McDonnell, Bradley, Blaine Billings & Yanti. to appear. The Languages of Sumatra and the Barrier Islands. In Bill Palmer & Ellen Smith-Dennis (eds.), *Languages and Linguistics of Island Southeast Asia and the Pacific.* Berlin: De Gruyter Mouton.

The Languages of Sumatra and the Barrier Islands

Bradley McDonnell
University of Hawai‘i at Mānoa

Blaine Billings
University of Hawai‘i at Mānoa

Yanti
Atma Jaya Catholic University of Indonesia

1 Introduction

This chapter provides an overview of the Malayo-Polynesian languages of Sumatra and its satellites. Sumatra is situated between mainland southeast Asia via Malaysia to its east and Java to its southeast, being separated from each by no more than forty miles at their closest points. It is the third largest island of Indonesia in terms of land mass and the largest island to be fully encompassed within Indonesia. Though it is more than three times the size of Java, the most populous island of Indonesia, it barely has a third of the population. The island is split vertically by the Barisan Mountain Range (*Bukit Barisan* in Indonesian) situated along its western half. While this geography creates vast areas of fertile highlands and swampy lowlands connected through a complex network of rivers in the east, the west coast reaches directly to the mountain range with no more than thirty miles separating the two, often much less. Island chains are located along both sides of Sumatra. Along the east coast lie Bangka and Belitung in the south with the Riau Archipelago further north. Running nearly the whole length of Sumatra's west coast are the Barrier Islands, a chain of seven major islands and their associated archipelagos, the largest and most populated of which is Nias.

The languages of Sumatra and the Barrier Islands, henceforth SBI languages, are exclusively Malayo-Polynesian but are nonetheless diverse, representing a complex history of migration and multilingualism.[1] SBI languages fall into one of four genetic groups: Malayo-Chamic, Sumatran, Lampungic, and Rejang (see Figure 1). Sumatran languages are most widespread and are located in the highlands of the northern half of Sumatra (Gayo [gayo1244, gay], the Batak languages), all of the Barrier Islands (Simeulue, Haloban, and Leukon, assigned a single language code [sime1241, smr]; Sigulai [siku1242, skh], Nias [nias1242, nia], Mentawai [ment1249, mwv], Enggano[2] [engg1245, eno]), and three coastal villages in southern Bengkulu province (Nasal [nasa1239, nsy]). The Lampungic languages are dominant both coastally and inland on the southern tip of Sumatra, while Rejang [reja1240, rej] is spoken in northern Bengkulu province from the southwestern coast to the Barisan highlands. Malayo-Chamic extends far beyond Sumatra. On Sumatra it includes one Chamic language, Acehnese [achi1257, ace], which is spoken in the coastal regencies of the northernmost parts of the island, and the

[1] A number of other languages are additionally spoken in the region, most notably Javanese in the southern part of Sumatra and Hokkien in eastern and northern Sumatra and the Riau Archipelago, but these are not discussed here given that they are not indigenous to the region.

[2] Although the inclusion of Enggano has historically been a matter of debate, it is now well established as an Austronesian language (Edwards 2015), with recent proposals placing it within Sumatran (Smith 2017; Billings & McDonnell 2024).

Malayic languages, which dominate the eastern half of the island, the central highlands, and much of the western coast, filling in the space between Lampungic and Rejang and the wide region between Rejang and the Batak languages.

