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

The Austroasiatic central riverine hypothesis[1]

The paper considers the vexing issues of the homeland and dispersal of the Austroasiatic languages. A critical analysis finds little firm support for nested sub-groupings among a dozen recognised branches, while lexical analyses suggest a long-term pattern of contact and convergence within mainland Southeast Asia. These facts are interpreted as consistent with a stable long-term presence in Indo-China, probably centred on the Mekong River. The most geographically distant branch — the Munda of India — is treated as a highly innovative outlier, and the evolution of Munda root structure is reconstructed, consistent with this theory.

Keywords: Austroasiatic, Munda, comparative method, language relationship, lexicostatistics.

Introduction

The question of localizing the Austroasiatic (AA) homeland is a crucial and increasingly topical one for scholars interested in the prehistory of Mainland Southeast Asia. A geneological language classification can inform discussion of this point, since the correlation of geography and phylogeentic distribution may permit inferences concerning migration routes, contacts, and time depths. In this context, it is both intriguing and frustrating that, after more than a century of comparative AA studies, scholars have yet to present an explicitly justified and comprehensive internal genetic classification of the phylum (see Sidwell 2009 for a general discussion).

For sure, there are various proposals in print, and in unpublished sources such as dissertations, conference presentations, and manuscripts circulating informally. But when these disparate sources are tracked down, compared, and analysed, it becomes abundantly clear that there is no scholarly consensus on:
— the relations between AA branches,
— the age or diversity of AA,
— an appropriate program for addressing these issues

Consequently, the field is yet to significantly benefit from multidisciplinary research. Scholars eager to pursue the synthesis of archaeology, genetics and linguistics are exasperated, such as recently expressed by Roger Blench:

Austroasiatic languages are the most poorly researched of all those under discussion. Many are not documented at all and some recently discovered in China are effectively not classified. The genetics of Austroasiatic speakers are almost unresearched. Austroasiatic is conventionally divided into two families, Mon-Khmer (in SE Asia) and Munnda (in India). Diffloth (2005, 79) now considers Austroasiatic to have three primary branches but no evidence for these realignments has been published. Indeed Austroasiatic classification has

been dogged by a failure to publish data, making any evaluation of competing hypotheses by outsiders a merely speculative exercise. (Blench 2008, 117–118)

In this paper it is argued that the thirteen[2] AA branches radiated more or less equidistantly from proto-AA. Consequently, since the greatest number of AA branches is spoken along an axis that runs roughly southeast to northwest along the middle course of the Mekong river, it is reasonable to suggest that the AA languages dispersed along and out from that axis. This theory is provisionally called the *Austroasiatic Central Riverine Hypothesis.*

Recent thinking — India and China

Since the second half of the 1990s there has been a resurgence of interest in AA linguistics and the homeland question. Among the various suggestions that have been offered over the years, there are two broad lines of inquiry that have been especially emphasized:

1. a western origin, in eastern India or about the Bay of Bengal, or
2. a northern origin, in central or southern China
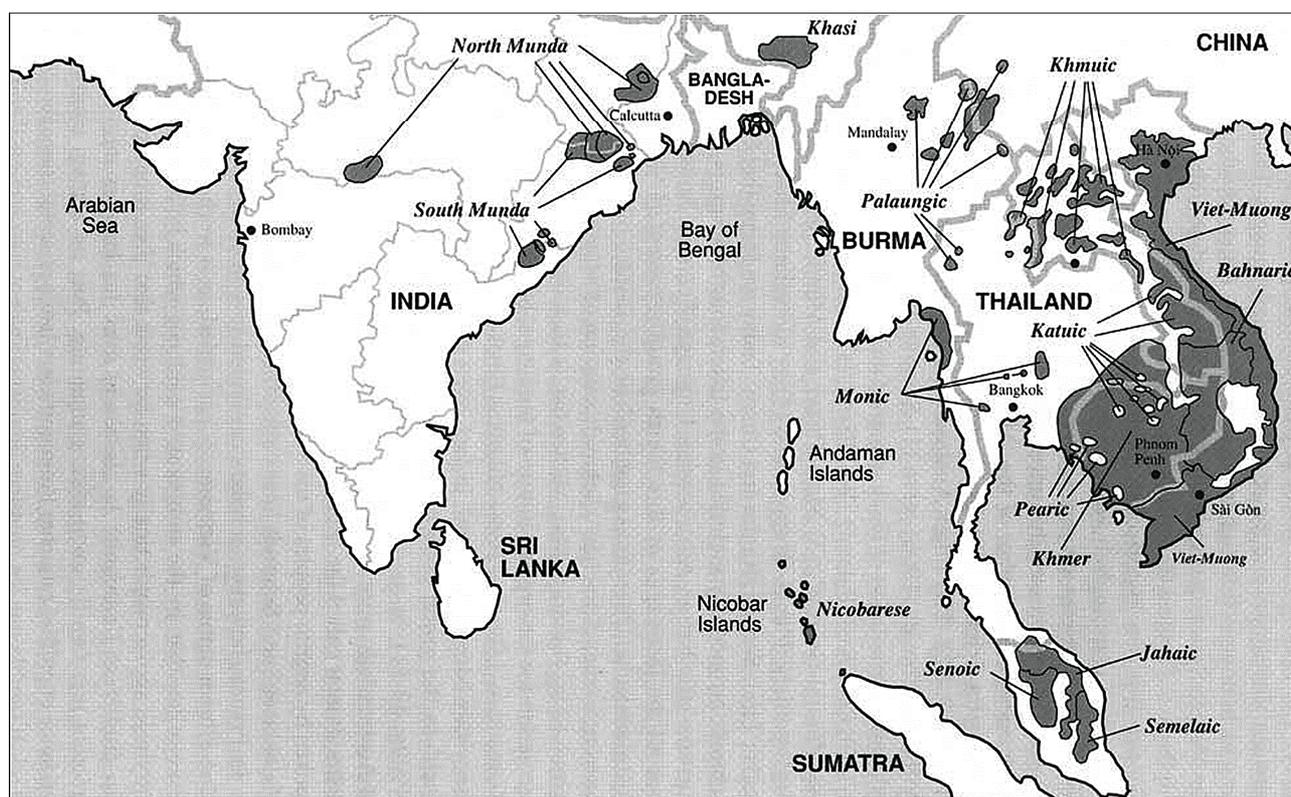
Figure 1: *Map of Austroasiatic languages (van Driem 2001)*

Gerard Diffloth (2005 and elsewhere) has argued that, on the basis of floral and faunal terms, the AA lexicon rules out a temperate zone (i.e. China) in favour of a tropical homeland. Assuming a primary split between Munda and Mon-Khmer within AA, he proposes a location proximal to the Bay of Bengal (having earlier proposed the Burma-Yunnan border zone).

---

[2] The figure thirteen assumes the following branches: Munda, Khasi, Palaungic, Khmuic, Mangic, Vietic, Katuic, Bahnaric, Pearic, Khmer, Monic, Aslian, Nicobarese.

George van Driem has been actively promoting this view in print (especially his 2001 handbook). Van Driem makes much of the literature which poports to identify an old Munda substrate in Vedic etc. (especially Kuiper 1955, 1967, 1991, and recently Witzel 1999).

