (side)'	**hma**:	**m**:**a**:	**cəma**:	***cma**:	
'finger'	**hpuac**	**p**:**uac**	**tərpuac**	***trpuac**	
'buttock'	**hbo**:**k**	**p**:**o**:**k**	**tərpɑk**	***trbo**:**k**	Khmer **trɑpouk**
'bean'	**hta**:**k**	**t**:**a**:**k**	**hnta**:**k**	***hnta**:**k**	Khmer **sɑndaek**
'turtle'	**ʔtʌ**:**k**	**t**:**ɐ**:**k**	**ʔntə**:**k**	***ʔntə**:**k**	Khmer **ʔɑndaək**
'hoof'	**kʲɟo**:**p**	**c**:**ɔ**:**p**	**-**	***kɟ(ɔ/o)**:**p**	Stieng **kənɟɔ**:**p**
'a small bean'	**kɲɛ**:	**ɲ**:**ɛ**:	**-**	***kɲɛ**:	

Table: Laven-Nyaheun comparions (1)

The Laven initial sequences written with initial /h/ represent preaspirated oral stops and partly devoiced nasals. The nasals are normal length but voicing commences about half way through the duration (these are described by Jacq 2001:62-65). The Nyaheun geminates are very long, with the stops around 200+ ms, and the nasals typically more than 300ms, actually a little more than twice as long as the equivalent unmarked segments.

In the next block of examples (below) we see Nyaheun words where lenition of prevocalic segments creates sequences with rising sonority, although in some cases there are also doublets with geminated onsets. Where the historical initial was ***s**, the outcome in Laven is a partly devoiced onset. Where the initial is a stop there is not a great change, but

it is noticeable that the epenthetic vowel which is normally a feature of these clusters is acoustically much less prominent than in other West Bahnaric languages, such as Brao. Note, in several examples below clusters are reconstructed with medial glottal stops; this was tentatively offered by Sidwell & Jacq (2003) to explain differential outcomes of structurally similar clusters in Laven and Nyaheun. Presently I reserve judgement on this hypothesis, and speculate that these items reflect a stratum of loans from Old/Middle Khmer, since it underwent a change in which prevocalic voiceless stops became implosives (Ferlus 1992 dates the shift to implosion to the 13th~14th C) although this would not explain why forms with prevocalic ***s** are also affected. Examples:

gloss	Laven	Nyaheun	Brao	pWBahnaric	other
'mortar'	**tʰpal**	**dwaw**	**təwa:w**	***tʔpal**	Khmer **tɓal**
'crab'	**kʰta:m**	**gra:m**	**kəda:m**	***kʔta:m**	Khmer **kɗa:m**
'ghost'	**kʰsɔk**	**gjɔk**	**kəjɑk**	***kʔsɔk**	
'to wash'	**hta:**	**hra:**	**səda:**	***sʔta:**	
'day'	**tʰŋaj**	**nie, n̥e:**	**təŋaj**	***tŋaj**	Khmer **tŋaj**
'year'	**kʰmɔ:**	**ŋwɔ:, m̥ɔ:**	**kəmɑ:**	***kmɔ:**	Katu **kamɑ:**
'house'	**n̥ə:m**	**n̥̊ra:m**	**hna:m**	***sna:m**	SurinKhmer **sna:m**

Table: Laven-Nyaheun comparions (2)

It is apparent that Nyaheun continues to an extreme a tendency that is partly manifest in Brao, another West Bahnaric language. In Brao there is a limited lenition in which clusters of stop+stop and stop+fricative see the second element become a voiced approximant. Yet in Nyaheun all initial clusters that where not already of the type stop+approximant were restructured either to that pattern or into geminates.

