Divergence and convergence in Sumatra and the Barrier Islands

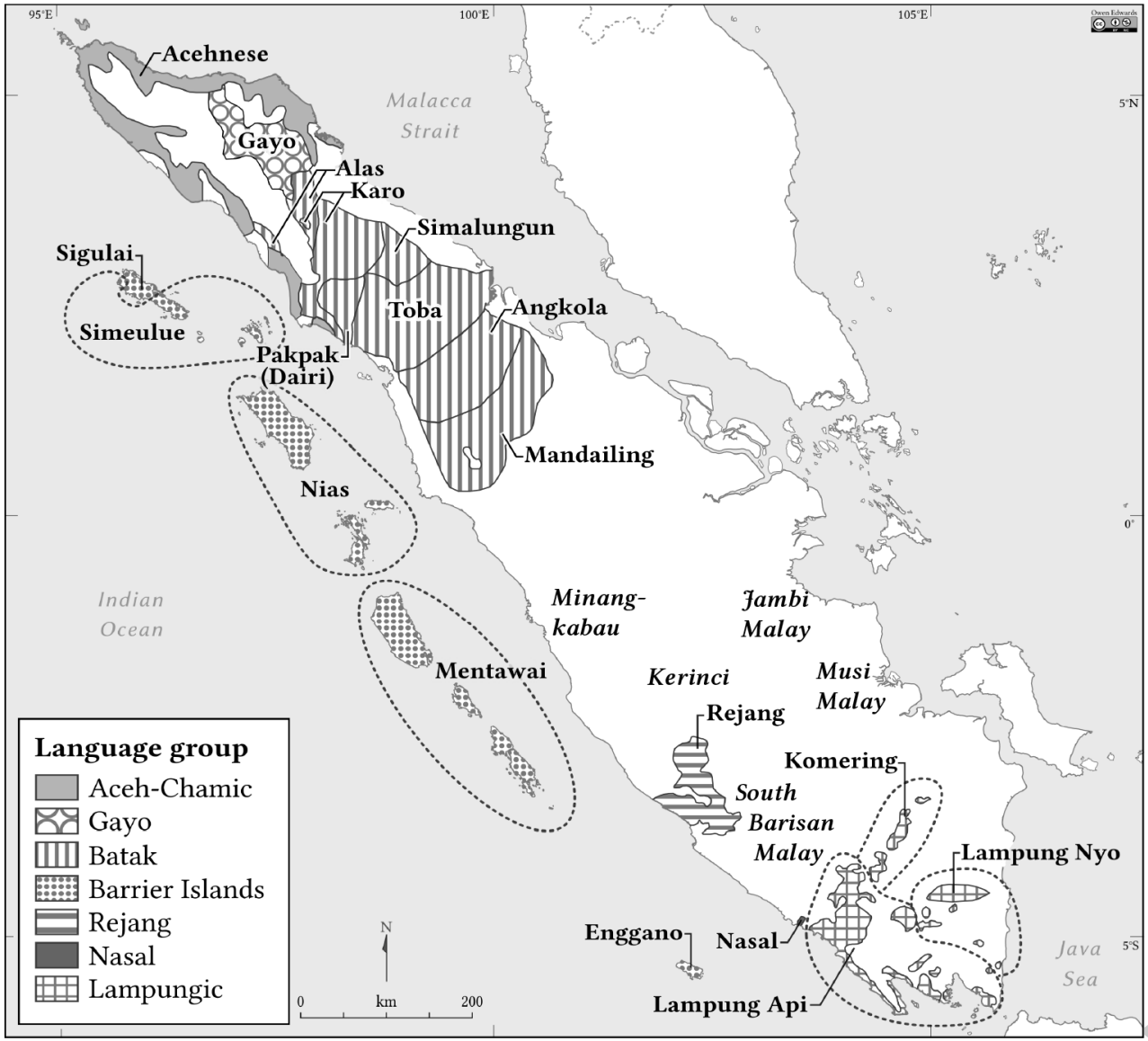
Bradley McDonnell and Blaine Billings
University of Hawai‘i at Mānoa

January 9, 2025

Goals

1. Provide an overview of the languages of Sumatra and the Barrier Islands (especially *Sumatran* languages)
2. Present examples of divergence and convergence in these languages
3. Discuss implications for understanding the history of the region

