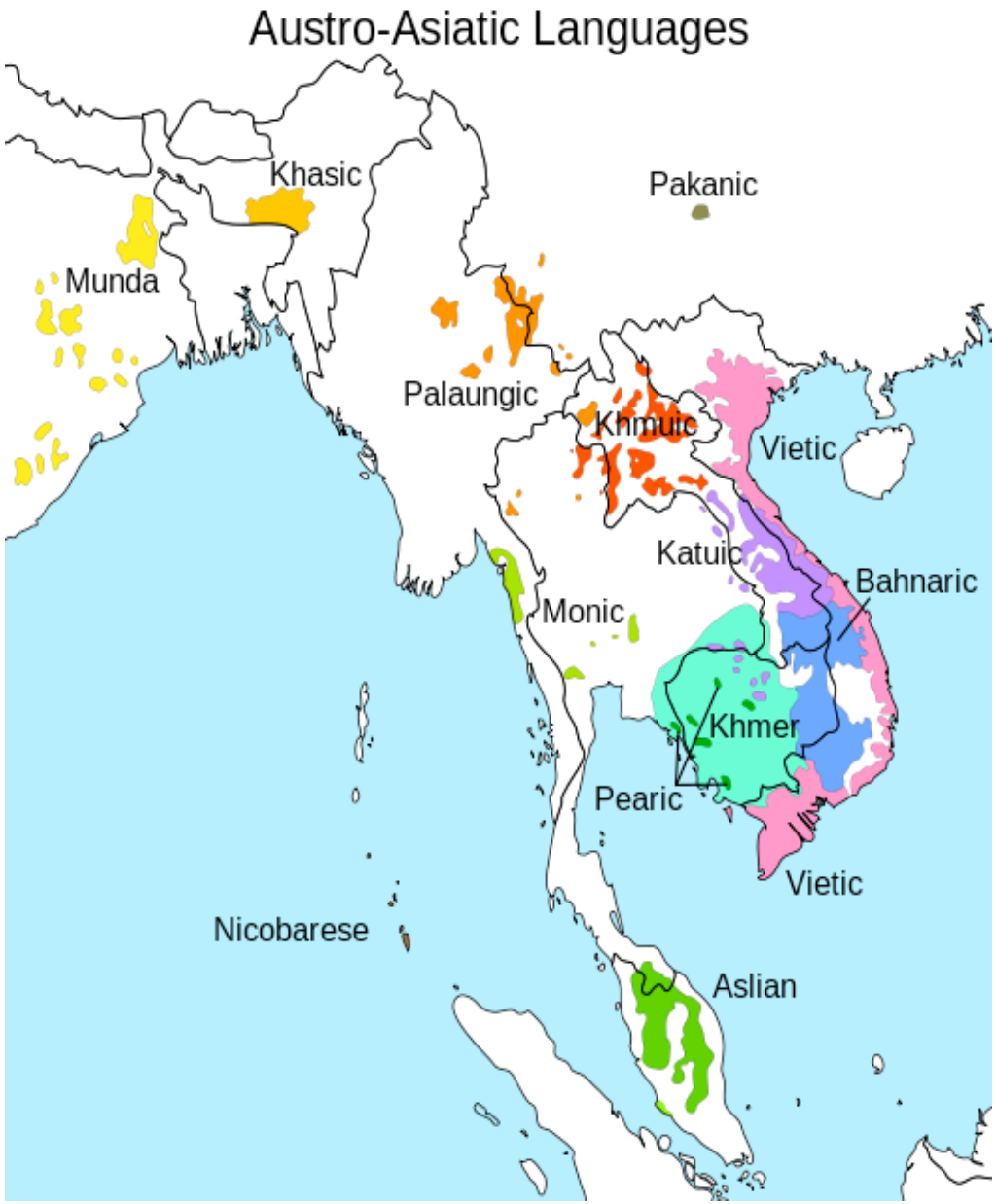
University of Zurich[UZH]

Towards a family portrait of Austroasiatic: phonological characteristics

Tobias Weber

toweb@spw.uzh.ch

SEALS 22, Agay, France, May 30 – June 2, 2012


Austro-Asiatic Languages
Khasic
Pakanic
Munda
Palaungic
Khmuic
Vietic
Katuic
Bahnaric
Monic
Khmer
Pearic
Vietic
Nicobarese
Aslian