By way of contrast, Peiros (1989, 1998) and Peiros & Shnirelman (1998), have asserted that the AA lexicon indicates a non-tropical, non-coastal location. Combining this idea with data about the mid-Holocene climate, and supposed lexical isoglosses with Hmong-Mien languages (some two dozen or more supposed resemblances), they propose an origin on the mid-Yangtse river. This view dovetails with philologists' suggestions of various AA etymologies for words of uncertain origins in classical languages of China. Mei & Norman (1976) especially are widely cited as proposing AA loans into Chinese, including even the name of the Yangste itself. Schuessler (2007) more recently proposes hundreds of Mon-Khmer-Old Chinese comparisons, so many that he writes (p.4): "When pursuing OC and TB/ST etyma down to their roots, one often seems to hit Austroasiatic bedrock, that is, a root shared with Austroasiatic."

Generally these philological arguments have the following character: there are a handful of close phonetic and semantic matches which are intriguing, plus there are scores or more of vague resemblances which cannot be organised into any system of regular correspondences. Unfortunately this is what one expects to find when comparing unrelated but typologically similar languages. Presently we have no clear reasons to favour either the sinophiles or the indophiles, indeed shouldn't we begin from the premise embedded in the traditional philologist's wisdom that the search for Latin etymologies should begin on the Tiber? After all, the various philological arguments suppose localising proto-AA in places where little or none of the diversity of the AA phylum is presently found. There are no AA speaking communities along the middle Yangste or in China's south-eastern provinces. There are no AA speaking communities around the Bay of Bengal, and the closest reasonably diverse group — the Munda — may not be so especially diverse as to justify giving their position any particular weight (see discussion below).

For about a century now, especially since Sapir (1916) it has generally been recognised that, since diversity within a language group tends to increase over time, a region of higher diversity is likely to have been settled longer, and therefore is more likely to be, or be proximal to, the homeland. Dyen (1956) and Diebold (1960) specifically formalized this (and related ideas) under the heading of 'Migration Theory', and this approach is applied here. The efficacy of the theory is readily confirmed in the real world; for example the south of Vietnam was settled by Vietnamese speakers more or less since the formal incorporation of Champa into Vietnam from 1693, and the diversity of Vietnamese dialects in the south is low, but markedly higher in the north, especially for example around Vinh.

The present task is to look at the facts of the AA family and decide whether we can fairly characterize its phylogenetic diversity. While it is uncontroversial that the great bulk of AA languages are spoken in what I am calling the central zone, it is another matter whethar they represent the greater proportion of coordinate branches. Consequently it is the issue of the higher branching structure of AA that one must focus..

AA Classifications since Pinnow (1959)

Presently the view is widely received that the AA phylum is composed of two coordinate families; e.g.:

> The primary split in the family is between the Munda languages in central and eastern India and the rest of the family. (Anderson 2006, 598)

Although a somewhat different conception is also reported, e.g:

> The Austroasiatic language family is conventionally divided into three branches or sub-families, viz. the Munda, the Nicobarese and the Mon-Khmer languages. (van Driem 2001, 262)

Both of these propositions give a special status to Munda, an idea that can be traced directly to the works of Pinnow (1959, 1960, 1963 etc.). In Pinnow's 1960 paper Munda was charactersied as preserving ancient morphological complexities that find only murky traces in other AA groups, such as apparent alternations of final consonants in Khmer. Shortly after in 1963 he presented the classification reproduced here at Figure 2, which effectively divides the AA into Western (Munda) and Eastern (Khmer-Nicobar) families, with Mon-Khmer further divided into Nicobarese and 'Palaung-Khmer' sub-families. Thus we see the origins of both views quoted above. Earlier studies, such as Schimdt (1906) actually gave Munda a much lower status, sub-grouping it with languages of Indo-China.

Western group (Nahali-Munda)
   (*A*) West : Nahali (?)
   (*B*) East : Munda
      (*a*) North
         Kherwari (Santali, Mundari, Korwa, etc.)
         Kurku
      (*b*) South
         1. Central : Kharia, Juang
         2. South-East : Sora, Pareng, Gutob, Remo
Eastern group (Khmer-Nicobar)
   (*A*) West : Nicobarese (Nancowry, Car, etc.)
   (*B*) East : Palaung-Khmer
      (*a*) West : Khasi
      (*b*) North : Palaung-Wa (Palaung, Wa, Riang, Lawa, etc.)
      (*c*) East : Mon-Khmer (Mon, Khmer, Bahnar, Sre, etc.)
      (*d*) South : Malacca
         1. Sakai
         2. Jakud
         3. Semang

Figure 2: *Austroasiatic classification by Pinnow (1963)*

However, Pinnow's approach was more typological than historical, and he made an explicit statement to the effect that his Eastern (Khmer-Nicobar) group might not be a single phylogenetic division in opposition to Munda; but this caution appears to have been almost entirely ignored in subsequent scholarship.[3] Subsequently the structure so clearly communicated by Pinnow's figure came to be widely received.

In the following decade Diffloth's (1974) Encyclopedia article became especially influential. There Diffloth presented the modified Pinnow model with three coordinate families (Munda, Nicobarese, Mon-Khmer). His Mon-Khmer family was further sub-divided according to the lexicostatistical studies of Thomas and Headley (1970 especially). Later (e.g. 1979, 1999) Diffloth merged Nicobarese into Mon-Khmer as a sister of Aslian, and more recently (2005) promoted Khasi-Palaungic-Khmuic to form a 'new' third coordinate family.

Such classifications characterize Munda as comparable in terms of historical diversity to the rest of AA, and thus suggest that the geographical centre of that diversity is unlikely to have been in Southeast Asia but further to the west.

Presently it is not possible to assess Diffloth's classifications, since he has not made his data and arguments available. Still, we are able to address the most salient aspect — whether

---

[3] I have found no explicit recognition of this point in my reading of relevent literature.