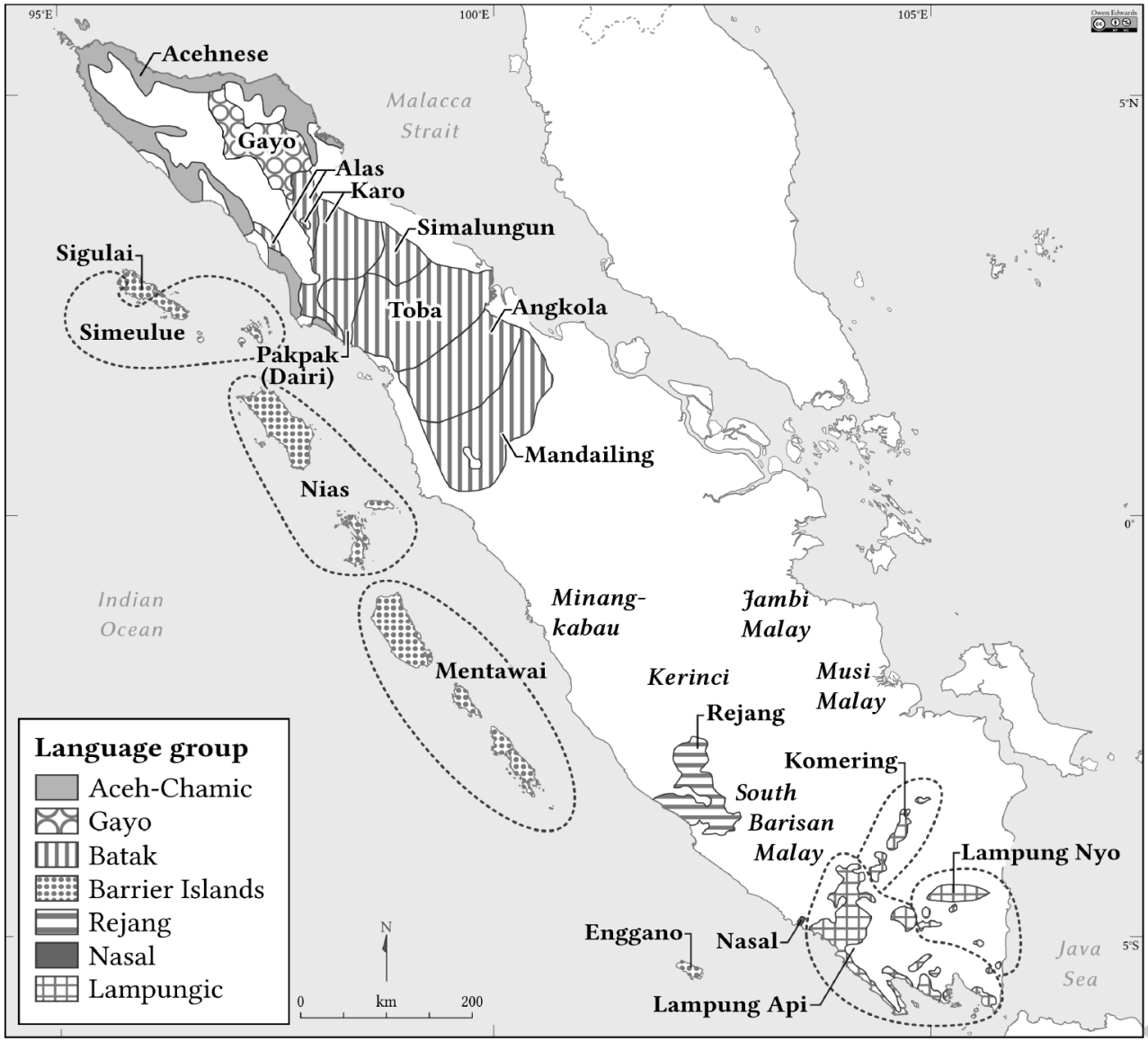
Sumatra and the Barrier Islands



Diverse Malayo-Polynesian languages

1. Sumatran
   - Batak
   - Barrier Islands
   - ...
2. Malayo-Chamic
   - Acehnese
   - Malay varieties
3. Lampungic
4. Rejang

Sumatra and the Barrier Islands



Diverse Malayo-Polynesian languages

- Little-to-no typological/areal features uniting these languages
- Any similar features are a result of contact w/ Malay (McDonnell & Truong 2024)