Methodology

- research questions:
  - What is typical for Austroasiatic (AA)?
  - What is uncommon in AA?
  - To what degree do the AA languages match the areal patterns of surrounding languages?
- data from 50+ AA languages and some dominant non-AA contact languages (taken from grammars and other descriptions)
- for some properties: Mon-Khmer Etymological Dictionary database (Sidwell and Cooper 2007-2012)
- for a broader comparison: UPSID (UCLA Phonological Segment Inventory Database, containing 451 languages, first presented in Maddieson 1984) and the WALS (World Atlas of Language Structures, latest version by Dryer et al. [eds.] 2011)

Size of the consonant inventory

small (20 consonants or less): languages of almost all AA branches

Munda: Gorum 18

Palaungic (e.g. P'uman 17, Riang 18)

Khmuic (e.g. Mal 17)

Katuic (e.g. Pacoh and Suei 18)

Bahnaric (e.g. Rengao 19, Jeh 19, Hrê 18)

Khmer 17, Pear 18

Aslian: Kensiu 18, Jahai 19, Temiar 19

Nicobarese: C Nicobarese 16, Car 17

average (21-29 consonants):

majority of the AA languages

majority of the languages worldwide, including: Thai 22, Malay 21+

large (30 consonants or more):

Munda: Santali 30, Kharia 31

Palaungic: Lawa 38

Khmuic: S Khmu' 30

Bahnaric: Sre 30, Loven 38

also: Bengali 31, Burmese 33

Size of the consonant inventory

Common consonant inventory of Austroasiatic languages

p	t	c	k	ʔ
b	d	ɟ	g	
	s			h
m	n	ñ	ŋ	
	r			
	l			
		y	w	

Stop series

1 series

C Nicobarese: only plain voiceless

2 series: usually voiceless vs. (sometimes implosive) voiced

Munda: Sora, Gutob
Palaungic: Riang, Shwe
Khmuic: S Khmu‘ (+/-aspirated), Mal
Katuic: Suei
Bahnaric: Rengao, Jeh, Hrê, Halang
Khmer, Pear
Aslian: majority of the languages
also: Malay

3 series: usually aspirated voiceless vs. nonaspirated voiceless vs. voiced

Khasi
Palaungic: Danau
Khmuic: S Khmu‘ (voiced implosive)
Vietic: Thavung, Vietnamese (vd. impl.)
Katuic: Kuy, Katu, Ngeq
Mon (vd. impl.)
Aslian: Mah Meri, Semelai
also: Burmese, Thai

3 series: voiceless vs. voiced vs. implosive

Bahnaric: Stieng, Chrau, Bahnar, Loven
Katuic: Katu

Stop series

4 series: aspirated vs. nonaspirated voiceless vs. voiced vs. breathy voiced

Munda: Korku, Santali, Mundari, Kharia

also: many Indo-Aryan lgs

4 series: aspirated vs. nonaspirated voiceless vs. voiced vs. implosive

Palaungic: Lawa (voiced is prenasalized)

Bahnaric: Sre

- the majority of the AA languages have two or three stop series, geographic picture unclear, but:
- one series is rare
- Aslian predominantly two, Vietic predominantly three series
- four series in Munda languages due to language contact with Indo-Aryan

Fricatives

- s and h only: vast majority of the AA languages
- v, s, h: Santali (Munda), other South Asian languages, Riang, Wa (Palaungic), Thavung (Vietic), Suei (Katuic), Khmer
- f/Φ, s, h: Lawa (Palaungic), Kuy (Katuic), Mintil (Aslian), C Nicobarese (also Thai)
- s, ʃ, h: Amwi (Khasic), Mon (also Bengali)
- s, z, h: Remo (Munda)
- more complex contrasts:
  - Danau (Palaungic): θ, s, ʃ, h (cf. Burmese)
  - Loven (Bahnaric): f, v, ‘v, s, ‘s, h, ‘h
  - Vietnamese: f, v, s, z , x, ɣ, h
- all the dominant languages in the AA speaking area have more fricatives than just s and h

➔ areal mismatch

Word-initial velar nasal /ŋ/

- present in 50% of the WALS languages (Anderson 2011)
- common in Eurasia only in the NE and the SE
- virtually all AA languages have a phonemic velar nasal
- virtually all Mon-Khmer languages also have a word-initial velar nasal
- in Munda languages the velar nasal does not occur word-initially (as in other languages of South Asia)

➔ matches the areal picture: word-initial velar nasal present in the languages of MSEA, largely absent in South Asia