it is justified to treat Munda as a primary branch which coordinates with just one or two other primary branches.
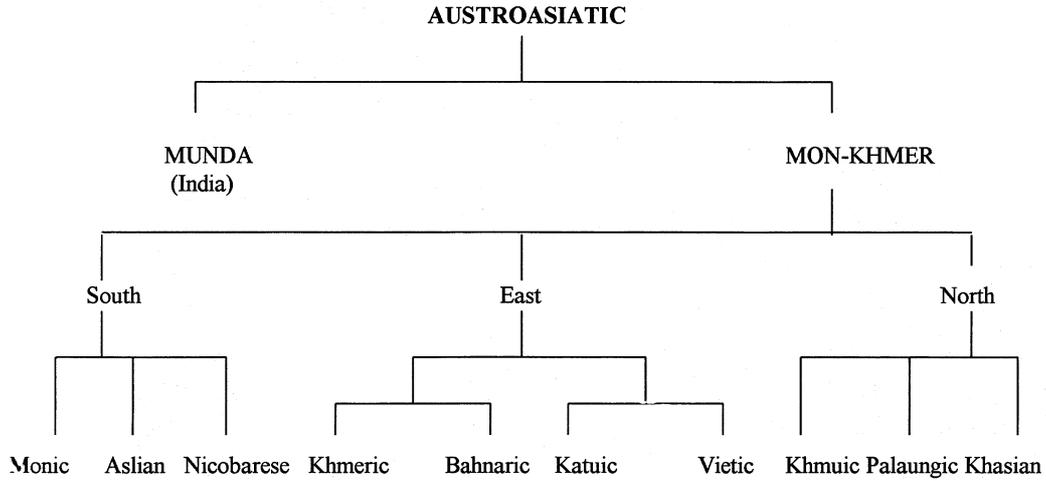
Figure 3: *Austroasiatic classification by Diffloth (reproduced from Chazée 1999)*
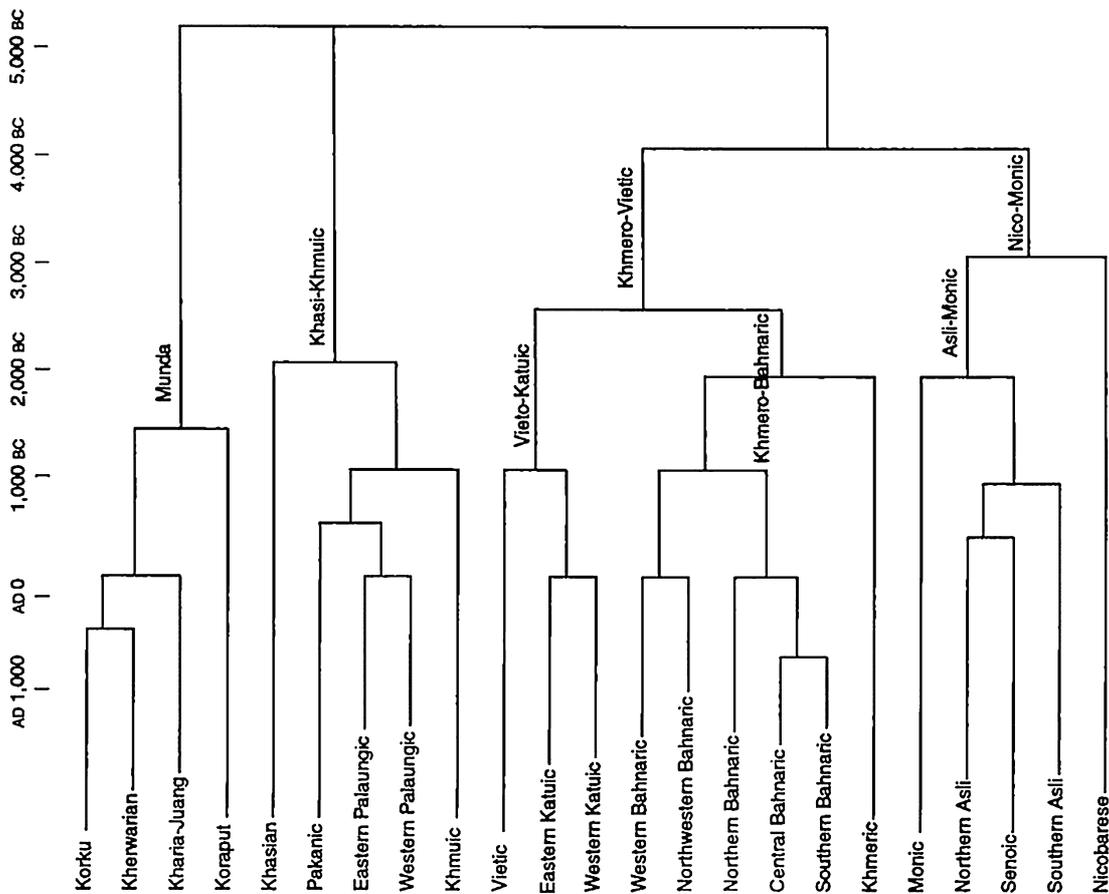
Figure 4: *Austroasiatic classification by Diffloth (2005)*

Munda as morphologically archaic?

Uniquely within AA, the Munda languages are highly agglutinative, making extensive use of suffixation. Characterizing the features of Munda as ancient, Pinnow asserted that:

> … the Munda languages are undoubtedly are more similar to Proto-Austroasiatic than the other members of the family. From a morphological viewpoint they are far more conservative than Nicobarese and Khasi, and from the standpoint of vocabulary they surpass the Mon-Khmer languages in their preservation of ancient word stems and word forms. (Pinnow 1963: 150)

Pinnow's assessment has precise and profound implications for the homeland problem. In asserting that Munda is more conservative, it strongly suggests that the Mon-Khmer languages innovated a typological restructuring away from suffixing etc. a change so profound that it probably did not happen more than once, and hence the Mon-Khmer languages are sprung from a common source, a single branching node in the AA tree. Such a model suggests that the ancestral Mon-Khmers migrating eastward, morphologically simplifying their languages the farther they went, until at the eastern extreme the Vietnamese (for example) eliminated any trace of historical affixes, recycling everything into compact monosyllables.

Some find this view persuasive. So confident is van Driem (2001:299) that he castigates Reid (1994), who attempted to resurrect the 'Austric' hypothesis by comparing Nicobarese and Austronesian, for not comparing with "the grammatically more conservative Munda languages". Van Driem went on to suggest that the intrusion of Tibeto-Burmans into Bengal stimulated the initial dispersal of AA to the west and east respectively, leaving the gap in the distribution of AA we see there today.

Yet it does not necessarily follow that the history of Mon-Khmer as been one of inexorable morphological loss — since the 1980s, Patricia Donegan and David Stampe have been variously arguing the case for a restructuring from isolating to synthetic typology, such that Pinnows formulation ought to be reversed, and Mon-Khmer characterised as structurally closer to proto-AA. The argument turns of the issue of phrasal accent. Munda accent is falling while in Mon-Khmer it is rising. This is crucial because the patterns of stress or accent have significant consequences over time, and tend to have strong structural correlates, particular in the distribution of clitics and affixes.

In falling accent languages morphemes that follow the head will tend to be phonetically reduced, and may grammaticalize into suffixes. On the other hand, in rising accent languages, morphemes that precede the head will tend to coalesce into prefixes. This suggests that one may reconstruct the direction of change based on typological observations. This is what Donegan (1993) had in mind when she pointed out that:

> Mon-Khmer has neither inflection nor suffixes, and neither did proto-Austroasiatic, but Munda have scores of them. […]
> Second, note the complex derivational and inflectional verb morphology, again largely suffixal. Pinnow, […], provided etymologies of many of these suffixes. Few of these reconstruct as suffixes to proto-Munda, i.e. they have been developed in the individual languages. (Donegan 1993:341)

And Stampe explained at the 2004 SEALS meeting in Bangkok:

> Only Pinnow (1960) and Zide & Anderson (2001) seem to have taken this view. Pinnow's argument was based on evidence for fossil suffixes in Khmer, none of which were supported by the exhaustive study of Khmer morphology by Jenner and Pou (1980–1981), and Pinnow's evidence was explained away by Jacob (1992),[4]

---

[4] The correct reference appears to be Jacob (1989–1990).

who showed that the variation such as Pinnow cited was expressive, not grammatical. Z & A's argument has two parts: (1) the occurrence of enclitic object pronouns and rarely nouns in Nicobarese and a few other Mon-Khmer languages, and (2) their reconstructions of the proto-Munda verb, which they with no specific evidence see as typical of Austroasiatic as well, with the implication that every trace was lost in Mon-Khmer. (Stampe 2004, 4)

Donegan & Stampe (2004:6) also compare Munda verb morphology to word-order equivalents in Mon-Khmer clauses, e.g:

Sora:	(*anin*)	*əd-*	*məl-*	*tiy*	*-dar*	*-iɲ*	*-da*	*-e*
	he	**not**	**want**	**give**	**rice**	**me**	-AUX	-3PR

Khmer:	*kŏət*	*ʔətcaŋ*	*ʔaoy*	*baay*	*kɲom*

Here we see a complete functional morpheme to morpheme correspondence, even though in this case only one morpheme is actually cognate (Sora **əd-** : Khmer **ʔət**). Unambiguously, Mon-Khmer VO word-order is preserved in Sora, grammaticalised as bound morphology, while at the clause level OV word-order prevails.

I would go even further, and propose that AA had had rising phrasal accent for a long time before the proto-language stage and the break-up of the family. Crucial is the observation that while only Munda has extensive suffixing, both Munda and Mon-Khmer have prefixes and infixes, and as Anderson (2004) and others have shown, this prefixing and infixing is cognate and reconstructable to the earliest times.