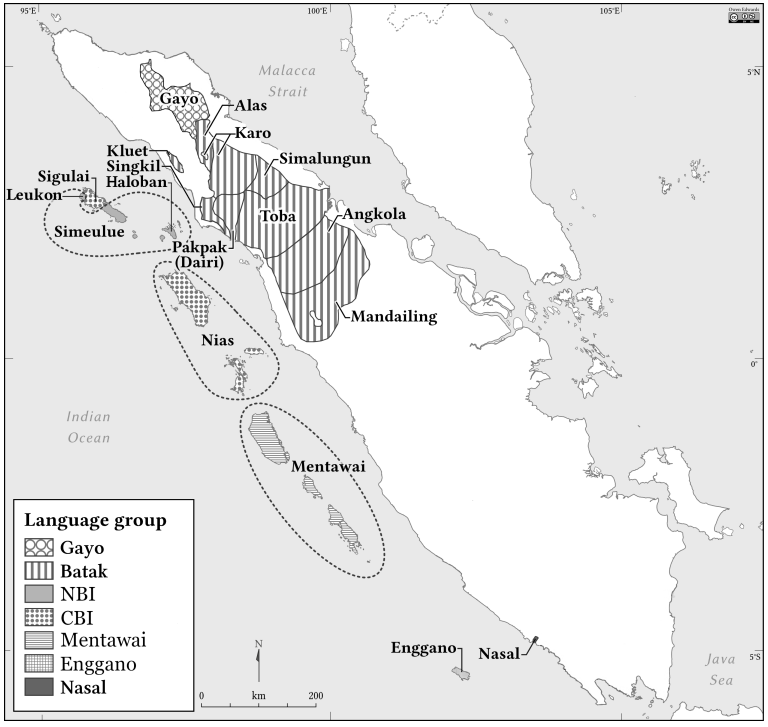
Malayo-Chamic

By far the largest population of speakers.

- Sumatra contains many, many varieties of Malay with complex networks of dialects.
- Acehnese is the only Chamic language in Sumatra
  - Undergone extensive changes due to contact with Austroasiatic languages
- Malayo-Chamic likely originated in Borneo.

Sumatran subgroup

Sumatran: Non-Malayo-Chamic languages of Sumatra *except* Lampungic and Rejang

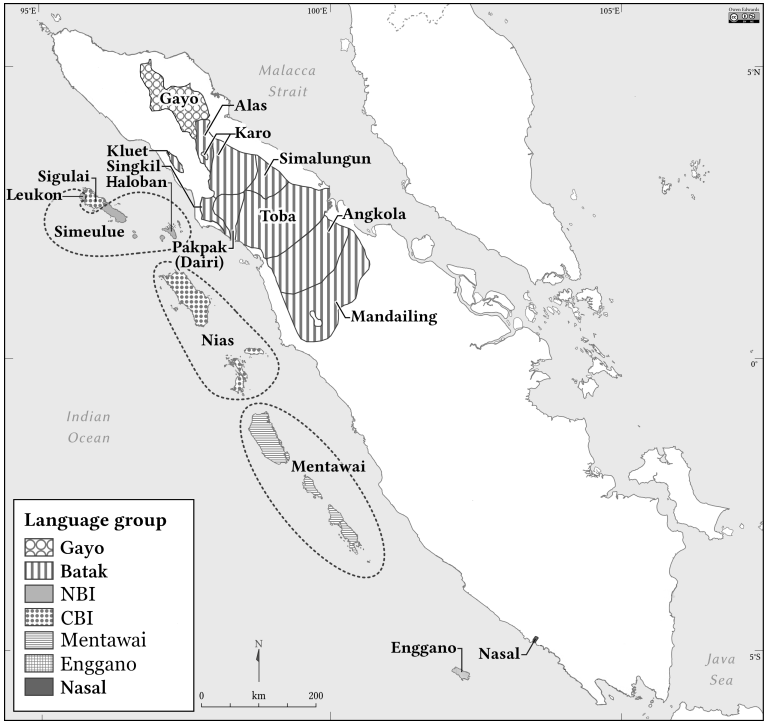


- *Batak-Barrier Islands* subgroup (Nothofer 1986)
- *Sumatran* (Smith 2017)

Billings, Blaine & Bradley McDonnell. 2024. Sumatran. *Oceanic Linguistics* 63(1). 112–174

Sumatran subgroup

Sumatran: Non-Malayo-Chamic languages of Sumatra *except* Lampungic and Rejang



1. Batak
2. Northern Barrier Islands
   - Haloban, Simeulue, Leukon
3. Central Barrier Islands
   - Sigulai, Nias
4. Mentawai
5. → Gayo
6. → Enggano
7. → *Nasal*

Billings, Blaine & Bradley McDonnell. 2024. Sumatran. *Oceanic Linguistics* 63(1). 112–174

Phonological Innovations PMP > Proto-Sumatran

Proposed shared innovations in Proto Sumatran:

Mergers

1. PMP *j, *g > Proto Sumatran *g
2. PMP *z, *d > Proto Sumatran *d
3. PMP *R, *r > Proto Sumatran *r
4. PMP *ñ, *n > Proto Sumatran *n
5. PMP *h > Proto Sumatran ø

Shift

1. PMP *q > Proto Sumatran *h

Billings, Blaine & Bradley McDonnell. 2024. Sumatran. *Oceanic Linguistics* 63(1). 112–174

Lampungic

- Lampungic languages spoken by 1,500,000 people in southwest Sumatra
  - in northern Lampung province,
  - parts of south South Sumatra.
- Three primary branches – Lampung Nyo, Lampung Api, Komering
  - Historically in contact with Nasal.

