Phonological Word Domains in Austroasiatic Languages

René Schiering, Franziska Crell & Thomas Goldammer

1. Austroasiatic Languages

(1) Language Sample

Munda	Northern Munda	Kherwarian		**Santali**
	Southern Munda			**Kharia**
Mon-Khmer	Northern Mon-Khmer	Khasic		**Khasi**
		Khmuic		**Khmu**
	Southern Mon-Khmer	Aslian	Jehaic	**Jahai**
			Semelaic	**Semelai**
		Nicobaric		**Car**
		Monic		**Mon**
	Eastern Mon-Khmer	Viet-Muong	Vietic	**Vietnamese**
		Katuic		**Pacoh**
		Khmeric		**Cambodian**
		Bahnaric		**Chrau**

(2) Typological features of Austroasiatic languages

i. Weak noun/verb-distinction in Munda (cf. Evans & Osada 2005 on Mundari)

ii. Tonogenesis in Mon-Khmer: -voice > high, +voice > low (cf. Haudricourt 1954)

iii. 'Unlikely cousins' (Donegan 1993, Donegan & Stampe 1983)

(3) Rhythmic Holism in Austroasiatic Languages

	MUNDA	MON-KHMER
Phrase Accent:	Falling (initial)	Rising (final)
Word Order:	Variable - SOV, AN, Postpositional	Rigid - SVO, NA, Prepositional
Syntax:	Case, Verb Agreement	Analytic
Word Canon:	Trochaic, Dactylic	Iambic, Monosyllabic
Morphology:	Agglutinative, Suffix-ing, Polysynthetic	Fusional, Prefixing or Isolating
Timing:	Isosyllabic, Isomoric	Isoaccentual
Syllable Canon:	(C)V(C)	(C)v̆ or (C)(C)v́(G)(C)
Consonantism:	Stable, Geminate Clusters	Shifting, Tonogenetic, Non-Geminate Clusters
Tone/Register:	Level Tone (Korku only)	Contour Tones/Register
Vocalism:	Stable, Monophthongal, Harmonic	Shifting, Diphthongal, Reductive

2. Phonological Word Domains in Austroasiatic Languages

2.1. Phonological Words in Munda

2.1.1. Minimal/Maximal Word Size

- The Proto-Munda bimoraic word requirement survives in Kharia and Santali. Accordingly, the vowel of monosyllabic words is generally long in both languages.

(4) Kharia
si /si/ → [si:] 'plow (v)'

(5) Santali
ti /ti/ → [ti:] 'hand'

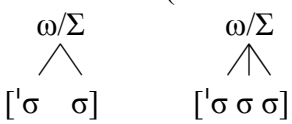
- With respect to maximal word size, both languages are said to prefer disyllabic domains, exceptionally allowing for trisyllabic domains (cf. the foot structure as outlined below).

2.1.2. Stress

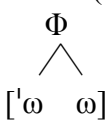
- The stress systems of both Kharia and Santali are still not accessibly described. However, the schematic description in (6) and (7) seem to capture the most general characterization.

(6) Kharia

Word stress (= trochaic feet)

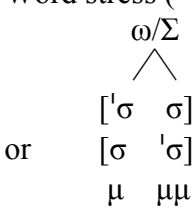


Phrasal stress (falling)

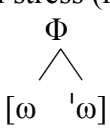


(7) Santali

Word stress (= trochaic feet)



Phrasal stress (rising)



- Foot structure overrides morphological structure in both cases, i.e. polysyllabic words which result from morphological processes such as affixation, reduplication and compounding are dissolved into stress units with preferably two syllables each.

2.1.3. Phonotactics

- A number of phonotactic restrictions on the distribution of certain phonemes and consonant clusters define the left and right edge of words.