On typological grounds we might also assume to derive AA infixes from metathesized prefixes. In the present case, we can even suggest the specific source: infixed -**m-** and -**n-** derived from the implosives ɓ, ɗ which are otherwise rare or missing as prefixes, but securely reconstructable as frequent initials in the proto-language (see for example Shorto 2006). Proto-AA must have had rising accent long enough not only to develop prefixes, but infixes as well. The shift to falling accent in pre-Munda triggered the rise of suffixation, where AA prefixes and infixes remain recognizably intact. Clearly this implies the historical primacy of rising phrasal accent.

The question of accent also offers an explanation for Munda root structure. The problem is to reconcile the monosyllables and trochaic disyllables of the Munda root cannon, with the monosyllables and characteristically iambic sesquisyllables of Mon-Khmer languages.

In the first place we can assume that a proportion of roots have always been monosyllables and these require no special explanation. Mon-Khmer languages like Vietnamese (e.g. Ferlus 1998), Nyaheun (e.g. Sidwell&Jacq 2003), U (Svantesson 1988) reduced their sesquisyllables to monosyllables. The question is the approximately half of the lexicon which, in conservative Mon-Khmer languages, is sesquisyllabic. Shorto (2006) reconstructed a sesquisyllabic root cannon *(C)CV(:)C, and we might also apply this formula to proto-AA.

According to Donegan and Stampe's typology the rightmost edge of the phonological word will show reduced phonological and morphological structure. Significantly, the coda position universally in AA languages lacks any contrast of VOT, i.e. in Mon-Khmer we find unreleased voiceless stops with weak glottal coarticulation, and in Munda we find voiced stops with strong glottal coarticulation — in both cases it is a single series without voicing contrast, consistent with the theory.

Additionally, within Mon-Khmer there is frequently contrastive vowel length, while it is uncommon in Munda. In Mon-Khmer sesquisyllabic words mainsyllables are assigned 2 mora, and presyllables have no effective rhythmic weight and their vowels no phonemic value. Consistent with mainsyllables carrying 2 beats, we have the strong tendency for diphthongization

and register splits in Mon-Khmer. By contrast, Munda vowel inventories tend to be small, perhaps only a couple (if any) preserve old length distinctions.

We may suppose that in pre-Munda the phrasal accent fronted, and speakers began to assign a rhythmic beat to the presyllable (and reduplicating or otherwise augmenting monosyllables), taking a mora from the mainsyllable and giving it to the new initial syllable. Reduced to one mora, there would be little structural motivation for length or diphthongization contrasts. We can still see the traces of this process by looking at etymologies that suggest proto-long vowels. Consider the following examples (data and reconstructions from Shorto 2006):

PAA long vowels echoed:

*bluuʔ 'thigh'

Khmer:	**phlɤ̀u**
Bahnar:	**blu:**
Palaung:	**blu**
Temiar:	**bəloʔ**
Nicobarese:	**pulɔ:**
Sora:	**bulu:-n**
Kharia:	**bhulu**
Mundari:	**bulu**
Ho:	**bulu**
Kurku:	**bulu**

*ɟliiŋ 'long'

Old Mon:	*jlīń*
Lawa:	**ʔleiɲ**
Nicobarese:	**cəliŋ**
Mundari:	**ɟiliŋ**
Bhirhor:	**ɟiliŋ**

*kluuʔ 'tortoise'

Mon:	**klao**
Stieng:	**blɔ:u**
Sora:	**ku(:)lu:**
Kharia:	**kulu**

*briiʔ 'forest'

Bahnar:	**bri:**
Praok:	**praj**
Jah Hut:	**bəriʔ**
Old Khmer:	**vrai**
Kharia:	**biru**
Mundari:	**bir**
Bhirhor:	**bir**
Ho:	**bir**

*rk[aw]ʔ 'husked rice'

Khmer:	**ʔɔŋkɔ:**
Palaung:	**rəko**
Vietnamese:	**gạo**
Nyahkur:	**ŋkɔ́ː**
Kharia:	**'rɔŋkuʔb**
Sora:	**'ruŋku:**
Juang:	**runku:**
Gutob:	**runku:**

*ɟhaam 'blood'

Khmer:	**chì:əm**
Bahnar:	**pha:m**
Nicobarese:	**maha:m**
Kammu-Yuan:	**mà:m**
Sora:	**miɲaˑm**
Kharia:	**ijãm**
Santali:	**mãjãm**
Mundari:	**mãjɔ̃m**

*kjaal 'air, wind'

Old Mon:	*kyāl*
Bahnar:	**kja:l**
Khmer:	**khjɔl**
Kuy:	**kja:l**
Kharia:	**kɔjɔ**
Juang:	**kojo**
Santali:	**hɔe**
Mundari:	**hɔjɔ**

*kmuuʔ 'dirty'

Khmer:	**khmau**
Kuy:	**kmau**
Palaung:	**kəmu**
Bhirhor	**humu**
Mundari:	**humu**
Ho:	**homu**
Kurku:	**kumu**

*smuul 'shadow, soul'

Khmer:	**srəmaol**
Mon:	**həmao**
Kuy:	**smɔːl**
Bahnar (infixed):	**pəhŋɔːl**
Sora:	**um'mul**
Santali:	**umul**
Mundari:	**umbul**

PAA short vowels — variable/conditioned new initial syllable vowels:

*kraʔ 'road, way'

Kuy:	**knaː**
Praok:	**kra**
Muong:	**khá**
Mundari:	**hora**
Bhirhor:	**hora**
Kurku:	**kora**

*klaʔ 'tiger'

Bahnar:	**klaː**
Old Mon:	*kla(')*
Khasi:	**khlaː**
Khmer:	**khlaː**
Santali:	**kul**
Mundari:	**kul ~ kula**
Korwa:	**kul ~ kula**

*[hj]muʔ 'name'

Old Mon:	*jamo', himo'*
Praok:	**mɔ**
Semai:	**muh**
Old Khmer:	**ɟmah**
Sora:	**ə'ɲam-ən**
Kharia:	**'ɟĩmi, 'ɲimi**
Mundari:	**num ~ nutum**
Ho:	**numu**
Kurku:	**ɟumu ~ ɟimu**

*tmiʔ 'new'

Mon:	**kəmɔeʔ**
Khmer:	**thmɤj**
Thin	**hme**
Vietnamese:	**mới**
Sre:	**təme**
Kharia:	**'tɔnmɛ**
Sora	**'tammeː-**
Remo:	**tammɛ**

*kn[i]ʔ 'rat, mouse'

Old Mon:	*kni(')*
Bahnar:	**kənɛː**
Kammu Yuan	**kənéʔ**
Khasi:	**khnai**
Kharia:	**kɔnɛ**
Mundari:	**huni**
Kurku:	**ho'ne'**

Occasional reduplication of all or part of monosyllable:

*ɓaʔ 'paddy' (***(kə)ɓa:ʔ** Diffloth 2005)

Bahnar:	ɓa:
Temiar:	ba:ʔ
Danaw:	bā
Khasi (prefixed):	kba
Kharia:	baʔ
Mundari:	baba
Sora:	ba:ba:
Kurku:	baba

*[b]uuk 'head'

Stieng:	bu:k
Sre:	bou
Kharia:	bɔʔ / bɔkɔʔb
Mundari:	bɔhɔʔ ~ bɔɔʔ
Sora:	bɔʔɔb
Juang:	bɔkɔʔ

Simple monosyllablic roots with short vowels have straightforward reflexes:

*kap 'bite'

Bahnar:	kap
Kensiu:	kap
Katu:	kap
Car Nicobar:	kap
Sora:	kab
Santali:	haʔb
Mundari:	haʔb
Kurku:	kap