How to characterise the phonological processes? In respect of Nyaheun we recognise that there is a phonotactic simplification, such that only three types of onsets became permissible in the language: 1) single consonantal segments, 2) long level sonority (geminate), and 3) rising sonority voiced onsets (voiced stop/nasal+approximant). Where a voicing change has occurred, it has involved an increase in voicing. In contrast, Laven has increased the overall complexity of possible onsets. While retaining a large proportion of otherwise typically Mon-Khmer clustered onsets, it also restructured a marked subset of clusters into preaspirates and partly devoiced nasals. The effect of the restructuring was to depress sonority at the left edge of onsets, thus creating a rising sonority over the onset. This can be characterised as a fortition in restructured onsets.

In a broad sense we might say that both Laven and Nyaheun have been restructuring characteristically complex AA initials into more compact rising sonority onsets, consistent with broad SEAsian trends. Change was most dramatic in Nyaheun, to an extent that is directly comparable to, for example, Vietnamese, which is known to have radically reduced all clustered and sesquisyllablic onsets (e.g. Gage 1985, Ferlus 1992). Consistent with our well established narrative of contact driven change, it is widely recognised that Vietnamese restructured in circumstances of prolonged intimate contact with (especially southern) Chinese (see Alves 2001 for a survey).

Yet no such history of prolonged contact conditioning phonological change can be invoked for either Nyaheun or Laven, both closely related and in intimate contact with each other for hundreds of years they manifest very different phonological histories. There

is nothing the phonologies of Lao, Khmer, or other known contact languages that models the changes we have discussed above. Additionally we do know that especially Laven has been under very strong Lao influences for several hundred years (see below).

Wider areal/historical context

There are good linguistic indications that the Boloven Plateau was closely integrated into the Angkorian empire, but had little or no interactions with Champa (see Sidwell & Jacq 2003). Consequently, with the collapse of Angkorian power in the 1300s, the locals enjoyed considerable autonomy until the Laotians and Siamese began settling in and exerting control over the area in the 1700s. But 1870 Harmand, the first European to visit the Boloven Plateau, was able to report that the Laven women were so culturally Laoisized that they dressed as Laos, wore their hair as Laos, and even used lipstick purchased in the markets of Pakse with money made from their plantations (Harmand 2002). Harmand also described meeting Nyaheun on the Plateau, and the sketchy details are consistent with the two groups living in close contact much as they do now.

It is apparent that language contact between Lao and Laven has been so extensive that we may almost characterise Laven as a relexified form of Lao; for example, a comparison of Lao clause structure as described by Enfield (2007:171) and Laven as described by Jacq (2001) shows almost complete congruity. In fact, the minor discrepancies between the two descriptions vanish under close examination:

- the Laven future marker corresponds to the Lao irrealis marker,
- the positions immediately following Subject/Agent in Laven can be filled by particles and/or adverbials equivalent to Lao,
- although the Lao Achievement morpheme (**dajø**) slot has no equivalent in Jacq's Laven, I have heard Laven speakers use a equivalent (**bic** 'obtain/achieve') in the same way,
- many Laven functors and adverbs are borrowed from Lao or Isaan Thai, and greetings/leave takings are calqued (some of these are discussed by Jacq (2001).[3]

However, Nyaheun presents a different situation, especially being much more flexible in word order. Compare:

Obligatory Lao, Laven word order						
S/A	NEG.	ASP.		MOD.	V(O)	
+ Available Nyaheun order						
S/A	NEG.	ASP.		MOD.	V(O)	
	NEG.	ASP.	S/A	MOD.	V(O)	
		ASP.	S/A	MOD.	V(O)	NEG.
S/A		ASP.		MOD.	$V_{INTRANS}$	
		ASP.		MOD.	$V_{INTRANS}$	S

Figure: Variation in Nyaheun word order (recorded by Sidwell) compared to Lao and Laven

[3] It can be somewhat confronting to first encounter Nyaheun greetings which translate literally as "you are not sick?"or "Aren't you dead yet?" Davis 1973), while the Laven /hbai rɨp/ "good health" transparently calques the Lao /saba:j3 di:3/.

In my own Nyaheun text collection I find many examples of Nyaheun clauses with distinctive word order, such as intransitive subjects after verbs, subject immediately before the modal and main verb so that negation and the aspectual marker precede the subject, even the negator can be relegated to the end of the clause (without it becoming a question marker as in Lao).