Rejang

- Rejang spoken by 350,000 people in southwest Sumatra
  - in northern Bengkulu province,
  - parts of northwest South Sumatra.
- McGinn (2005) identifies five dialects:
  - Rawas dialect may not be mutually intelligible with other Rejang dialects.

(Linguistic) history of Sumatra

> *"the linguistic history of Sumatra has not been one of continuous undisturbed differentiation, but probably included a major episode of language leveling in which incoming Malayic speakers replaced earlier languages and so 'reset the clock' of linguistic evolution"*
>
> (Blust 2013: 77-78)

Four apparent primary migrations:

1. Pre-Austronesian groups (see Blench 2014 for Enggano specifically)
2. Arrival and dispersal of speakers of Proto Sumatrans (← Nasal)
3. Displacement by Malay groups
4. Recent (< 100 years) population movements

Notions of "vergence"

Divergence

Languages/dialects from a common source increasingly become dissimilar

- typically with reduced contact over time

Convergence

Languages/dialects from a different sources become *similar*

- typically with sustained contact over time

Both notions are scalar: Languages or bundles of features in a language or languages can be said to be highly divergent or even highly convergent.

Let's compare two Sumatran languages at either end of the spectrum.

Enggano, Austronesian's "most divergent language"?

Enggano perhaps serves as one pole in this scale Domains of divergence:

- **Lexicon**: Apparently small number of inherited lexical items
- **Phonology**: Large number of innovative phonological developments
- **Morphology**: Large number of innovative morphemes
- **Syntax**: Case marking, alignment shift, ...

Seemingly all domains?

Nasal as a convergent language

Nasal perhaps serves as the opposing pole Domains of convergence:

- **Lexicon**: Largely overlapping with neighboring Malayic and Lampungic
- **Phonology**: Inventory leveled to that of surrounding languages
- **Morphology**: Few clear inherited, corresponding to neighbors
- **Syntax**: Indonesian-type voice, *te-* with explicit agent, ...

The picture is a near polar opposite

Divergence

Divergence

Divergent features within and across languages/dialects are often found in language ecologies characterized by geographical isolation.[1]

[1] Acehnese is the one exception.

Mentawai lexicon

> *"the only other example in western ISEA of a language that might display a level of aberrancy vaguely approaching the level seen in Enggano is its neighbor Mentawai."*

(Edwards 2015: 91)

Mentawai varieties on Siberut remarkably diverse:

- Siberut Island: Area of 110km x 45km
- Wide range of lexical difference: 11%–71%
- Even "basic vocabulary" varies greatly

Simalegi	**Sikabaluan**	**Tailelew**	
ama	*mae*	*uk:uy*	'father'
puro	*bata?*	*buk:u?*	'stone'
rid:it	*sigigiŋ*	*uma?*	'bird'
sa?ra?	*leyti?*	*bala?*	'earthworm'
maeje?	*marai?*	*maregre*	'near'
mara?	*masepu*	*malep:et*	'cold'

(see Langgang 1992/1993)

Mentawai phonology

Throughout Siberut, wide variety in phonology...

- Distinguishing between final nasals and final voiceless stops and leveling to one or the other

Sagulubbe'	**Paipaijet**	**Saibi**	
uran	*uran*	*urat*	'rain'
kulin	*kulit*	*kulit*	'skin'

- Nasal-stop sequences and gemination

Sikabaluan	**Simatalu**	
antelu	*attelu*	'egg'
maniŋki	*manikki*	'thin'
makempu	*makeppu*	'thick'

- Stops versus trills, diphthong realizations, ...

(see Pampus 1989)

Mentawai phonology (Simalegi)

...as well as some remarkable divergence:

- Simalegi:
  - /s/, /t/ > /s/ _V[+front], /t/ else
  - /k/ > /s/ _[+front]
  - /g/ > /j/
  - /j/ > /ɖ/

Simatalu	**Simalegi**	
sapew	*tapew*	‘house’
ekew	*esew*	‘2s, you’
magila?	*majila?*	‘scared’
mabaja?	*mabaɖa?*	‘to buy’

(see Pampus 1994)

Nias phonology (South Nias)

Exhibits consonants not found elsewhere in Sumatra

		lab	alv	pal	vel	glot
	voiceless		t		k	ʔ
stop	**voiced**	b	d		g	
	trilled		d^r			
affricates	**voiceless**			ʧ		
	voiced			ʤ		
fricative	**voiceless**	f	s		x	h
	voiced	v				
nasal		m	n			
trill		ʙ	r			
lateral			l			
approximant		ʊ		j	w	

(see Brown 2001)

Nias morpho-phonology

Initial consonant mutation is used for...