*mat 'eye'

Bahnar:	mat
Kensiu:	mɛd
Katu:	mat
Car Nicobar:	mət
Sora:	mʔo:d
Kharia:	mɔʔɖ
Mundari:	mɛʔd
Kurku:	met

*tɔh 'breast'

Bahnar:	tɔh
Semelai:	tuh
Katu:	tɔh
Car Nicobar:	tɛh
Santali:	toa
Bhirhor:	toa
Mundari:	toa

The following etymon has Eastern reflexes indicating both short and long variants, and Munda reflexes are similarly variable:

*ɟə[ə]ŋ 'foot/leg'

Old Mon:	*juṅ*
Palaung:	ɟɯŋ
Bahnar:	ɟəŋ
Khmer:	cə̀:ŋ

Kharia:	guɟuŋ
Sora:	ɟʔe:ŋ
Mundari:	ɟaŋga
Kurku:	ɟaŋga

I suggest that such examples indicate that Munda roots are readily derived from sesqui-syllabic proto-AA roots, more or less consistently with the root-canon reconstruction of Shorto (2006). Consequently, we cannot characterize the so-call Mon-Khmer languages as forming a sub-group in opposition to Munda. Rather they are relics that maintain (subject to various local changes) AA rising phrasal rhythm and its typological correlates. Thus there is no structural basis for thinking that Munda is "more similar to proto-Austroasiatic", and we need to turn to other methodologies to investigate this problem.

Linguistic Phylogeny: phonological reconstruction

For classification of a language family, especially of the structurally isolating type, the gold standard is a phonological reconstruction that identifies innovations with branching nodes. In the case of AA a comprehensive reconstruction has not been presented, rather we have only various branch level studies, and the attempt by Shorto (2006) which does not analyse Munda correspondences (Sidwell in press offers a comprehensive review of such studies). The reconstruction of the proto-AA vowels is very problematic, and results are accruing only gradually. However, the broad outlines of AA consonantism have been fairly well understood for a century already, being largely sorted out by Wilhelm Schmidt (1905, 1906 etc.). The basic consonant system preserved in conservative Mon-Khmer languages (such as Katu) and in the epigraphic record of Mon and Khmer before they underwent devoicing and restructuring, appears to adequately account for the proto-AA consonants without needing to posit additional segments or articulation types.

I reproduce here at Table 1 AA consonant correspondences for initial/prevocalic stops and fricatives — it is my working model at the time of writing. The table compares the oral stops reconstructed for each branch, plus the proto-MK (effectively proto-AA) reconstructions from Shorto (2006). Consonant mergers are shaded; splits are also indicated although without conditioning factors given (these are not yet thoroughly worked out). The table can be compared with the correspondences tabled by Shorto (2006: 52–54).

The most common change we can see is a merger of voiced and implosive stops, which occurred in six out of 12 branches.[5] This is a well distributed and rather trivial change which does not suggest any sub-grouping, not even among neighbouring groups. For example, it occurs in both Pearic and Khmer, although with the former there is a (partial only?) shift to aspiration among the voiceless stops.[6]

Perhaps the most important type of change, the devoicing on initials that is closely associated with restructuring of vowel systems, is not evident at all. It is clear that a voicing contrast is reconstructable for all branches, and such restructuring must have occurred independently after individual branches began diversifying internally.

Thus there is no patterning among the evident splits and mergers which suggest nested sub-branching among branches — their phonological histories appear to be quite independent, so far as I can tell. Consequently we have one other place to look for sub-grouping indications — the lexicon.

---

[5] Mangic is not included.

[6] There was discussion about this apparent Pearic split at the ICAAL4 meeting in Bangkok (October 28 2009); Michel Ferlus propsed that there was no conditioned split, rather there was borrowing of words with plain voiceless stops after the original voiceless series had become regularly aspirated.

PMK	Munda	Khasi	Palaungic	Khmuic	Vietic	Katuic	Bahnaric	Pearic	Pre-Khmer	Old Mon	Aslian	Nicobaric
*p-	*p-	*pʰ- / *p-	*p-	*p-	*p-	*p-	*p-	*pʰ- / *p-	*p-	*p-	*p-	*f-
*b-	*b-	*p-	*b-	*b-	*b-	*b-	*b-	*b-	*b-	*b-	*b-	*p-
*ɓ-	*b-	*b-	*ɓ-	*b-	*ɓ-	*ɓ-	*ɓ-	*ɓ-	*ɓ-	*ɓ-	*ɓ-	*p- (~*ʔ)
*t-	*t-	*tʰ- / *t-	*t-	*t-	*t-	*t-	*t-	*tʰ- / *t-	*t-	*t-	*t-	*t-
*d-	*d-	*t- / *d-	*d-	*d-	*d-	*d-	*d-	*d-	*d-	*d-	*d-	*t-
*ɗ-	*ɗ-	*d-	*d-	*d-	*d-	*d-	*d-	*d-	*d-	*d-	*d-	*d-
*h-	*G- (Ø- ~ h-)	*h-	*h-	*h- / -Ø-	*h-	*h-	*h-	*h-	*h-	*h-	*h-	*h-
*s-	*s-	*ɟ- / *s-	*s-	*s-	*s-	*s-	*s-	*s-	*s-	*s-	*s-	*h-
*c-	*s-	*s-	*c-	*c-	*c-	*c-	*c-	*cʰ- / *c-	*c-	*c-	*c-	*s-
*ɟ-	*ɟ-	*ɟ- / *ɟ-	*ɟ-	*ɟ-	*ɟ-	*ɟ- / *ɟ-	*ɟ-	*ɟ-	*ɟ-	*ɟ-	*ɟ-	*c-
*k-	*q- / *k-	*kʰ-	*k-	*k-	*k-	*k-	*k-	*kʰ- / *k-	*k-	*k-	*k-	*k-
*g-	*g-	*k-	*g-	*g-	*g-	*g-	*g-	*g-	*g-	*g-	*g-	*k-

Table 1: *Phonological correspondences for initial/pre-vocalic oral stops between Austroasiatic branches (PMK from Shorto (2006), proto-Munda follows Pinnow, other proto-branch reconstructions by Sidwell); (working model at 1/2/10)*

Lexicostatistics

Southeast Asian linguistics has frequently embraced the lexicostatistical method. Typically it is employed as a useful discovery tool, preliminary to a comparative reconstruction, and AA has been the subject of numerous lexicostatistical analyses. The most recent and extensive of these studies is Peiros (2004, in Russian), which utilizes data from over 100 languages. Peiros applies Starostin's glottochronological method which is controversial.[7] His AA classification is reproduced here at Fig. 5.;[8] he proposes an elabourate nested branching structure for the phylum; interestingly similar to Pinnow (1960) and Diffloth (1974) in that the highest branches involve Nicobarese, Munda and Mon-Khmer (although in a somewhat different relation).
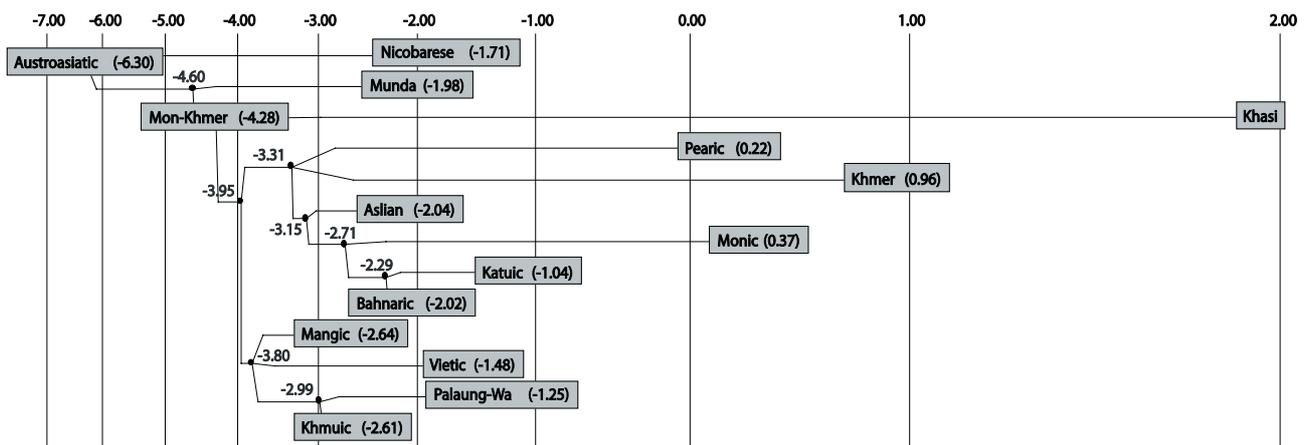
Figure 5: *Peiros (2004) Lexicostatistical classification*