(8) Kharia

word-initial	word-final
*C onset if /ŋ/	?*C coda if /h/
?*C onset if /ɾ, ʈ/	?*C coda if /s/
?*C onset if /w/	?Superheavy syllables
?V-initial syllables	
[C Prenasalization]	[Schwa insertion with C final words]

(9) Santali

word-initial	word-final
*C onset if /ŋ/	*C coda if /h/
*C onset if /ʈ/	*C coda if /w/
*C onset if /w/	*C coda if /y/
*C onset if /y/	*CC coda (unless C_1 is nasal)
	[Final Consonant Lengthening]

- Note that in these cases, information on morphological domains is usually not discussed. In most of the cases, these restrictions could be interpreted as phonotactic restrictions on stems; in some cases they seem to include bound morphemes.

2.1.4. Morphophonology

- Since both languages disallow vowel-initial syllables within words, several morpho-phonological processes are employed to repair hiatus resulting from affixation or cliticization.

(10) Kharia

a. Glide insertion (prefix+stem+enclitic)

u=aʔ (DEM=GEN) → *uwaʔ* 'of this'

b. Front vowel deletion (stem+enclitic)
koŋ=te=iɲ (know=ACTIVE.PRS=1s) → *koŋtiɲ* 'I know'

(11) Santali
a. Glide insertion (stem+suffix+enclitic)
ho-ok^{ʔ}-a-e (become-MID-IND-3sS) → *hoyok^{ʔ}ae* 'he will become'
[b. Diphthongization, no example but see word-final /ae/ above]

- Additionally, consonants which appear in word-medial position are subject to segmental rules such as lenition and flapping.

(12) Kharia
Intervocalic Flapping of /ɖ/ (stem+enclitic)
ɖoɽ=e (take=ACTIVE.IRR) 'take'

(13) Santali
Lenition (stem+suffix+enclitic)
dɛc^{ʔ} 'pick' → *dɛj-a-e* 'he will pick'

2.1.5. Summary

- With respect to phonological word domains, Munda languages are characterized by the minimal bimoraic word constraint and a preference for maximally disyllabic words.

- Whereas both languages are trochaic at the word level, they differ with respect to falling vs. rising phrase prosody.

- A number of phonotactic restrictions allow the identification of word boundaries, as well as morphophonological rules of hiatus resolution and consonantal alternations.

2.2. Phonological Words in Mon-Khmer

2.2.1. Minimal/Maximal Word Size

- The Mon-Khmer languages in our sample differ with respect to the minimal and maximal word size. Although minimal words are usually bimoraic and maximal words usually disyllabic, there are a number of deviations, not only in loan word phonology.

(14)

	Minimal word	Maximal word
Khasi	/k^{h}la/ → [k^{h}la(˙)] ‘tiger’	*kr̥teŋ* ‘name’ (> 2 σ = loans)
Khmu	CV: or CVC	*rm̀-hɔ́:m* (NMLZ-tie) ‘tying’
Jahai	CVC	σ σ σ
Semelai	/t^{h}i/ ‘hand’	[kə.ru.wan.ceŋ] ‘coral snake sp.’
Car	σ	σ σ σ σ
Mon	*ʔa(:)* ‘go’	*pə-lac* ‘tear down’ (> 2 σ = loans)
Vietnamese	*đi* ‘go’	σ σ σ (> 1 σ = loans)
Pacoh	*dɔ:* ‘3s’	*ʔa-maj* ‘2s.DAT’
Cambodian	CV: or CVC	σ σ σ σ σ σ (> 1 σ = loans)
Chrau	*hwi* [hwi:] ‘wide’	*panang* ‘room’

2.2.2. Stress/Tone

- The ten Mon-Khmer languages discussed here all have final stress at the word and at the phrase level (as far as we can tell from the available information). Note that the lack of word stress is peculiarity of Vietnamese.