So far as I can tell, this variability in Nyaheun word order does not impact on information structure, rather it facilities considerable stylistic manipulation. For example, speakers value a 2-4 beat, especially with rhyming/alliterating pairs, often realised with reduplications, deletions and reordering elements within the generous constraints indicated above. This kind of flexibility is apparently not available in Laven, which is apparently syntactically remodel after Lao, an example of metatypy (in the sense of Ross 2006).

In other important ways both Laven and Nyaheun are untouched by Lao; there is no apparent accommodation to Lao style tonal system, and the robustly complex vocalism and range of consonant articulations remain essentially intact without collapsing them to the simpler phonotactics of Lao. Apparently, Laven and Nyaheun have mixed for a considerable time, and both have been exposed to Lao, with Laven significantly affect, and in these conditions speakers have effected very different phonological and syntactic restructuring.

Gta' (Munda) creating initial clusters/sesquisyllables

Anderson (2008) presents a sketch grammar of Gta', a small Munda language of southern Orissa, India with less than 4,500 speakers. Several distinct varieties are spoken, including a Hill Gta' and a Plains Gta'. The language and people are also called Didayi or Didei. According to Anderson the classification is open to question, although it is generally regarded as belonging to the South Munda sub-family. A Didayi dictionary by Chatterjee et al. is available electronically form the Stampe digital Munda archive.[4] Forms labelled Gta' in the discussion that follows are extracted from Anderson (2008) while Didayi forms are from Chatterjee et al.

As Anderson reports, Gta' is phonotactically somewhat Mon-Khmer looking:

> Gta' is an unusual language from the perspective of syllable structure or phonotactics It has an enormous number of 'clusters' found in word-initial position but a restricted number of consonants found in coda position. A small number of words with syllabic nasals and prenasalized stops may also be found.
> [.....]
> This feature of Gta' appears to be very similar to syllable structure constraints found in other AA languages that are distant relations of Gta'. However, rather than representing a retention of an archaic phonotactic feature directly inherited from Proto-Austroasiatic lost in all other Munda languages, this feature of Gta' is more likely to be a pseudo-archaism, [....].
> Although, so-called sesquisyllabic words are found frequently in Mon-Khmer languages, their presence in Gta' arose through, among other processes, the loss of unstressed vowels in word-initial syllables The specific form of the vowel found in such languages as Gutob or Remo is not predictable based on the Gta' form, but the reverse is largely true.
> (Anderson 2008:684-5)

[4] http://www.ling.hawaii.edu/austroasiatic/AA/Munda/Dictionaries/Gta_Chatterji

To illustrate something of these phonotactics before going further I present some lexical examples, with other South Munda comparisons and pMK (effectively pAA) reconstructions by Shorto:

Shorto (2006) PMK	Gloss	Gta'	Didaji	Remo	Gutob
***ɓa:r** (#1562)	'two'	**mbar**	**mbar**	**bar**	**ba:r-ɟu:**
***bri:ʔ** (#181)*	'uncultivated land'	**bri**	**-**	**biri**	**bjroŋ**
***cɔʔ** (#41)	'dog'	**gsuʔ**	**gusu**	**guso:**	**guso?**
***ci(ə)m** (#1324)**	'chicken'	**gsæŋ**	**gesaiŋ**	**gi-siŋ**	**gi-siŋ**
***cʔa:ŋ** (#488)	'bone'	**ncia**	**ncja**	**saŋ**	**sisaŋ**
***ɗa:k** (#274)	'water'	**nɖiaʔ**	**ɖia**	**ɖag**	**da**
***ɗak** (#330)	'trap'	**dnoʔ**	**-**	**ɖonok**	**-**
***ɟla(i)ŋ** (#740)	'long/tall'	**clæʔ**	**-**	**sileŋ**	**sileɟ**
ɟmu:l** (#1777)	'seed'	**cmu**	**-**	**sumu**	**-**
***ɟru:ʔ** (#172)	'deep'	**cri**	**ciri**	**siri**	**gaɽia**
***ɟu(ə)ŋ** (#538I)	'foot'	**nco**	**co**	**suŋ**	**susuŋ**
***kaʔ** (#16)	'fish'	**haʔɽo**	**-**	**aʔ**	**aʔɖoŋ**
***klaʔ** (#197)	'tiger'	**ŋku**	**-**	**kisa:**	**gikkil**
***ks(i)ʔ** (#246)	'rope'	**ghæʔ**	**-**	**gieʔ/gije**	**geʔ**
***mat** (#1045)	'eye'	**mmwaʔ**	**moa**	**m'o:**	**mo:**
***muh** (#2045)	'nose'	**mmu**	**mu**	**se:-mi**	**mi**
***ris** (#1927)	'root'	**nɖræʔ**	**nɖrɛ**	**regi**	**-**
***rk(aw)ʔ** (#1820)	'uncooked rice'	**rkoʔ**	**-**	**rŋku**	**rukuʔ**
***ru(ə)j** (#1534)	'fly'	**n(d)rwe**	**conɖroe**	**-**	**uroj**
sŋiʔ** (#37)*	'village'	**hni**	**hini**	**suŋ**	**-**
***ti:ʔ** (#66)	'hand'	**nti, tti**	**ti**	**ti**	**titi**
***tŋi:ʔ** (#31)	'sun'	**sni**	**sini**	**siŋi**	**siŋgi ~ sĩĩ**

Table: Proto-Angkuic phonetic tones

* 'forest', ** 'bird'; *** 'to dibble'; **** 'house'

Anderson is confident that these initial clusters in Gta' are secondary, observing that, "The specific form of the vowel found in such languages as Gutob or Remo is not predictable based on the Gta' form, but the reverse is largely true." (p.685). There are additional factors that point strongly toward Gta' being innovative rather than conservative:

- historically AA lexicon have clustered onsets in Gta', even when proto-forms did not, with gemination and prenasalisation being common augments;
- the vowel system is very simple, including the lack of a length contrast in monophthongs.

The simple vocalism is particularly Munda in character; we assume that the original AA vocalism, with its richer vowel inventory and quantity distinction, was simplified in the restructuring of protoMunda, leaving some traces in the accentual system. This is largely captured by Donegan & Stampe's "rhythmic drift" model of syllable canon change. The diagrammatic representation from their seminal 1983 paper is reproduced below:

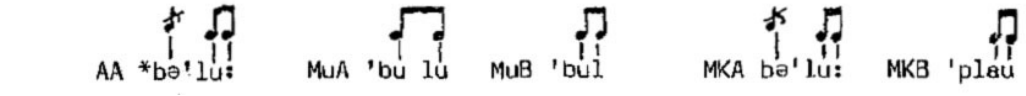

Figure: Rhythm and opposite rhythmic drift in AA
(from Donegan & Stampe 1983:346).

The model requires further explanation, in particular it must be noted that it should not be read as suggesting that protoMunda shifted primary stress leftward within the phonological word cannon. Munda languages still predominantly put stress on the second syllable of disyllabic words (with exceptions relating to Munda languages with quantity sensitive systems such as Mundari, Anderson:personal communicatiuon 6/10/2012). But there was a prosodic shift that changed sesquisyllables (and even some monosyllables) into disyllables, creating the equivalent of the MundaA stage in the schema above, which we see roughly in Didayi examples above such as **ciri** 'deep', **sini** 'sun'etc. This is the Didayi of Chatterjee et al., with its penultimate copy vowels and non-geminate onsets, and has the appearance of being transitional to Gta'. Broadly, it is apparent that Gta' has restructured words of both CV´CV(C) and CV(C) types into CCV(C), in what seems to be a very local innovation.