I decided to replicate the study, but with a smaller, more managable set of data, in this case 36 languages. A major concern being to use only languages where both the sound correspondences and the major contact languages are reasonably well understood/documented. My data set, with cognate assignments and commentary is publicly available via my website at http://people.anu.edu.au/~u9907217. The matrix at Fig. 7 was generated using Glotpc.exe (devised by Jacque Guy, and freely downloaded via link at http://sil.org). The results came out quite different to those of Peiros — it is my analysis that he underestimated both borrowings between branches (especially from Mon and Khmer into neighbouring languages) and placed too much importance on some very low cognate percentages (such as Nicobarese) where isolation and social factors must affect the rate of lexical change.[9] Thus I contend that a somewhat flatter tree is indicated, in fact, lexical signals that might support sub-branching are so confused that it is difficult to argue for something other than a simple rake-like tree.

What especially struck me were remarkably parallels with my matrix and the one offered more than 30 years ago by Frank Huffman (1978), reproduced here at Fig. 6.[10] Although my calculations yeiled lower overall cognate densities between branches, there is a structural

---

[7] The software, and Peiros' data and reconstructions, are accessible via: http://starling.rinet.ru/.

[8] Peiros (2004) doe snot provide a matrix of percentages, only various dendrograms and an appendix of data and numbered cognate assignments.

[9] Peiros mentions the possibility of faster rate of change in Nicobarese, but did not take it into account. Readers can compare Peiros' data with my own cognate assignments and analyses at: http://people.anu.edu.au/~u9907217/lexico/AAclassification.html.

[10] Huffman's data is available online via: http://sealang.net/archives/huffman/

similarity. Both Huffman and I found the highest interbranch percentage is between Katuic and Bahnaric, plus there is a strong tendency for other branches to show a higher percentage between themselves and Katuic-Bahanric than with other branches, with the effect declining as one moves geographically further way from Katuic-Bahnaric. Logically it does not seem possible to represent these relations as nested branching relationships.

Chart 1: Interbranch Cognate Percentages and Averages

	K-B	Kme	Mon	Pal	V-M	Pea	Asl	Kmu	Kha	Mun	Nic
Katuic-Bahnaric		47	42	34	33	35	31	32	24	27	20
Khmer	47		33	31	30	39	28	25	24	24	23
Mon	42	33		33	30	22	24	28	22	20	19
Palaungic	34	31	33		26	25	22	26	26	18	23
Viet-Muong	33	30	30	26		24	22	24	23	20	18
Pearic	35	39	22	25	24		24	22	17	19	16
Aslian	30	28	24	22	22	24		24	20	17	19
Khmuic	32	25	28	26	24	22	24		24	20	17
Khasi	24	24	22	26	23	17	20	24		18	14
Munda	27	24	20	18	20	19	17	20	18		17
Nicobarese	20	23	19	23	18	16	19	17	14	17	
Interbranch totals	324	304	273	264	250	243	231	222	212	200	186
Interbranch averages	32.4	30.4	27.3	26.4	25.0	24.3	23.1	22.2	21.2	20.0	18.6

Figure 6: *Austroasiatic lexicostatistical matrix by Huffman (1978)*

In Huffman's study, inter-branch averages are shown, which makes the effect very clear. In my matrix individual languages are distinguished, and another potentially interesting pattern is noted: the higher percentage scores with Katuic-Bahnaric actually fall away within, rather than just between, branches. For example: within Palaungic Danaw shows markedly lower scores; within Aslian Jahai scores lower; within Khmuic Mal scores very high, Khmu' moderately, and Mlabri very low. Even within Khasian Wa scores significantly higher against Katuic-Bahnaric than do Standard Khasi or Pnar. Such variations appear to be marginal or absent in respect of Munda and Nicobarese.

How can one explain these patterns? It is possible that to some extent the data of individual languages includes some cases of more rapid lexical replacement that are complicating the picture. This said, we still have a stark geographic correlation: the closer any language or branch is to Katuic or Bahnaric, the higher percentage of cognates will be found with Katuic and/or Bahnaric. At the same time, I have not identified shared innovations that would compel one to sub-group Katuic-Bahnaric. Thus I suggest that we are seeing the effects of prolonged contact, centred about the middle Mekong. Branches that moved further away from this contact area were less affected (such as Palaungic or Aslian), and some (such as Mangic and Nicobarese) came into quite different language contact areas or other social conditions, such that their lexicons were likely subject to even more change than might have otherwise occurred.

Huffman suggested that:

> This would seem to argue for an eastern (Central Vietnam) center of dispersal and a separate westward migration for each branch of Austroasiatic. (Huffman 1978:5)

The idea would be that various branches, when they were still basically unitary languages, wandered out of the central zone at different times, hence there are no indications of