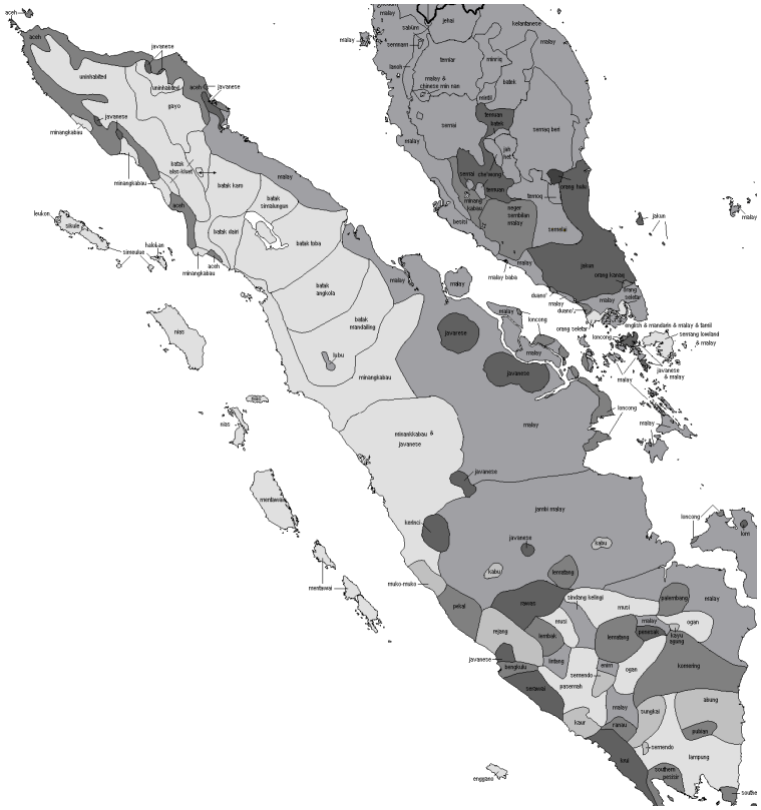

Figure 1. Languages of Sumatra and the Barrier Islands.

The Malayic languages comprise large networks of dialects throughout Sumatra, representing the largest language group both in terms of number of speakers and number of languages. Given its strategic location alongside the lucrative trading strait of Malacca, Sumatra has historically been home to some of the most powerful maritime kingdoms of the region, most notably Melayu (7th century) and Srivijaya (7th–13th centuries). It is from this success that Malay spread from Sumatra in all its diverse regional varieties and became prevalent especially in the eastern half of the island. The prevalence of the Acehnese language in northern Sumatra is likely the result of the influence of the Aceh Sultanate from the 16th–18th centuries. In present day Sumatra, both Standard Indonesian [indo1316, ind] and other varieties of Indonesian, which are often described as colloquial, regional, or koiné varieties, play an important role in society. The prestige and broad usage of Indonesian has additionally led to a rising number of native speakers, whether alongside or in replacement of regional languages (Adelaar & Prentice 1996; Adelaar 2005a; Tadmor 2018).

The rest of this chapter is organized as follows. Section 2 describes the phylogenetics of the SBI languages, including recent proposals for a large Sumatran subgroup. Section 3 discusses language contact, multilingualism, and vitality, and Section 4 presents a brief typological overview of the languages of the region. The final three sections present sketch grammars of three languages of the region, namely Jambi Malay [jamb1236, jax] (Section 5), Rejang (Section 6), and Mentawai (Section 7).

2 Phylogenetics

SBI languages are all well established as members of the Malayo-Polynesian branch of the Austronesian language family. They are further classified as Malayo-Chamic (Thurgood 1999), Sumatran (following Smith 2017; formerly Nothofer's 1986, 1994 Batak-Barrier Islands), Lampungic (Anderbeck 2007a), and Rejang (McGinn 2005). Aside from Malayo-Chamic, these have not been conclusively subgrouped with other languages outside the region (cf. McGinn 2005 et seq. on Rejang). The distribution of Sumatran, Rejang, and Lampungic on the one hand and Malayo-Chamic on the other suggests an intriguing history of the region, wherein the former groups represent an earlier settlement and the latter a later one. Blust (2013: 77–78) hypothesizes that "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".

The history of the non-Malayo-Chamic SBI languages is largely unknown. As Blust's quote points out, though, the present-day distribution of the languages is more readily explained if non-Malayic languages had been present in Sumatra prior to the arrival of Malayic languages (see also Smith 2017: 460–461). For languages of mainland Sumatra, the arrival of Malayic languages was followed by a period of extensive contact that continues to this day (Blust 1984, 1994; Anderbeck & Aprilani 2013). In addition to these broadly accepted theories about the linguistic history of Sumatra, there have also been multiple proposals suggesting that pre-Austronesian (possibly Austroasiatic) groups first inhabited the region and that their languages form a substrate for the Austronesian languages now present (Weber 2019; see Blench 2014 for Enggano specifically). Though much work still remains to be done in order to validate such proposals, the possible effects of earlier non-Austronesian-speaking inhabitants of Sumatra and the Barrier Islands on later Austronesian settlers cannot be immediately ruled out.

The proposals that form the basis for each of the groups in the four-way division of the SBI languages are discussed in turn below. We begin first with those language families which have relations outside the region (Malayo-Chamic) and then proceed with the three remaining subgroups in the order of greatest to least linguistic diversity within each (Sumatran, Lampungic, Rejang).