- S for intransitive verbs

(1) a-ukhu nidanɣ.
STAT-hot water.MUT

'The water is hot.'

- P for transitive verbs

(2) la-bunu mbaβi.
3P.REAL-kill pig.MUT

'They killed a pig.'

- Possessors, prepositional arguments, ...

(see Brown 2001)

Nias morpho-phonology (cont.)

Mutation of nominals' initial consonants

Unmutated	Mutated	Unmutated	Mutated
/b/	/ʙ/	/f/	/v/
/d/	/dʳ/	/t/	/n/
/k/	/g/	/s/	/ʤ/
/g/		/ʧ/	

Unmutated	Mutated	
faxe	*vaxe*	'rice'
siʔo	*ʤiʔo*	'stick'
baβi	*ʙaβi*	'pig'
doi	*dʳoi*	'thorn'

(see Brown 2001)

Rejang morphology

- Rejang has *no suffixes*
  1. No applicative
  2. No nominalizer
  3. No circumfixes
- AV is marked by *m-/me-/-em-*
- PV is marked by *n-/ne-/-en-*

(3) a. Si k-əm-lea? imuo.
3SG <AV>see tiger

'He saw tigers.'

b. Imuo k-ən-lea? nə.
tiger <PV>see 3SG

'He saw tigers.' (McGinn nd)

Rejang morphology

Form	Type	Description
bə-	INTR	Intransitive verbs, ‘have [noun]’
mə(N)-	INTR	Intransitive verbs expressing locomotion
te-	NVOL	Non-volitional intransitive verbs
-əm-	AV	A-voice transitive verbs
-ən-	PV	P-voice transitive verbs
kə-$_1$	IMP	Imperative (transitive), intensive (intransitive)
kə-$_2$	NMZR	Abstract nouns
kə-$_3$	ORD	Ordinal numbers
kə-$_4$	ADVR	Verbs with adversative meanings
də- / sə-	NUM	Numeral ‘one’, ‘together with’, ‘same as’
pə-	NMLZ	Agent nominalizer
pe(N)-	NMLZ	Instrumental nominalizer

(McGinn nd)

Kerinci (Malayic) rime alternation

Alternation of final syllable rime:

Absolute (AB)	**Oblique (OB)**
1 gɨ ‘to hold’	1 gʌŋ ‘a handle’
2 mandâe͡ ‘bathe’	2 mandɨy ‘bathe someone’

1. Nominalizer
2. Applicative

→ Syntactic functions...

(see McKinnon 2011, Yanti, Mckinnon, Cole & Hermon 2018)

Kerinci (Malayic) rime alternation

(4) a. batɨ
tree.AB

‘tree.’

b. batʌŋ pina
tree.OB areca.nut.AB

‘areca nut tree.’ (McKinnon, Cole & Hermon 2011: 719)

Kerinci (Malayic) rime alternation

(5) a. ɲo tidə͡o lamo
3 sleep.AB long.AB

‘He slept for a long time.’

b. ka ndə͡o? makən nasae minɨn lah
1s FUT eat.OB rice.AB now just

‘I am going to eat rice right now.’

c. ka ndə͡o? maka minɨn lah
1s FUT eat.AB no just

‘I am going to eat right now.’ (meaning ‘eat s.t. unspecified’)
(McKinnon, Cole & Hermon 2011: 721-724)

Convergence

Phonological convergence (phoneme inventories)

Sustained contact with Malayo-Chamic languages have led to convergence in certain aspects of phonological inventories:

- **Near Universal**: Import and (re-)innovation of contrastive /c/ and /j/
- **Kluet**: Palatal nasal in nasal substitution
- **Gayo**: Nasal-stop sequences; /b/, /d/, /g/
- **NBI**: Loss of final /s/ with Minangkabau

Phonological convergence (Kluet)

Kluet occupies middle-ground between Acehnese and Minangkabau:

- Acehnese: Final /l/ > /y/ (retained elsewhere in Batak)

Karo Batak	**Kluet**	**Acehnese**	
akal	*ake*	*akai*	'intellect' (from Arabic)
kawil	*kawe*	*kawe*	'fishhook' (from PAN)
takal	*take*	—	'head' (from PBatak)

- Minangkabau: Final /a/ > /o/

Karo Batak	**Kluet**	**Minangkabau**	
bərita	*bəʀito*	*barito*	'news' (from Arabic)
mata	*mato*	*mato*	'eye' (from PAN)
sada	*sado*	—	'one' (from PBatak)