	Munda			Khasi			Palaung			Mangic			Khmuic			Vietic			Katuic			Bahnar			Pearic			Khmer		Monic		Aslian			Nico.	
	Ko	Mu	So	Kh	Pn	Wa	Da	Wa	U	Bu	Ma	Pa	Ml	Km	Ma	Mu	Vi	Ru	Ku	Pa	Ka	Ba	Jr	St	Ch	Ka	Pe	Su	Kh	Mn	Ny	Se	Sl	Ja	Ca	Na
Korku		45	29	9	9	12	9	12	11	10	14	13	14	17	18	10	13	13	22	19	18	15	19	22	12	12	15	12	13	21	19	14	14	8	9	11
Mundari	45		36	10	10	14	13	16	15	12	15	15	12	15	20	14	17	16	21	17	15	15	19	20	15	16	18	15	16	26	24	13	15	8	8	11
Sora	29	36		11	9	11	8	11	14	15	11	15	13	15	15	16	17	17	18	15	14	12	14	20	10	11	13	9	10	17	17	13	10	8	4	8
Khasi	9	10	11		84	52	16	18	19	14	15	16	12	14	16	13	14	15	14	11	12	10	13	10	12	12	13	13	12	13	13	9	8	8	5	10
Pnar	9	10	9	84		55	18	19	20	14	16	18	11	15	15	13	14	15	15	10	13	9	11	10	12	12	13	13	14	13	13	11	10	8	5	10
War-J	12	14	11	52	55		15	20	22	12	15	16	12	16	20	13	13	13	19	16	16	12	15	16	13	15	15	15	14	16	16	13	16	9	7	11
Danaw	9	13	8	16	18	15		25	28	13	14	14	10	17	15	8	10	10	20	15	18	13	13	17	12	11	13	10	10	17	19	12	12	10	4	10
Wa	12	16	11	18	19	20	25		53	18	21	21	14	24	20	17	15	19	19	17	20	17	19	19	16	17	18	17	17	23	25	19	16	14	11	16
U	11	15	14	19	20	22	28	53		18	20	20	18	30	23	11	14	19	23	20	21	15	19	20	17	17	16	17	17	24	26	17	17	15	7	13
Bugan	10	12	15	14	14	12	13	18	18		24	46	13	19	14	14	15	18	16	15	15	14	19	17	12	16	15	10	11	17	20	14	14	13	8	10
Mang	14	15	11	15	16	15	14	21	20	24		31	12	20	15	18	17	20	18	17	19	13	17	19	15	15	15	13	13	21	23	17	14	12	10	15
Paliu	13	15	15	16	18	16	14	21	20	46	31		13	17	13	17	15	19	18	18	17	15	18	18	15	18	18	14	15	19	22	14	15	10	7	11
Mlabri	14	12	13	13	11	12	10	14	18	13	12	13		29	31	9	12	12	18	18	17	15	18	21	10	11	14	13	13	13	16	13	12	11	6	10
Khmu	17	15	15	14	15	16	17	24	30	19	20	17	29		31	14	16	18	24	15	19	17	22	22	20	20	22	17	17	21	23	21	16	13	9	13
Mal	18	20	15	15	15	20	15	20	23	14	15	13	31	31		16	16	19	30	21	21	22	23	14	21	22	24	19	20	20	24	22	22	15	11	13
MuongKoi	10	14	16	13	13	13	8	17	11	14	18	17	9	14	16		65	51	20	18	19	17	16	16	12	13	13	8	8	17	17	14	13	11	7	10
Viet.	13	17	17	14	14	13	10	15	14	15	17	15	12	16	16	65		53	21	15	19	17	15	21	13	13	13	9	9	16	16	17	13	13	7	12
Ruc	13	16	17	15	15	13	10	19	19	18	20	19	12	18	19	51	53		22	21	21	18	19	22	13	16	16	11	11	20	20	16	15	13	11	15
Kui	22	21	18	14	15	19	16	20	19	23	16	18	18	24	30	20	21	22		58	53	31	31	32	25	26	27	18	18	27	28	26	24	16	14	16
Pacoh	19	17	15	11	10	16	14	15	17	20	15	17	18	22	27	18	15	21	58		62	29	26	29	24	24	27	18	18	23	26	22	24	15	14	15
Katu	18	15	14	12	13	16	16	18	20	21	15	19	17	22	26	19	19	21	53	62		33	31	32	26	26	28	20	20	26	27	24	26	14	13	15
Bahnar	15	15	12	10	9	12	13	17	15	14	13	15	15	17	21	17	17	18	31	29	33		50	47	24	22	24	20	23	22	25	29	26	12	13	16
Jru	19	19	14	13	11	15	13	19	19	19	17	18	18	22	22	16	15	19	31	26	31	50		41	22	24	23	20	22	29	30	23	24	14	12	14
Stieng	22	20	20	10	10	16	13	17	19	20	17	19	18	21	21	14	16	21	32	29	32	47	41		20	20	23	23	23	31	30	25	20	12	13	15
Chong-H	12	15	10	12	12	13	12	16	17	12	15	15	10	20	21	12	13	13	25	22	26	24	22	20		76	59	17	19	21	21	17	18	10	13	15
Kasong	12	16	11	12	12	15	11	17	17	16	15	18	11	20	22	13	13	16	26	24	26	22	24	20	76		67	19	21	25	24	19	20	11	13	13
Pear-B	15	18	13	13	13	15	13	18	16	15	15	18	14	22	24	13	13	16	27	27	28	24	23	23	59	67		23	25	24	23	22	21	13	14	15
Surin	12	15	9	13	13	15	10	17	17	10	13	14	13	17	19	8	9	11	18	18	20	20	20	23	17	19	23		87	22	22	19	22	13	10	14
Khmer	13	16	10	12	14	14	10	17	17	11	13	15	13	21	20	8	9	11	18	18	20	23	22	23	19	21	25	87		23	23	21	22	12	10	13
Mon	21	26	17	13	13	16	17	23	24	17	21	19	13	21	24	17	16	20	27	23	26	22	29	31	21	25	24	22	23		80	22	23	15	15	15
Nyakur	19	24	17	13	13	16	19	25	26	20	23	22	16	23	24	17	16	20	28	26	27	25	30	30	21	24	23	22	22	80		22	23	15	14	14
Semai	14	13	13	9	11	13	12	19	17	14	17	14	13	21	22	14	14	16	26	22	24	29	23	25	17	19	22	19	19	22	22		40	30	15	17
Semelai	14	15	10	8	10	16	12	16	17	14	14	15	12	16	22	13	13	15	24	24	26	26	24	20	18	20	23	19	21	22	23	40		24	12	15
Jahai	8	8	8	8	8	9	10	14	15	13	12	10	11	13	15	11	7	13	16	15	14	12	14	12	10	11	13	13	12	15	15	30	24		9	12
Car	9	8	4	5	5	7	4	11	7	8	10	7	6	9	11	7	7	11	14	14	13	13	12	13	13	13	14	10	13	15	14	15	12	9		46
Nancawri	11	11	8	10	10	11	10	16	13	10	15	11	10	13	13	10	12	15	16	15	15	16	14	15	15	13	15	14	13	15	14	17	15	12	46	

Figure 7: *Austroasiatic lexicostatistical matrix compiled by Sidwell 1/2/10. Branches are boxed, anomolous high percentages are shaded.*

sub-grouping between these geographically peripheral groups.[11] Within the central zone there was ongoing differentiation tempered by contact. The 'least moves' needed to account for the distribution of the phylum is a simple radiation, rather than than a nested tree of mulitiple nodes. This would seem to be consistent with the Migration Theory approach.

In the absence of a detailed analysis of AA lexical innovations, this is where the available evidence has taken us. The lexicostatistics provide broad indications of the lexical diversity of the phylum and its branches. At least in respect of the basic vocabulary, we cannot say, for example, that Munda is more diverse than, say, Khmuic or Aslian or Mangic. While some branches, such as Khmer and Monic, are quite small, with the data at hand we cannot say that any of the larger branches looks especially older than the others. Thus we have only a weakly branching tree or rake-like radiation.

Conclusion:

Three independent lines of inquiry — morphological, phonological, and lexical — have failed to provide decisive indications of nested sub-groupings among AA branches, while the lexical data strongly suggests that there is a contact area centred on Katuic and Bahnaric. Until other indications are forthcoming, the most reasonable hypothesis is a simple radiation out of the Mekong valley.

Harry Shorto, who was inclined to accept a Yangtze origin for AA, struggled to reconcile this with what he knew of Mon and Khmer, and reached a similar conclusion. Writing 30 years ago he remarked:

> The Northern Mon-Khmers and Khasis are likely to have followed what became a Chinese trade route to India, as the Mundas may well have done before them. But there seems no over-riding reason to trace routes for the Mons and Khmers, and other groups who occupied the river-plains, down the rivers from the hinterland rather than up them from the coast. (Shorto 1979:278)

Literature

ANDERSON, G.D.S. 2006. Advances in Proto-Munda reconstruction. *Mon-Khmer Studies Journal*, 34:159–184.

BLENCH, Roger. 2008. Roger Blench Stratification in the peopling of China: how far does the linguistic evidence match genetics and archaeology? In: Alicia Sanchez-Mazas, Blench, R.M., Ross, M.D., I. Peiros & Marie Lin eds. *Human migrations in continental East Asia and Taiwan. Matching archaeology, linguistics and genetics.* 105–132. London: Routledge.