Other uncommon consonants

- retroflex consonants: Munda (as in other South Asian languages)
- preglottalized nasals: S Khmu‘, Loven, Halang, Sedang (analyzed as clusters)
- preglottalized liquids: Loven (l and r), Bahnar (l)
- preglottalized glides: Lawa (y and vl. y), Bahnar (y), Hrê (y, vl. y, w)
- voiceless sonorants: Lawa (N, l, r, y), S Khmu‘ (N, l, r, w), Sre (ñ, y, w), Loven (N, l, y), Sedang (N, y, w (analyzed as clusters)), Mon (N, l), Che‘ Wong (N), (also Burmese (N, l))
- other glides: palatal rounded glide (Wa), back unrounded glide (Mal)
- preploded nasals (non-phonemic): several Aslian; also some Austronesian languages of Borneo and Sumatra

Size of the vowel inventory

Average vowel system of AA languages

(not taking into account diphthongs and distinctions in length, nasalization, and phonation types):

i	ɨ/ɯ	u
e	ə/ɤ	o
ɛ		ɔ
	a	

- majority of the AA more complex than what is commonly found wordwide (5-6 vowels)
- less complex: most Munda, Khasic, and also Malay (6)

Diphthongs

- complex distinctions in AA languages; also involving only non-high vowels
  - Mundari (Munda): 7: ae, oe, oa, ua, ia, ea, eo (cf. South Asia: not that prominent)
  - Shwe (Palaungic): 9: eo, ɛə, aɛ, oə, iə, eə, uə, (aə), (ɔɛ)
  - Khmu‘: 4: ia, ɨa, ua, əa
  - Bruu (Katuic): 14: ia, ea, ɯa, oa, oʌ, ua, ie^{h}, ia^{h}, ɯo^{h}, oa^{h}, uo^{h}, ua^{h}, ũə, ũa
  - Sedang (Bahnaric): 9: ia, ua, ea, oa, io, uo, eo, ie, oe
  - Khmer: 13: ei, əɨ, uə, ou, æɛ, aə, ao, ĕə, ŭə, ŏə, iə, ɨə, ɔə
  - less complex in Aslian and Nicobarese (cf. neighboring Austronesian: not that complex, except Acehnese and some Chamic)
- also fairly complex in Thai/Lao, not that complex in other MSEA

➔ complex diphthong distinctions are fairly prominent in MSEA; the most complex distinctions are found in AA

Nasalized vowels

- 22.62% of the UPSID languages
- not very common in MSEA, common in South Asia
- Munda: majority (but not Kharia)
- Khasic: e.g. Amwi (but not Khasi)
- Katuic: Bruu, Suei
- Bahnaric: only few, such as Jeh and Sedang
- Aslian: common
- Nicobarese: common, present in C. Nicobarese and Car

➔ matches the areal patterns fairly well; Aslian also needs to be compared to Austronesian languages of Borneo and Sumatra

Long vowels

- length distinction in 11.3% of the UPSID languages
- presence common in AA languages and also generally in MSEA
- Munda: at least in Ho and Gorum (but not in many other Munda languages) (cf. Hindi: present, Bengali: absent)
- Khasic: Khasi
- Palaungic: at least in Shwe and Wa yes (cf. distinction absent in Burmese)
- Khmuic: generally present (cf. Thai/Lao: present)
- Vietic: Thavung (but not Vietnamese)
- Katuic: presence common
- Bahnaric: majority (but not Sedang and Hrê)
- Khmer and Pear: present
- Aslian: at least in Temiar, Mah Meri, and Semelai (cf. Malay: absent)
- Nicobarese: C Nicobarese absent, Car present

Mid vowel series

- values: one vs. two (i.e. /e/ vs. /ɛ/ and/or /o/ vs. /ɔ/)
- worldwide: one series more common (incl. Hindi, Malay)
- MSEA: two series common (incl. Thai, Burmese, and also Bengali)
- majority of the AA languages have two series
- the following languages only have one series:
  - Munda: majority (but Santali and Sora have two)
  - Khasic: Khasi (but Amwi has 2)
  - Bahnaric: Chrau, C Mnong, Jeh, Halang

➔ matches the areal pattern: one series common in South Asia, two series common in MSEA east of Burmese

Central/back unrounded vowels

Presence of phonemic schwa (/ə/ or /ɤ/) and/or high central/back vowel (/ɨ/ or /ɯ/)