2.1 Malayo-Chamic

The Malayo-Chamic subgroup is a family consisting of two distinct branches, Malayic and Chamic. Although they likely originated in Borneo, the two branches have diverged considerably since the migration of Chamic to mainland Southeast Asia. The connection between the two branches is found primarily in phonological evidence, where ten shared innovations have been identified (Adelaar 2005b: 373–374), as well as a number of lexical innovations. Most important of these shared lexical innovations are the numerals *tujuh 'seven', *dua-lapan 'eight', and *sa-lapan/*sa-ambil-an 'nine', these last two both being subtractives formed from the roots *alap and *ambil meaning 'to fetch/take' (Blust 1981: 467). While there have been multiple proposals arguing for other languages/subgroups being closely related to Malayic (Nothofer's 1975 Malayo-Javanic) or being included within this family (Adelaar's 2005b Malayo-Sumbawan), the only proposal that has been conclusively argued is the close relationship between Malayic and Chamic.[3]

[3] More recent proposals place Malayo-Chamic under a Greater North Borneo family, itself nested under a Western Indonesian branch of Malayo-Polynesian (Blust 2010, Smith 2017).

The Malayic languages in Sumatra, as elsewhere, form a broad web of interconnected isolects that, due to sustained contact and resulting overlapping phonological and lexical innovations, have proven difficult to subgroup. Within Malayic, a distinction has sometimes been made between the terms *Malayic* and *Malay* (Blust 1988; Nothofer 1988; Adelaar 1992), the first referring to all languages of the subgroup as characterized by fourteen shared phonological innovations (Adelaar 2005b), the second to the lower-level group within Malayic (often including, for example, languages like Jambi Malay and South Barisan Malay [cent2053, pse] while excluding others such as Iban [iban1264, iba] and Urak Lawoi' [urak1238, urk]). Though this distinction is important elsewhere, for the sake of the Malayic languages of Sumatra, all are firmly classified under Malay, with the exception of more aberrant varieties such as Mingangkabau [mina1268, min] and Kerinci [keri1250, kvr]. Convincing arguments have been made for further subgrouping of Malay isolects within Sumatra, as with, for example, Jambi Malay and its Upstream and Downstream varieties (Anderbeck 2008) as well as South Barisan Malay and Musi Malay [musi1241, mui] spoken in southern Sumatra (McDowell & Anderbeck 2020). Though much progress has been made in recent years on the classification of Malay isolects, the topic remains widely discussed, and many questions remain on the historical development of Malay in Sumatra.

Since Thurgood's (1999) reconstruction of Proto-Chamic, the status of Acehnese as a Chamic language has been widely accepted. As explained by Thurgood (1999: 22–23), it is clear from shared innovations that as early as 1000 years ago, Chamic speakers migrated from mainland Southeast Asia to northern Sumatra (cf. Sidwell 2005, 2006). Achenese has since gained regional prominence and is used by more speakers than all other Chamic languages combined. As such, Acehnese has had significant effects on the bordering languages of Gayo, Kluet (assigned the same language code as Alas [bata1292, btz]), and the variety of Minangkabau spoken in Aceh referred to as Aneuk Jamee. Additionally, since "re-entering the Austronesian world", as Thurgood (1999: 23) put it, Acehnese has itself been influenced by contact and trade relationships with Malayic-speaking peoples. These influences are clear when compared against the Chamic languages still present in mainland Southeast Asia, where Austronesian influence has been largely absent and Austroasiatic influence has progressed further since the separation of Acehnese.

2.2 Sumatran

Sumatran is a diverse family of languages spoken in the highlands of northern Sumatra (Gayo, the Batak languages), the Barrier Islands off Sumatra's west coast (Simeulue/Devayan, Haloban, Leukon, Sigulai, Nias, Mentawai, Enggano), and in three coastal villages in southern Bengkulu province (Nasal). Close relationships between the Batak languages and some languages of the Barrier Islands have long since been proposed (Lafeber 1922), but more thorough argumentation was provided by Nothofer (1986, 1994) who referred to the subgroup as Barrier Island-Batak. Though the proposal was supported by Smith (2017: 459), it was noted that evidence for such a subgroup "rests on a single sound change", Nothofer's (1986) PMP *j, *g > Batak-Barrier Islands *g. Billings & McDonnell (2024) have since argued that the family can be established on the basis of six shared phonological innovations and includes both Nasal (excluded from Nothofer 1986) and Gayo (excluded from Nothofer 1986 and Smith 2017). Notably, this subgroup is argued to also include Enggano. Though definitively established as an Austronesian language only recently by Edwards (2015), contra Capell (1982) and Blench

(2014), Enggano's relationship with other Malayo-Polynesian languages has long been a topic of debate (Dyen 1965; Nothofer 1986; Edwards 2015; Smith 2017).