CHAZEE, Laurent.1999. *The Peoples of Laos: Rural and Ethnic Diversities*: Bangkok White Lotus.

DIEBOLD, A.R. 1960. Determining the centers of dispersal of language groups. *International Journal of American Linguistics* 26: 1–10.

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. 1999. Austroasiatic Classification appearing in Chazée, Laurent (1999) *The Peoples of Laos: Rural and Ethnic Diversities*: Bangkok White Lotus.

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. Pp. 79–82.

---

[11] This may turn out to be too strong a claim. There are weak lexical indications of a connection between Khasi and Palaungic, I expect to explore these in the near future.

DONEGAN, Patricia and David STAMPE. 1983. Rhythm and the Holistic Organization of Language Structure. In RICHARDSON, MARKS and CHUKERMAN (eds.), *Chicago Linguistic Society Papers from the Parasession on the Interplay of Phonology Morphology and Syntax*. Pp. 335–353.

DONEGAN, Patricia and David STAMPE. 2004. Rhythm and the Synthetic Drift of Munda, The Yearbook of South Asian Languages and Linguistics. Berlin and New York, De Gruyter. Pp. 3–36.

DONEGAN, Patricia. 1993. "Rhythm and vocalic drift in Munda and Mon-Khmer." *Linguistics of the Tibeto-Burman Area* 16.1:1–43

DYEN, Isidore. 1956. Language distribution and migration theory. *Language* 32/4: 611–26. (Reprinted in *Linguistic Subgrouping and Lexicostatistics*. The Hague: Mouton. Pp. 50–74.)

FERLUS, Michel. 1998. Les systèmes de tons dans les langues viet-muong. *Diachronica* 15:1.1–27.

GRIERSON, George Abraham. 1904. *Mon-Khmer and Siamese-Chinese Families*. Vol. II, *Linguistic Survey of India*: Delhi.

HUFFMAN, Franklin E. 1978. *On the centrality of Katuic-Bahnaric to Austroasiatic*. Unpublished handout distributed at the 3rd ICAAL meeting in Mysore.

JACOB, Judith. 1989–1990. Some comments on the relationship between Khmer words having identical vowel nuclei and final consonants. *The Mon-Khmer Studies Journal*, 18–19:67–76.

JENNER, Philip & Saveros POU. 1980–81. A lexicon of Khmer morphology. *Mon-Khmer Studies* 9–10.

KUIPER, F. B. J. 1955. Rigvedic loanwords, in: Studia Indologica, ed. Spies, Bonn

KUIPER, F. B. J. 1967. The genesis of a linguistic area. Indo-Iranian Journal 10: 81–102

KUIPER, F. B. J. 1991. Aryans in the Rigveda, Rodopi (1991).

MEI TSU-LIN and NORMAN, Jerry. 1976. The Austroasiatics in Ancient South China: some lexical evidence. *Monumenta Serica* 22:274–301

PEIROS, I. and V. SHNIRELMAN. 1998. 'Rice in Southeast Asia: a regional interdisciplinary approach', in R. BLENCH and M. Spriggs (eds) *Archaeology and Language, archaeological data and Linguistic hypotheses* Vol. II: 379–389, London: Routledge

PEIROS, Ilia J. 2004. *Genetičeskaja klassifikacija avstroaziatskix jazykov*. Moskva: Rossijskij gosudarstvennyj gumanitarnyj universitet (dissertacija).

PEIROS, Ilia. 1989. Dopolnenie k gipoteze S. A. Starostina o rodstve nostraticheskix i sinokavkazskix jazykov. *Lingvisticheskaja rekonstrukcija i drevnejshaja istorija vostoka*: Materialy k diskussijam na mezhdunarodnoj konferencii (Moskva, 29 Maja — 2 Ijunja 1989g.) Chast' 1. Moskva, Nauka.

PEIROS, Ilia. 1998. *Comparative Linguistics in Southeast Asia*. Pacific Linguistics Series C-142. Canberra.

PINNOW, Heinz-Jürgen. 1960. Uber der Ursprung der voneinander abweichenden strukturen der Munda- und Khmer-Nikobar-Sprachen. *Indo-Iranian Journal* 4:81–103.

PINNOW, Heinz-Jürgen. 1963. The position of the Munda languages within the Austroasiatic language family. In *Linguistic Comparison in Southeast Asia and the Pacific*, edited by H. L. Shorto: London: SOAS.

PINNOW, Heinz-Jürgen. 1963. The position of the Munda languages within the Austroasiatic language family. In: H.L. Shorto (ed.). *Linguistic Comparison in Southeast Asia and the Pacific.* London: SOAS. 140–52

REID, Lawrence A. 1994. Morphological evidence for Austric. *Oceanic Linguistics* 33(2):323–344.

SAPIR, Edward (1916). *Time Perspective in Aboriginal American Culture, A Study in Method*. Ottawa: Government Printing Bureau.

SCHMIDT, Pater Wilhelm. 1905. Grundzüge einer Lautlehrer Khasi-Sprache in ihren Beziehungen zu derjenigen der Mon-Khmer-Sprachen. Mit einem Anhang: die Palaung-Wa-, und Riang-Sprachen des mittleren Salwin. *Abh. Bayrischen 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* 5:59–109.

SCHUESSLER, Axel. 2007. *ABC Etymological Dictionary of Old Chinese.* University of Hawaii Press.

SHORTO, Harry L. 1979. The linguistic proto-history of mainland South East Asia. in: R. B. Smith & W. Watson (eds.), *Early South East Asia.* New York, Kuala Lumpur, Oxford University Press. Pp. 273–278.

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. 2009. *Classifying the Austroasiatic languages: history and state of the art*. Munich, Lincom Europa.

STAMPE, David. 2004. Was Proto-Austroasiatic like Munda or like Mon-Khmer. Conference handout, 14th Southeast Asian Linguistics Society meeting, Bangkok, May 2004.

SVANTESSON, Jan-Olof. 1988. U. *Linguistics of the Tibeto-Burman Area*, 11.1:64–133.

VAN DRIEM, George. 2001. *Languages of the Himalayas Volume One*: Leiden Brill.

WITZEL, Michael (1999), "Substrate Languages in Old Indo-Aryan (Rgvedic, Middle and Late Vedic)", Electronic Journal of Vedic Studies 5 (1), http://www.ejvs.laurasianacademy.com/ejvs0501/ejvs0501article.pdf

ZIDE, Norman H. & G. D. S. ANDERSON. 2001. "Recent Advances in the Reconstruction of the Proto-Munda (Autroasiatic) Verb" L. Brinton (ed.). *Historical Linguistics 1999*. Amsterdam: John Benjamins. 13–30.

Статья посвящена вопросам прародины и последующего расселения австроазиатских языков. Традиционный «интуитивистский» подход пока не позволяет убедительно классифицировать более мелкие языковые подгруппы внутри уже установленных ветвей этой лингвистической семьи. При этом лексический анализ словарного состава анализируемых языков указывает на наличие длительных контактов и, возможно, случаев языковой конвергенции на территории континентальной Юго-Восточной Азии. По мнению автора, это указывает если не на исконность, то, по крайней мере, на существенную древность присутствия австроазиатских языков в Индокитае (с возможным центром расселения, расположенным в бассейне р. Меконг). Наиболее географически удаленная ветвь австроазиатской семьи — языки мунда в Индии — демонстрирует исключительно высокое число инноваций. В соответствии с этой теорией в статье реконструируется эволюция корневой структуры языков мунда.