In terms of language contact, Anderson remarks that there is evidence of Remo influence, and borrowings from Dravidian, Desiya (the local Indic language) are also noted. None of these influences can be credited as providing a model for the changes in Gta', and we are left wondering how and why - in the face of evidently overwhelming areal pressure - Gta' restructuring took place. As is evident, it was not just a matter of losing unstressed vowels, but there was also a comprehensive addition of augments that created a general pattern of initial clusters/geminates that are not supported etymologically.

Concluding remarks

In this paper I have discussed several examples of phonological restructuring in languages from across the breadth of the AA phylum, each apparently running counter in some ways to local and wider areal trends in their details. Yet at the same time, some of the changes may be broadly characterised as consistent with the 'spirit' of areal tendencies (tonogenesis, monosyllabism etc) such that in a typological overview they might just be counted as yet more data supporting an argument for areality or similar.

Looking at the case of Angkuic tonogenesis, it would be easy to simply count the languages as being tonal and lacking length contrast, and assume that speakers had achieved this by imitating the phonology of Shan, yet this cannot be the explanation. North Bahnaric languages superficially have a register system phonetically similar to Mon, Cham or Middle Khmer, yet neither they nor other important language of the area could have been a model for imitation in this respect. Laven and Nyaheun have both taken to creating monosyllables, and yet show no commonalities in how they have done it. And Gta' speakers actively eschewed CV(C) syllables in direct contradiction to what all indications would lead us to expect.

This leads me to make 2 observations:

1. Linguistic typology is at great risk of making well intentioned but misleading generalisations if insufficient attention is paid to the correct characterization of data; and

2. None of the above cases of language change can be explained by the ".....imitation of the observable behaviour of others," as Post (2011:219) characterises contact driven change. Rather, speakers seem to have adopted behaviours that are in contradistinction to what they observe in others.

Having regard to social correlates in so far as they are recognised in this short study, we can tentatively suggest that the more marked examples are found where speaker communities are in a subordinate or marginal relation to more dominate groups, and the population is very small, only a few thousand or so. Perhaps these are speaker efforts to assert distinctive or emblematic characteristics in the face of external pressures. In the circumstances, would it be too bold to talk about *contact driven divergence*?

References

Alves, Mark J. 2004. *A grammar of Pacoh: a Mon-Khmer language of the central highlands of Vietnam*. Canberra. Pacific Linguistics.

Alves, Mark J. 2001, "What's so Chinese about Vietnamese?", in *Papers from the Ninth Annual Meeting of the Southeast Asian Linguistics Society*, ed. G.W. Thurgood, pp. 221-242. Arizona State University, Program for Southeast Asian Studies.

Anderson, Gregory (ed.). 2008., *The Munda Languages.* London/New York: Routledge.

Bradley, D., R. J. LaPolla, B. Michailovsky and G. Thurgood, Eds. 2003. *Language Variation: Papers on Variation and Change in the Sinosphere and in the Indosphere in Honour of James A. Matisoff*. Canberra, Pacific Linguistics.

Brunelle, Marc. 2005. Register in Eastern Cham: phonological, phonetic and sociolinguistic factors. Doctoral Dissertation, Cornell.

Comrie, Bernard. 2007. Areal Typology of Mainalnd Southeast Asia: What We Learn from the WALS Maps. Manusya 13.18-47

Cooper, James & Nancy Cooper. 1966. Halang Phonemes. *Mon-Khmer Studies* 4:87-98.

Davis, J.J. 1973, "Notes on Nyaheun grammar", in *The Mon-Khmer Studies Journal*, vol. 4, pp. 69-75.

Davis, John J. 1968. *Nyaheun Phonemes.* unpublished manuscript. Copy held at David Thomas Library (Bangkok) file 495.955(6).

Donegan, P. J. and D. Stampe. 1983. 'Rhythm and the holistic organization of linguistic structure.' *Chicago Linguistic Society* 19(2: Papers from the Parasession on the Interplay of Phonology, Morphology and Syntax): 337-353.

Donegan, P. J. and D. Stampe. 2004. 'Rhythm and the synthetic drift of Munda.' In R. Singh, Ed. *The Yearbook of South Asian Languages and Linguistics*. Berlin/New York, Mouton de Gruyter: 3-36.