2.3 Lampungic

The three Lampungic dialect clusters of Lampung Api [lamp1243, ljp], Lampung Nyo [lamp1242, abl], and Komering [kome1238, kge] spoken in southern Sumatra form a subgroup on the basis of fifteen shared defining characteristics (Anderbeck 2007a: 42). While these three distinct Lampungic varieties are well established as close relatives, the Lampungic family as a whole has no such clear affinities with the languages spoken around it. Initially argued to be closely related to Malayic, Javanese, Madurese, and Sundanese (Dyen 1965; Nothofer 1985), this proposal has since been largely disfavored (Blust 1994). Arguments for its relationship with these and other languages were summarized and thoroughly reviewed by Anderbeck (2007a) in his reconstruction of Proto-Lampungic. He concludes that current evidence is insufficient for claiming that Lampungic forms any close subgroup of PMP with Malayic, Batak (and via the same argumentation, Sumatran), or Rejang as well as nearby Sundanese or Javanese. Based on current evidence, the only feasible post-PMP subgrouping proposal is Smith's (2017: 459) broader Western Indonesian family which includes Lampungic on the basis of a merger of PMP *j, *d > Western Indonesian *d and a few shared lexical innovations. Even this high-level subgrouping, however, is contentious, as Smith (2017: 459) himself states "[i]f Lampung is a Western Indonesian language, it shows little evidence of the fact".

2.4 Rejang

The most significant work in determining Rejang's place in Malayo-Polynesian has been carried out by McGinn (2003, 2005, 2009) and Blust (1984). With a thorough evaluation of Rejang's phonological history, McGinn (2009) concludes that Rejang is related to the Bidayuh languages spoken in Western Borneo and posits a migration out of Borneo around 1000 years ago. This conclusion is based primarily on the striking similarity between vowel changes undergone by Rejang and Bukar-Sadong as well as Rejang's verbal morphology (see Section 6.2), which, though unusual when compared to all other languages of Sumatra, finds correspondences in Bukar-Sadong. Despite such arguments, however, subgrouping relationships between Rejang and the Bidayuh languages remain contentious for three primary reasons. First, given the "prolonged history of unilateral lexical borrowing" from Malay into Rejang (Blust 1984: 423), it has proven difficult to identify native Rejang versus Malay influence, notably in the identification of reflexes of PMP *z, *j, *R, and *r. Second, the astounding number of sound changes Rejang has undergone, especially in its vowel inventory, have further complicated the process of identifying PMP inheritances. Third, though the vowel correspondences described by McGinn (2009) in sum are many, none of the shared innovations independently provides strong evidence for the Rejang-Bidayuh relationship. Thus, though McGinn (2009) provides an extensive argument for the connection between Rejang and the Bidayuh languages, such a proposal remains controversial, as evident in the response by Adelaar (2007). This is also apparent by the fact that both Blust (2010) and Smith (2017) discuss Rejang, though considered a member of the Greater-North Borneo family, independently of the languages of Borneo.

3 Language contact and language vitality

In present-day Sumatra, the use of Malay is ubiquitous. Standard Indonesian is used to different degrees in education, government, media, and a variety of formal situations, such as

weddings, funerals, and sermons. In urban centers and market towns throughout Sumatra, regional varieties of Malay-Indonesian act as lingua franca. Some of these varieties, such as Riau Indonesian [riau1234, zlm], Jambi Malay, and Palembang Malay [pale1264, mui], have been codified and can be described as koinés (Tadmor 2018; Gil 2020). In areas of western Sumatra and the northern and central Barrier Islands, it is varieties of Minangkabau — such as Aneuk Jamee, spoken on the west coast of mainland Aceh province as well as on Simeulue and the Banyak Islands — that act as lingua franca (Aziz & Amery 2016).