(15)

	Word stress	Phrasal stress
Khasi	*ŋa-léŋ* (1s-go) ‘I go’	*kà-ʃa:-sá:w* (fs-tea-black) ‘black tea’
Khmu	*rm̀-ˈhɔ́:m* (NMLZ-tie) ‘tying’	[n.a.]
Jahai	[n.a.]	[n.a.]
Semelai	kə.ru.wan.ˈceŋ ‘coral snake’	[n.a.]
Car	[n.a.]	(phrase-final)
Mon	*pə-ˈlac* ‘tear down’	*kwan ˈmòa* ‘a village’
Vietnamese	[NO]	*Tôi không ʹbiết*. ‘I don’t know’
Pacoh	*ʔa-ˈmaj* ‘2s.DAT’	[n.a.]
Cambodian	*qa.khòo.sáq* ‘future’	[n.a.]
Chrau	*paˈnang* ‘room’	[n.a.]

- Note that in most cases, the weak unstressed syllable results from prefixation or infixation, the only surviving forms of affixation. Reduplications and compounds seem to behave like phrases with respect to phrase-final stress assignment.

- Vietnamese is the only real tone language in the sample. Here, tonal Sandhi affects reduplications and compounds in such a way that disyllabic domains agree in high vs. low tone register. The other languages are quasi-tonal in showing different phonation types on vowels depending on the initial consonant within the syllable (cf. light vs. heavy register in Mon).

2.2.3. Phonotactics

- The most salient feature of syllable structure in Mon-Khmer is the dichotomy of minor syllables and major syllables, sometimes also referred to as pre- and main syllables, which together constitute the maximal word shell.

(16)

	Minor syllable	Major syllable
Khasi	CC̣/Cə	(C)CCVVC
Khmu	CC̣ (Cə)	CCVVC
Jahai	[n.a.]	[n.a.]
Semelai	CəC/CuC	CVC
Car	[NO]	CVC
Mon	Cə	C(C)V(C)
Vietnamese	[NO]	C(w)V(V)(C)
Pacoh	CV(C)	(C)CV(V)(C)
Cambodian	Cə	C(C)(C)V(V)(C)
Chrau	CV	(C)(C)CV(C)

- In all languages discussed here, the (unstressed) minor syllable is characterized by a number of phonotactic restrictions on onset and coda consonants. The nucleus of the presyllable is schwa, an unspecified vowel or a syllabic consonant.

- Consonant clusters are only possible in word-medial position within disyllabic words or word-initially in monosyllabic words. The combination of presyllable and major syllable potentially increases the complexity of consonant clusters.

- The (stressed) major syllable is usually characterized by fewer restrictions on onset and coda consonants. Unlike the presyllable, the nucleus of the major syllable can be filled by any full vowel of the language.

2.2.4. Morphophonology

- Morphophonology rules show a rather erratic distribution across the languages of the sample. There is not a single type of rule which is traceable in all representatives of the branch.

- Illicit consonant clusters are resolved by rules of deletion and epenthesis in Khasi, Khmu and Semelai.

(17) Khasi

a. V insertion in consonant clusters	b. C deletion in consonant clusters
/lber/ → [ləbɛːr] 'March'	/ka-/ 'fs' + /ʔm̩-/ 'NEG' → [kam-]

(18) Semelai

V insertion in consonant clusters

a. /d.rɛ/ → [də.rɛ] 'rattan'	b. /rp-rəp/ → [rəp-rʌp] 'to be threshing'

- In Khmu, nasal consonants are subject to morphophonological rules of nasal assimilation.

(19) Khmu

/rn-hɔ́ːm/ (NMLZ-tie) → [rǹ-hɔ́ːm] ~ [rm̀-hɔ́ːm] 'tying'

- Disyllabic words constitute the domain for a number of morphophonological rules which operate across two adjacent syllables. Such rules include nasalization, tone/register assimilation and vowel harmony.