(see Neumann 1951, Ismail, Hanoum, Hanafiah & Nurdin 1990, Daud & Durie 1999, *Kamus bahasa Indonesia-Minangkabau* 2013)

Phonological convergence (Sigulai phonotactics)

Sigulai:

- Spoken alongside Simeulue (and Leukon) on Simeulue Island
- Closely related with Nias (coda consonant loss, loss of *r, ...)
- Phonotactics undergoing restructuring to match Simeulue

Nias	**Siguali**	**Simeulue**	
beβe	*befe*	*bifil*	'lip' (< *bibir)
ilu	*ilo*	*ilul*	'spit' (< *ilur)
ud^{r}e	*undin*	*udil*	'turmeric' (< *kundir)
—	*alun*	*salul*	'valley' (< *salur)

Nasal Syntax

Symmetrical Voice

- 'Indonesian-type' symmetrical voice system
  1. A-Voice (AV): A oriented construction
  2. P-Voice (PV): P oriented construction

Symmetrical voice in Nasal is basically the same as Malay (Kaur, SBM) (see McDonnell 2016) and differs slightly from Way Lima.

*Unclear of similarities/differences in other Lampungic languages (e.g., Krui).

A-Voice in Nasal and Kaur

Nasal AV *N-*

(6) *Nyak* hago n-(t)etuk pempang kayu sudi.
1SG want AV-chop log wood DEM.DIST

'I am going to chop those wood logs.'

Kaur AV *N-*

(7) *Teguh* lah ‘adu m-beli de‘ian.
T. PFV finish AV-buy durian

'Teguh already bought durian.'

P-Voice in Nasal and Kaur

Nasal PV – 1/2

(8) *Pempang kayu sudi* khadu ku=tetuk.
log wood DEM.DIST already 1SG=[PV]chop

'Those wood logs were chopped by me.'

Kaur PV – 1/2

(9) *De'ian ni* ku=beli.
durian DEM.PROX 1SG=[PV]buy

'I bought the durian.'

P-Voice in Nasal and Kaur

Nasal PV – 3 (*di-*)

(10) *Pempang kayu sudi* khadu (di-)tetuk=nyo.
log wood DEM.DIST already PV-chop=3

'Those wood logs were chopped by him.'

Kaur PV – 3 (*di-*)

(11) *De'ian* lah 'adu (di-)beli Teguh.
durian PFV finish PV-buy T.

'Teguh already bought the durian.'

Nasal Voice

Symmetrical voice for Nasal, SBM, and Kaur can be summarized in the same way.

Symmetrical Voice Summary

Voice		Form	Primary Argument	Secondary Argument
AV		*N*-ROOT (homorganic nasal)	*A*	P
PV	1-2:	PRO(=)V	*P*	A
	3:	(*di-*)ROOT=pro (*di-*)ROOT noun		

Convergence in Nasal morphology

The inventory of affixes in Lampungic and Malay is similar

→ The vast majority of forms can be attributed to either language

Exceptions

- Nasal prefix *m(e)-* still found in Lampungic *m(a)-*
- Nasal infix *-em-* with limited productivity found in frozen forms in Lampungic
- Nasal prefix *te(k)-* maintains a lexically-conditioned allomorph *tek-* that appears to come from Lampungic.
- Frozen Nasal prefix *mi-* attaches to kinship terms likely from Lampungic.

Comparison of verbal morphology

Nasal	**Way Lima**	**Kaur**	
N-	*N-*	*N-*	A-VOICE
di-	*di-*	*di-*	P-VOICE
te-	*ta-*	*te-*	NON-VOLITIONAL
tek-†	*taka-*/*kata-*†	—	NON-VOLITIONAL
—	*ti-*	—	PASSIVE
—	*ka-*	—	PASSIVE
m(e)-	*ma-*	—	INTRANSITIVE
-em-†	—	—	INTRANSITIVE
be-	*ba-*	*be-*	MIDDLE VOICE
-i	*-i*	*-i*	LOCATIVE APPLICATIVE
-kun	*-ko*	*-ko*	BENEFACTIVE APPLICATIVE
mi-†	*mi-*†	—	KINSHIP

† = not productive

Nasal lexicon

Nasal open-class lexical items...

Nasal	**Lampungic**	
kaci	*kaci*	‘dog’
lapah	*lapah*	‘walk’
bahun	*bahun*	‘house’
balak	*balak*	‘big’

Nasal	**Kaur**	
hujan	*hujan*	‘rain’
kekho	*keʻe*	‘monkey’
bukhung	*buʻung*	‘bird’
gekheman	*geʻeman*	‘molar tooth’

...and closed-class lexical items...