In areas where indigenous vernacular Malay(ic) varieties or non-Malayic languages are spoken, people are often multilingual (or multilectal). Standard Indonesian is often used alongside a colloquial/regional variety of Malay-Indonesian in addition to the vernacular. These situations can be characterized as polyglossic, where the indigenous language is the basilect, a regional lingua franca is the mesolect, and Standard Indonesian is the acrolect. For example, in the highlands of southwest Sumatra, Besemah is the basilectal vernacular Malay, which is used in daily life among Besemah people. A mesolectal regional variant of Palembang Malay, which could be referred to as Pagaralam Malay, is used in the market town of Pagaralam and in some cases for interethnic communication throughout the region. Standard Indonesian acts as the acrolect, being used in education and formal ceremonies like weddings and funerals. In Sumatra's southernmost province, Lampungic varieties are used in the home for most daily communication, and Indonesian is used in formal settings, such as school and religious ceremonies, and to some extent for inter-ethnic communication (Katubi 2007).

In a few communities, there is a mix of multilingualism and polyglossia. In the Nasal speech community, which is concentrated in three villages along the coast of southwest Sumatra, Nasal, a Sumatran language, is spoken alongside Kaur [kaur1269, vkk] and Semende [seme1248, sdd], two closely related vernacular Malays, and Standard Indonesian. On the island of Simeulue, Leukon speakers are also able to use Sigulai and Simeulue (all Sumatran languages), but the inverse is not true. The lingua franca Aneuk Jamee has been growing in use for trade and business over the past few decades, both on the island of Simeulue and the Banyak Islands where Haloban and Nias are also spoken. Standard Indonesian here, as elsewhere, is used in other official capacities (Aziz & Amery 2016).

This contact, both among Malay(ic) varieties and between Malay(ic) and non-Malayic SBI languages, is not a recent phenomenon. Contact among Malay(ic) varieties is extremely difficult to detect due to the genetic closeness of the isolects. In urban centers, contact among Malay(ic) isolects has resulted in the process of koineization. In some cases, the contact is so extensive that it has led a number of linguists to classify a non-Malayic language as Malayic (see e.g. Blust 1984 for discussion of Voorhoeve's 1955 classification of Rejang). In southern and central Sumatra, use of Malay was associated with powerful early states, Melayu and Srivijaya, dating back to the seventh century (Blust 2013: 18), and the influence of Malay among Rejang, Nasal, and Lampung-speaking groups is evident in their lexicons and likely their grammars (Anderbeck 2007a; Anderbeck & Aprilani 2013; Blust 1984). In northern Sumatra, Malay was the official language of government and education under the Acehnese Sultanate from at least the 15th century (Durie 1985). This intensive contact with Malay is present in Acehnese but even more so in Gayo, which was ruled by the Acehnese sultanate. It appears that Batak populations, on the other hand, were never firmly incorporated under Malay or Acehnese rule. However, while the Batak languages show less early influence from Malay than Gayo and Acehnese, they still show significant Malay influence (Adelaar 1995). Even in Enggano, there are a number of

examples of early borrowings from Malay, as in Enggano *e-dayada* 'sail' < Malay *layar* and Enggano *e-añãmũ* 'chicken' < Malay *ayam*, among many others (Kähler 1987).

There is little evidence of contact among the non-Malayic languages. Clearest among these is Nasal, which has historically been in close contact with Lampungic. This is evident in a large number of lexical borrowings in the language (Anderbeck & Aprilani 2013). The Sumatran languages Gayo and Kluet have both, to varying degrees, been influenced by neighboring Acehnese. This is again clearest in lexicon borrowings, although the change of Kluet word-final *l > /i̯/, unique among the Batak languages, was likely also brought on by contact with Acehnese. Two interesting contact situations wherein Malay varieties have come in contact with non-Malayic languages include Haji [haji1235, hji], a Malay variety spoken in Lampung province, and Duano [duan1242, dup], a Malay variety spoken on the east coast of Sumatra in Riau and Jambi provinces. A full third of the vocabulary in Haji is the result of borrowing from Lampungic (Anderbeck 2007b), and according to Anderbeck (2012: 273), "Duano is likely the most aberrant Malayic lect in terms of its lexicon", most likely as a result of non-Malayic substratum.

The majority of SBI languages appear to be vital with continued intergenerational transmission. These languages are spoken by relatively large populations, most commonly in tens or hundreds of thousands and, in several cases, by millions of people. However, there are a few notable exceptions: Leukon, Haloban, Enggano, and Nasal. All of these are spoken by smaller populations and appear to be undergoing language shift (Aziz & Amery 2016; Hemmings & Arka 2023). Other issues of vitality concern the ongoing process of koineization that is leading to the loss of distinctive features of Malay(ic) varieties.