Donegan, P. J. and D. Stampe. 2004. 'Rhythm and the synthetic drift of Munda.' In R. Singh, Ed. *The Yearbook of South Asian Languages and Linguistics*. Berlin/New York, Mouton de Gruyter: 3-36.

Enfield, N. J. (2007). *A grammar of Lao*. Berlin: Mouton de Gruyter.

Fagan, Joel L. 1988. Javanese intervocalic stop phonemes. *Studies in Austronesian Linguistics* 76:173–202.

Ferlus, Michel. 1974. Delimitation des groupes linguistiques austroasiatiques dans le centre indochinois. *Asie du Sud-Est et Monde Insulindien* 5.1:15-23.

Ferlus, Michel. 1992, "Essai de phonétique historique du khmer", in *The Mon-Khmer Studies Journal*, vol. 21, pp. 57-89.

Ferlus, Michel. 1998. *Nhaheun-French-English Lexicon*, München, Lincom Europa.

Ferlus, Michel. 2011. 'Toward Proto-Pearic: Problems and Historical Implications.' In Sophana Srichampa, Paul Sidwell and Kenneth Gregerson (eds.) *Austroasiatic Studies: papers from ICAAL4. Mon-Khmer Studies Journal Special Issue No. 3, part 1*. Dallas, SIL International; Salaya, Mahidol University; Canberra, Pacific Linguistics. pp.38-51.

Gage, W.W. 1985, "Vietnamese in Mon-Khmer perspective", in *Southeast Asian Linguistic Studies Presented to André-G. Haudricourt*, ed. S. Ratanakul et al., pp. 493-524. Institute of Language and Culture for Rural Development, Mahidol University.

Gordon, Matthew, and Peter Ladefoged. 2001. Phonation types: a cross-linguistic overview. *Journal of Phonetics* 29:383–406.

Gregerson, Kenneth J. 1976. "Tongue-Root and Register in Mon-Khmer." Philip N. Jenner, Laurence C. Thompson, and Stanley Starosta (eds.). *Austroasiatic Studies, Two Volumes.* Honolulu: University of Hawaii (Oceanic Linguistics, Special Publication, No. 13), I:323-369

Hall, Elizabeth. 2010. *A Phonology of Muak Sa-aak*. MA thesis, Payap University Thailand.

Harmand, F. J. 2002. *Laos and the Hill Tribes of Indochina: Journeys to the Boloven Plateau* (translated by Walter Tips). Bangkok, White Lotus.

Haudricourt, A.-G. 1965. "Mutation consonantique en Mon-Khmer." *Bulletin de la Société Linguistique de Paris* 60:160-72.

Huffman, Franklin (1978). *Khmer-English dictionary*. New Haven and London, Yale University Press.

Huffman, Franklin E. 1985. Vowel Permutations in Austroasiatic Languages. *Linguistics of the Sino-Tibetan Area: The State of the Art.* Pacific Linguistics Series C-No.87. Canberra: Australian National University, pp141-45.

Jacq, Pascale. 2001. *A Description of Jruq (Loven): a Mon-Khmer language of the Lao PDR*. MA thesis, Australian National University.

Kingston, John. 2011. Tonogenesis. In *Blackwell Companion to Phonology*, v. 4, edited by Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume, and Keren Rice, pp. 2304-2333.

Ladefoged, Peter, and Ian Maddieson. 1996. *The Sounds of the World's Languages*. Oxford: Basil Blackwell.

Lindell, Kristina, Jan-Olaf Svantesson and Damrong Tayanin. 1978. Two dialects of the Romeet (Lamet) language. *Cahiers de Linguistique, Asie Oriental* 4:5-22.

Manley, Timothy M. 1972. *Outline of Sre Structure*. Oceanic Linguistics Special Publication No.12. Honolulu, University of Hawaii Press.

Matisoff, James. 1991. 'Sino-Tibetan linguistics: Present state and future prospects.' *Annual Review of* Anthropology 20:469-504.