Body parts	✓	TAM	✓
Personal pronouns	✓	Numerals	✓
Basic interogatives	✓	Negation	✓

... historically borrowed from Lampungic and Malayic
Ongoing developments leading to greater convergence with Malayic (replacement of relativizer, 3s pronoun)

Implications

‘Vergence’ in Sumatra & the Barrier Islands

Divergent

Generally found in language ecologies characterized by geographical and linguistic isolation.

- Not simply genetic differences
  - e.g. divergent Kerinci features, convergent Nasal features
- Difficult to pin down what linguistic/extra-linguistic factors lead to convergence/divergence.
- Language contact may lead to divergence as in Acehnese
  - e.g. has been argued for Enggano

Convergent

Found in language ecologies characterized by multilingualism or otherwise higher degrees of ‘contact’.

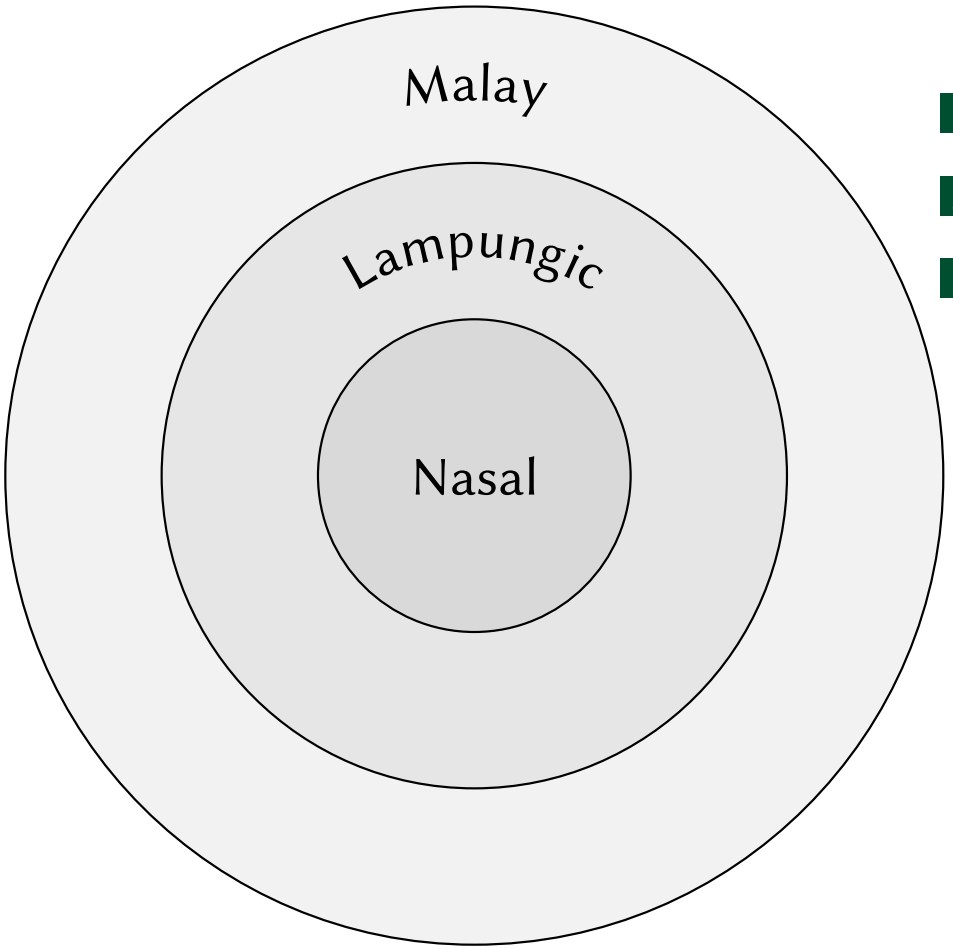
> *“while language diversification is a matter of diffusion and divergence, diffusion and divergence (or convergence, for that matter) among language varieties is a matter of language contact.”*
>
> (Espree-Conaway 2022: 58–59)

‘Vergence’ and the history of Sumatra

Divergence informs the historical development of individual languages and language groups.

Convergence informs the history of language contact and multilingualism.

Language Contact and Nasal



Nasal essentially has three layers

1. Nasal (core)
2. Lampungic (earlier influence)
3. Malay (more recent influence)

Rejang's contact with Malay

Quote from McGinn (nd):

"Rejang is indeed "heavily Malayicized" as consequence of recent contact history, and yet it remains a linguistic isolate. It has closer typological affinity with distant Acehnese, Bidayuh and Melanau than with its geographical neighbors, including Bahasa Indonesia, numerous contemporary Malay dialects, and the Malayic languages Kerinci and Minangkabau."