4 Typological overview

SBI languages are highly diverse, making it difficult to provide a typological profile in any succinct way. This diversity is the result of several factors. First, four distinct language subgroups are represented in Sumatra, all having undergone unique innovations particular to each. Second, several languages are outliers, presumably caused by geographical isolation. These may include Kerinci (a cluster of Malayic varieties spoken in the highlands of central-western Sumatra), Rejang, and the languages of the central and southern Barrier Islands (Nias, Mentawai, Enggano).[4] Third, Acehnese stands unique among all other languages of Sumatra due its historical contact with Mon-Khmer languages on mainland Southeast Asia (Thurgood 1999). Finally, if pre-Austronesian speaking peoples were present on Sumatra and the Barrier Islands, linguistic interactions would have left their effects primarily on earlier Austronesian settlers (Sumatran and potentially Lampungic and Rejang) and to a much lesser degree on the more recently settled Malayo-Chamic languages. Unless otherwise noted, the descriptions below draw on the following sources: Besemah (McDonnell 2016), Kerinci (McKinnon 2011; Ernanda 2017), Acehnese (Durie 1985, Thurgood 1999), Lampung (Walker 1976), Alas (Soravia 2007), (Southern) Nias (Brown 2001), Enggano (Yoder 2014). For references for Jambi Malay, Rejang, and Mentawai, see the respective grammar sketches.

4.1 Phonology

The majority of the languages of Sumatra have similar phonological inventories, and that of the Batak language Alas in Tables 1 and 2, can be taken as emblematic of the region. Deviations from this inventory are most frequently encountered in (i) the phonetic realization of

[4] Another explanation for Rejang is that it originates from western Borneo (see McGinn 2009).

the rhotic /r/ (see, e.g., Jambi Malay in Section 5 where the rhotic is realized as [ʁ]), (ii) the presence of phonemic /ʔ/ (see, e.g., Jambi Malay in Section 5 and Rejang in Section 6), and (iii) the differences between the quantity and quality of mid vowels (e.g., Mentawai lacks /ə/, Section 7) and diphthongs (see Section 6 with numerous diphthongs in Rejang). Additionally, a number of languages across the region are cited as having distinctive post-occluded nasals, also known as *funny nasals* or *barred nasals* (see e.g. Jambi Malay in Section 5 and Rejang in Section 6; see also Durie 1985: 15 for Acehnese).

Acehnese, Enggano, and Nias present perhaps the greatest divergence from this inventory. Acehnese is more in line with other Chamic languages and stands out against the rest of Sumatra primarily in its vowel inventory, where distinctions are made between high-mid and low-mid vowels, oral and nasal vowels, and ten diphthongs (Durie 1985: 16). Enggano has phonemicized nasality (though on a morpheme rather than segmental level) and has a contracted consonant inventory with only seven phonemic contrasts (Yoder 2014: 32). Nias follows yet a different direction, employing three cross-linguistically uncommon phonemes: a voiced bilabial trill /ʙ/, a voiced alveolar stop with trilled release /dʳ/, and a voiced labiodental approximant /ʋ/ (contrastive with both /w/ and /v/; Brown 2001: 22).

Table 1. Alas consonant inventory.[5]

		Labial	**Dental~Alveolar**	**Palatal**	**Velar**	**Glottal**
Stops	**Voiceless**	p	t		k	
	Voiced	b	d		g	
Affricates	**Voiceless**			t͡ʃ <c>		
	Voiced			d͡ʒ <j>		
Fricatives			s			h
Nasals		m	n	ɲ <ny>	ŋ <ng>	
Liquids			l			
Rhotics					ʀ <r>	
Glides		w		j <y>		

Table 2. Alas vowel inventory.

	Front	**Central**	**Back**
High	i		u
Mid	e	ə	o
Low		a	

The typical root structure for the languages of Sumatra is disyllabic with optional onsets and codas. Palatal consonants (excluding the glide /j/) and voiced stops are rarely found syllable-finally, though Simalungun Batak [bata1288, bts] (Adelaar 1981: 10) and Enggano are notable exceptions to the latter. The distribution of consonants and vowels is typically similar in both

[5] Though phonologically the distinctions between /p/~/b/ and /t/~/d/ primarily concern voice, unique among SBI languages, these are also phonetically differentiated by the fact that /b/ and /d/ are implosive [ɓ] and [ɗ].