- worldwide: 20.81% of the UPSID languages have schwa, 15.52% have the high central/back vowel
- MSEA: presence of both common
- Munda: majority none, Remo only schwa, Santali and Sora both (cf. Hindi: schwa only, Bengali: none)
- Khasic: Khasi none, Amwi schwa
- Palaungic: Riang, Shwe and Wa schwa only, P'uman and Lawa both (cf. Burmese: schwa only)
- Bahnaric: Jeh, Halang, Modra, Didra none; C Mnong and Sedang schwa only, rest both
- remaining branches generally both
- Thai both, Malay schwa

➔ majority of the AA languages have both, again except Munda, some diversity in the northern branches and in Bahnaric

➔ areally only minor mismatches between AA and neighboring languages

Vowel phonation types

- two-register distinction very common in AA languages
- do occur elsewhere in the world (e.g. some Otomanguean, some "Khoisan", Gujarati) but not frequently
- Munda: generally no registers
- Khmer: recently lost the register distinctions
- Aslian: uncommon, but present in Mah Meri
- more complex systems (cf. Thongkum 1988):
  - Didra: 3 registers: creaky, modal, breathy
  - Chong: 4 registers: modal, modal-creaky, breathy, breathy-creaky

➔ register distinctions without tone distinctions are clearly a family feature, generally not adopted in languages of other families in MSEA

Tones

- fairly common worldwide, very common in MSEA but rare elsewhere in Eurasia
- only in a minority of AA languages, but not that rare; examples:
- Munda: Korku 2 (cf. neighoring South Asian lgs: no tonal distinctions)
- Khasic: Khasi none, Amwi 4
- Palaungic: Danau 3, Riang 2, Shwe 3, P'uman 2, Wa 4 (cf. Burmese: 4)
- Khmuic: N Khmu' 2, Mal 2 (cf. Thai: 5)
- Vietic: Thavung 4, Vietnamese 6 (cf. other languages of N Vietnam and S China: complex tone systems)
- Aslian: Jahai 2?

➔ most AA are non-tonal, although the neighboring Tibeto-Burman, Tai-Kadai and Hmong-Mien languages are tonal

Maximal syllable structure

- Munda: generally CVC → less complex than Indo-Aryan (CCVCC)
- Khasic: CCVC
- Palaungic: often CCCVC → more complex than Burmese (CGV?)
- Khmuic: N Khmu‘ CCVC, Mal CCCVCC (cf. Thai/Lao CCVC)
- Vietic: Thavung CVC, Vietnamese CVC(C)
- Katuic: Bruu CCVCN, Kuy CCCVC, Pacoh CCVC
- Bahnaric: Chrau, Sre CCCVC, CCCVCC, E Mnong CCVC, Loven CCVC, Jeh CCVC, Halang CCVC, Sedang CCVC
- Khmer CCCVC
- Aslian: e.g. Jah Hut CCVC (cf. Malay (C)CVC)
- Nicobarese: C Nicobarese CVC

→ areal mismatches esp. in the cases of Munda and Palaungic

Prosodic word structure

- most languages of MSEA have a iambic stress pattern (cf. Brunelle & Pittayaporn forthcoming)
- most frequent word types in MSEA: disyllabic > sesquisyllabic > monosyllabic (“>”=diachronic development)
- sesquisyllabicity very prominent in Mon-Khmer
- monosyllabicity fairly rare (found in only 9 out of 235 AA languages in the Mon-Khmer Etymological Dictionary database (Sidwell & Cooper 2007-12) according to Brunelle & Pittayaporn fc), although MK languages are in close contact with dominant monosyllabic languages
- Munda: mostly trochaic, as other South Asian languages
- trochaic stress pattern more common in Austronesian; notable exceptions: Chamic, Acehnese

➔ areal mismatches in the case of MSEA, language contact might account for the Munda and Chamic cases

Results

- Fairly homogeneous areal patterns can be found for some properties (e.g. word-initial velar nasal, broadly vowel inventory size)
- But in many cases there are mismatches between AA and neighboring non-AA languages
- …and also a lot of diversity even within low-level genealogical groups of AA
- Language contact cannot account for this diversity

➔ Although WALS-like projects are crucial for today's (areal) typology, their (often fairly broadly defined) values often cannot account for finer-grained distinctions in smaller areas

➔ Language-specific (and, in some cases, even family-specific) internal diachronic developments might offer better explanations for those cases