Mortenson, David (2006). "Tonally conditioned vowel raising in Shuijingping Hmong." Handout from the LSA 80th Annual Meeting, Albuquerque, New Mexico. January 6, 2006.

Narumol Charoenma 1980, "The sound systems of Lampang Lamet and Wiangpapao Lua", Master thesis, Mahidol University.

Oostendorp, Colin J. Ewen, Elizabeth V. Hume, Keren Rice (eds.). 2011. The Blackwell Companion to Phonology. Wiley-Blackwell.

Post, Mark W. 2011. 'Prosody and typological drift in Austroasiatic and Tibeto-Burman: Against "Sinosphere" and "Indosphere".' In Sophana Srichampa, Paul Sidwell and Kenneth Gregerson (eds.) *Austroasiatic Studies: papers from ICAAL4. Mon-Khmer Studies Journal Special Issue No. 3*. Dallas, SIL International; Salaya, Mahidol University; Canberra, Pacific Linguistics. pp.198-221.

Ross, Malcolm D. 2006. "Metatypy". In K. Brown. *Encyclopedia of language and linguistics* (2nd ed.). Oxford: Elsevier.

Scott, James George and J.P. Hardiman. 1900. *Gazetteer of Upper Burma and the Shan States* 1.1. Rangoon, Superintendent, Government Printing.

Shorto, Harry L. 2006. *A Mon-Khmer Comparative Dictionary.* Canberra, Pacific Linguistics 579.

Sidwell, Paul & Pascale Jacq. 2003. *A Handbook of Comparative Bahnaric: Volume 1, West Bahnaric*. Canberra, Pacific Linguistics 551.

Sidwell, Paul. 1998. *A Reconstruction of Proto-Bahnaric*. PhD thesis, University of Melbourne.

Sidwell, Paul. 2002. Genetic classification of the Bahnaric languages: a comprehensive review. *Mon-Khmer Studies* 32:1-24.

Sidwell, Paul. 2005. *The Katuic Languages: classification, reconstruction and comparative lexicon*. Munich, Lincom Europa.

Sidwell, Paul. 2011. *Proto Bahnaric.* http://sealang.net/monkhmer/database/.
Smith, Kenneth, D. 1967. *A Phonological Reconstruction of Proto Central North Bahnaric.* University of North Dakota Summer Institute of Linguistics Workpapers.
Smith, Kenneth, D. 1972. *A phonological reconstruction of Proto-North-Bahnaric.* Dallas, Language Data Series, Summer Institute of Linguistics.
Smith, Kenneth, D. 1979. *Sedang Grammar*. Pacific Linguistics. Series B–50. Canberra, Australian National University.
Svantesson, J.-O. 1991, "Hu - a language with unorthodox tonogenesis", in *Austroasiatic Languages, Essays in honour of H. L. Shorto*, ed. J.H.C.S. Davidson, pp. 67-80. School of Oriental and African Studies, University of London.
Svantesson, Jan-Olof. 1988. U. *Linguistics of the Tibeto-Burman Area*, 11.1:64-133.
Svantesson, Jan-Olof. 1989. Tonogenic Mechanisms in Northern Mon-Khmer. *Phonetics*, 46:60-79.
Thurgood, Graham. 1999. *From Ancient Cham to Modern Dialects: two thousand years of language contact and change.* Oceanic Linguistics special Publications No. 28. Honolulu, University of Hawaii Press.
Thurgood, Graham. 2007. Tonogenesis revisited: revising the model and the analysis" *Studies in Tai and Southeast Asian Linguistics,* edited by Jimmy G. Harris, Somsonge Burusphat, and James E. Harris. Bangkok: Ek Phim Thai Co. Pp. 263-291.
Thurgood, Graham. 2002. Vietnamese and tonogensis. *Diachronica* 19. 333–363
Watkins, Justin (2002). *The phonetics of Wa: experimental phonetics, phonology, orthography and sociolinguistics*. Canberra, Australian National University