syllables with some deviations from this discussed below. First, phonetic diphthongs are often phonemically analyzed as vowel-glide sequences (Clynes 1997; Blust 2013) and are typically restricted to /aw/, /aj/, and /uj/. The glides pattern as other non-nasal consonants and thus only occurs as coda in final syllables. This is the case, for example, with Lampungic (Anderbeck 2007a: 51). In several languages, however, the inventory of true diphthongs has expanded greatly with some also occurring in closed syllables, as in Kerinci (12 diphthongs; Ernanda 2017: 44), Acehnese (10 diphthongs; Durie 1985: 17), Enggano (six diphthongs; Yoder 2014:32), and Rejang (see Section 6). Second, root-medial clusters are typically restricted to homorganic nasal-obstruent sequences (including *ŋs*) as well as rhotic-obstruent *r*C and liquid-obstruent *l*C clusters. Acehnese is particularly unique in that it has a clear distinction between pre-syllable and main syllable structure, as is the case with other Chamic languages (Thurgood 1999: 67, 104). Acehnese main syllables permit complex consonant cluster onsets of the form obstruent + /h/, /r/, /l/ not encountered elsewhere in Sumatra. Third, mutation of PMP *ə in final closed syllables and PMP *a in final open syllables has occurred all across the region, thus creating an asymmetry in the penultimate/ultimate distribution of these two phonemes (see Tadmor 2003 for an in-depth discussion of final /a/ mutation). Finally, Enggano and the Central Barrier Islands languages, Sigulai[6] and Nias, all historically lost coda consonants, presumably via multiple sequential rather than simultaneous changes, though only Nias still reflects this in its phonotactics. Sigulai has since regained coda consonants via an influx of loanwords from Northern Barrier Islands languages and Minangkabau (e.g., Sigulai *tenap* < Simeulue *tenap* 'nail', Sigulai *rabung* < Minangkabau *rabuang* 'bamboo shoot'), and Enggano has further lost word-final vowels and re-innovated word-medial codas (e.g., Modern Enggano *'akər* < Old Enggano *'akolu* < PMP *təlu 'three').

The most emblematic phonological process for Austronesian languages is nasal substitution, and, as such, is likewise encountered all across Sumatra via voice-alternating and nominalizing affixes (Blust 2004; see the sketch grammars in Sections 5–7). In addition to this, in languages that disfavor or disallow *r*C sequences, some vowel-final prefixes will surface with a final *-r* only in vowel-initial bases (Eades 2005: 35 for Gayo; Soravia 2007: 29 for Alas; McDonnell 2016: 33 for Besemah; Ernanda 2017: 63 for Kerinci).

Nias and Kerinci both exhibit morphophonological processes not seen elsewhere in the region. Nias nouns and pronouns undergo a process of initial consonant mutation to denote a variety of grammatical relations (see below). This mutation typically involves voicing of voiceless segments and trilling for voiced segments (e.g. /b/ → /ʙ/, /d/ → /dʳ/, /g/ remains unchanged), though the process is mostly irregular for pronouns (Brown 2001: 73).

Varieties of Kerinci are well known for possessing a typologically unusual alternation whereby the rime of the final syllable alternates between what have been termed *absolute* and *oblique* forms. While the choice of absolute or oblique is determined by apparent morphological and syntactic properties (see below), the phonological shape of the final syllable rime is the result of a historical sound change that depended upon the shape of the rime and consonant phonation (i.e., the presence of voiced obstruents /b/, /d/, /g/, or /ɟ/ in the word). The alternations are complex and synchronically idiosyncratic, but Prentice & Usman (1978) found that words with voiced obstruents have higher vowels than those that do not meet this condition, as in Table 3. See McKinnon (2011: Ch. 3) for detailed discussion of the historical development of this alternation in the Tanjung Pauh Mudik (TPM) variety.

[6] Word-medial nasal codas, however, were not entirely lost in Sigulai as they were in Nias and Enggano.

Table 3. Syllable rime in TPM Kerinci.

	Proto-Malayic	Oblique	Absolute	Gloss
Voiced obstruent	*-*o*	*mudu*	*mudɨw*	'young'
	*-*il*	*dʌe*	*dɨn*	'pistol'
No voiced obstruent	*-*o*	*kato*	*katəo*	'word'
	*-*il*	*kaɲčɛ*	*kaɲčəe*	'mouse deer'

4.2 Morphology

SBI languages are predominantly prefixing, and the prefixes employed are greatly varied in form and have a wide range of meaning. Nearly all languages of the region, with the notable exception of Acehnese, utilize two prefixes that trigger nasal substitution. These prefixes frequently have a non-nasal counterpart with similar form but different meaning (e.g., Simalungun Batak *maN-* vs. *mar-*, Malayic *peN-* vs. *per-*, Lampungic *paN-* vs. *pa-*, Rejang *peng-* vs. *pe-*). Among the most common meanings associated with the SBI prefixes are: voice (e.g., Malayic *meN-* and *di-*, Karo Batak [bata1293, btx] *N-* and *i-*); causative (e.g., Acehnese *peu-*, Nias *fa-*); non-volitional (e.g., Lampungic *ti-*, Malayic *ter-*); intransitive (e.g., Lampungic *ba-*, Rejang *be-*); and nominalizing (e.g., Malayic *peN-*, Rejang *peN-*).