(20) Jahai

Nasalization

a. /b-n-awc/ → [mə̃nawət˺] 'pig tailed macaque'

b. /haʔɛ̃t/ → [hãʔɛ̃t˺] 'stench'

(21) Semelai

Nasalization

/mham/ → [mãhãm] 'blood'

(22) Khmu

Tone Sandhi

lìan 'to come out' → *p-lían* 'to take out'

(23) Mon

Vowel Register Assimilation

a. /ʔuʔcàn/ → [ʔùʔcàn]

b. /ʔuʔpətè/ → [ʔùʔpətè]

(24) Vietnamese
Register Harmony in Reduplication
a. /trang/ (*sac1*) ‘white’ → /trang trang/ (*ngang-sac1*) ‘rather white’ (HIGH)
b. /lanh/ (*nang1*) ‘cold’ → /lanh lanh/ (*huyen-nang1*) ‘rather cold’ (LOW)

(25) Cambodian
Vowel harmony
/muc/ ‘to submerge’ → [prɔ-mɔc] ‘to put under’

2.2.5. Summary

- Mon-Khmer languages differ with respect to minimal word domains: whereas bimoraic domains are minimal in the majority of languages, a couple of languages show only weak evidence for this constraint (Khasi, Mon) or have lost it completely (Semelai, Vietnamese).

- Maximal words are usually disyllabic, however, a number of languages violate this principle, especially due to the influence of contact languages (loan word phonology).

- Iambic word stress and rising phrasal prosody is a common characteristic. Vietnamese lacks word stress and only exhibits regular phrase-final stress. Tone is at best incipient in the different vowel registers found in the different languages, Vietnamese sticks out again as being fully tonal.

- The phonotactics of minor and major syllables is evident across all languages, though with differences in detail with respect to the concrete constraints on consonants and vowels. Vietnamese and Car lack this characteristic since they lack the minor syllable.

- There is no common morphophonology across the languages discussed. Consonant clusters are subject to repair strategies in some languages. In some languages, the disyllabic word forms the domain for syllabic processes such as tone harmony or vowel harmony.

3. Summary and outlook

- Common traits for Austroasiatic word domains are hard to pin down. Bimoraic minimal words and disyllabic maximal words seem to be more reliable candidates. Note that those languages which lost the bimoraicity constraint are at the Eastern margin of the area (Semelai and Vietnamese). In most cases language contact is the reason for the loss the disyllabicity constraint.

- With respect to stress, Munda (trochaic) and Mon-Khmer (iambic) in fact behave like unlikely cousins. Note, however, that iambic feet are also evidenced in Santali, as well as rising phrasal prosody.

- This observation is also true with respect to phonotactics. The asymmetry of weak and strong syllables is characteristic for Mon-Khmer (again with certain traces already evident in Santali). In the Eastern margins of Mon-Khmer (Vietnamese) this asymmetry is lost due to the lack of the weak syllable.

- Overall, different types of morphophonological rules show a very erratic distribution across Austroasiatic languages and are therefore very bad indicators for common traits.

➢ Although the synchronic data compiled here suggests common traits in the phonological word domains of the Austroasiatic languages in general, the significant differences between Munda and Mon-Khmer (being gradual on a west-east axis) call for an explanation which might be found in diachronic or language contact scenarios.

References

Donegan, Patricia 1993. “Rhythm and vocalic drift in Munda and Mon-Khmer.” *Linguistics of the Tibeto-Burman Area* 16: 1-43.

Donegan, Patricia J. & David Stampe 1983. “Rhythm and the holistic organization of language structure.” In: Richardson, John F.; Mitchell Marks & Amy Chukerman (eds.). *Papers from the Parasession on the Interplay of Phonology, Morphology and Syntax*. Chicago: Chicago Linguistic Society: 337-353.

Evans, Nicholas & Toshiki Osada 2005. “Mundari: The myth of a language without word classes.” *Linguistic Typology* 9: 351-390.

Haudricourt, André-G. 1954. “De l’origine des tons en vietnamien.” *Journal Asiatique* 242: 69-82.