Conclusions and future prospects

1. Linguistic diversity in Sumatra and the Barrier Islands is often overlooked, but this region is anything but straightforward.
2. Language ecologies in the region are complex and, at the same time, rapidly changing.
3. Divergent patterns in these languages have the potential to reveal important patterns of language evolution.
4. Convergent patterns uncover complex histories of language contact and, given the present trajectory, provide a window into the future linguistic development of the region

Acknowledgements

We would like to thank the following organizations and people.

1. The Nasal community from Tanjung Betuah and Gedung Menung, especially Johan Safri and Wawan Sahrozi.
2. Center for Language and Culture Studies, Atma Jaya Catholic University of Indonesia, especially Yanti.
3. National Research and Innovation Agency (BRIN) for supporting this research in Indonesia, especially Obing Katubi and Dendi Wijiya.

This material is based upon work supported by the National Science Foundation under grant BCS-1911641 to the University of Hawaiʻi at Mānoa. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References I

Billings, Blaine & Bradley McDonnell. 2024. Sumatran. *Oceanic Linguistics* 63(1). 112–174.

Blust, Robert. 2013. *The Austronesian languages*. Revised (Pacific Linguistics 602). Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, Australian National University. 824 pp.

Brown, Lea. 2001. *A grammar of Nias Selatan*. Sydney: University of Sydney dissertation.

Daud, Bukhari & Mark Durie. 1999. *Kamus basa Acèh*. Canberra: The Australian National University.

Edwards, Owen. 2015. The Position of Enggano within Austronesian. *Oceanic Linguistics* 54(1). 54–109.

Espree-Conaway, DeAndré A. 2022. *Evolutionary cartographies of language diversification: quantitative approaches to the geolinguistic mapping of the kayanic languages (central borneo)*. Grand Forks: University of North Dakota Master's thesis.

Ismail, M. Ridwan, Syarifah Hanoum, Anwar Hanafiah & Bakhtiar Nurdin. 1990. *Sistem morfologi bahasa Kluet*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa.

Kamus bahasa Indonesia-Minangkabau. 2013. Padang: Balai Bahasa Sumatera Barat.

Langgang, Zainuddin HR. 1992/1993. *Laporan penelitian: Geografi dialek bahasa Mentawai*. Padang: Pusat Pembinaan dan Pengembangan Bahasa.

McDonnell, Bradley & Christina L Truong. 2024. The Non-Malayic languages of Sumatra and the Barrier Islands. In Alexander Adelaar & Antoinette Schapper (eds.), *The Oxford Guide to the Malayo-Polynesian Languages of Southeast Asia*, 409–430. Oxford ; New York: Oxford University Press.

References II

McDonnell, Bradley J. 2016. *Symmetrical voice constructions in Besemah: A usage-based approach*. Santa Barbara: University of California, Santa Barbara dissertation.

McGinn, Richard. 2005. What the Rawas Dialect Reveals about the Linguistic History of Rejang. *Oceanic Linguistics* 44(1). 12–64.

McGinn, Richard. nd. A grammar of Rejang.

McKinnon, Timothy. 2011. *The Morphology and Morphosyntax of Kerinci Word Shape Alternations.* Newark, Delaware: University of Delaware dissertation.

McKinnon, Timothy, Peter Cole & Gabriella Hermon. 2011. Object Agreement and 'Pro-Drop' in Kerinci Malay. *Language* 87(4). 715–750. (16 February, 2023).

Neumann, J. H. 1951. *Karo-Bataks-Nederlands woordenboek*. Medan: VAREKAMP & Co.

Nothofer, Bernd. 1986. The Barrier Island Languages in the Austronesian Language Family. In Paul Geraghty, Lois Carrington & S A Wurm (eds.), *Focal II: Papers from the Fourth International Conference on Austronesian Linguistics* (Pacific Linguistics: Series C 94), 87–109. Canberra: Department of Linguistics, Research School of Pacific Studies, The Australian National University.

Pampus, Karl-Heinz. 1989. Zur dialektgeographischen Gliederung des Mentawai-Archipels. In Wilfred Wagner (ed.), *Mentawai: Identität im Wandel auf indonesischen Außeninseln*, 61–102. Frankfurt am Main: Frobenius-Institut.

Pampus, Karl-Heinz. 1994. Sabirut-Taburit 'Siberut': Die Lautgesetze der Simalegi-Sprache im Mentawai-Archipel (westsumatra). *Zeitschrift der Deutschen Morgenländischen Gesellschaft* 144(1). 114–129.

References III

Smith, Alexander D. 2017. The Western Malayo-Polynesian Problem. *Oceanic Linguistics* 56(2). 435–490.

Yanti, Timothy Mckinnon, Peter Cole & Gabriella Hermon. 2018. The Phonological Basis of Syntactic Change in Kerinci. *Oceanic Linguistics* 57(2). 433–483.