Suffixes, albeit a smaller set, are similarly widely employed in SBI languages with a variety of meanings. The most prominent and widespread suffixes are nominalizers and applicatives. One or two nominalizing suffixes are typically encountered in the SBI languages, and where a distinction is made between two, one normally nominalizes patients and the other locations (e.g., Toba Batak [bata1289, bbc] *pangan* 'to eat', *pangan-on* 'food', *pangan-an* 'plate'). SBI languages, like many of the region (McDonnell & Truong 2024), have two applicative forms: one that marks beneficiaries and instruments among other roles and another that marks locations and goals among other roles (e.g., Lampungic *-ko* vs. *-i*). In Nias, many of these suffixes have a wide variety of allomorphs with varying initial consonants, apparently tracing back to the root-final consonants that have since been lost (Brown 2001: 337–338). Kerinci, Acehnese, and Rejang are outliers in the region for wholly lacking suffixes.

Infixes are only productively employed by Acehnese, the Sumatran languages, and especially Rejang (see Section 6) and are few in number, all being derived from PMP *<um> (e.g., Acehnese <*eum*>, Rejang <*əm*>, Simeulue <*um*>) and PMP *<in> (e.g., Acehnese <*eun*>, Rejang <*ən*>, Toba Batak <*in*>). Derivatives of the former typically denote an imperfective meaning, mark certain intransitive verbs or an A-oriented voice, whereas the latter are used to derive nominals from verbs or mark a P-oriented voice.

Affixes can frequently be used in conjunction with one another with relatively transparent semantics, although some combinations might be properly called circumfixes in that, while their form consists of productive affixes, their semantics are not discernable from the constituent forms. This can be seen, for example, in the Lampungic reciprocal *sa- -an* (Walker 1976: 30) and the Karo Batak intensive comparative *ter- -en* (Woollams 1996: 55).

Enggano again stands out against other languages of the region. While many of its affixes are typical of the SBI languages (e.g., a prefix that triggers nasal substitution) and traceable to

PMP, many have innovative semantics. For example, while the PMP patient nominalizer *-*ən* is inherited with its nominalizing function, it also becomes a marker of obligation as Enggano -*Co*. Other affixes have been innovated entirely, such as the instrumental *V*- which takes the form of a prefixal reduplication of the vowel of the initial syllable (Edwards 2015: 73; Crowley n.d.: 12–13).

Finally, the alternation in final rimes in Kerinci is in some cases a derivational process wherein the oblique form corresponds to words that contain applicative, nominalizing, or possessive suffixes in other varieties of Malay, as in the examples in Table 4. In other cases, the alternation is based on the word's phrasal position (see below).

Table 4. Derivational processes in syllable rime alternation in TPM Kerinci (McKinnon 2011).

	Absolute	**Indonesian**	**Gloss**	**Oblique**	**Indonesian**	**Gloss**
Nominalizer	*gɨ*	*pegang*	'to hold'	*gʌŋ*	*pegang-an*	'a handle'
Possessive	*umah*	*rumah*	'house'	*um͡əoh*	*rumah=nya*	'his/her/the house'
Applicative	*mand͡ae*	*mandi*	'bathe'	*mandɪy*	*mandi-kan*	'bathe s.o.'

4.3 Syntax

The syntax of SBI languages is diverse. One unifying characteristic is that they are largely head-initial. In terms of clausal word order, SBI languages are described as verb-initial or verb-medial. However, such descriptions are complicated by the fact that the majority of the languages have a simplified *symmetrical voice* system in which there is a two-way distinction between A-oriented and P-oriented constructions. Because these constructions show little difference in terms of morphosyntactic complexity and syntactic transitivity, neither can be considered the basic transitive construction. The former, referred to as A-Voice (AV), privileges A, while the latter, referred to as P-Voice (PV), privileges P. These privileged arguments pattern with S and are typically referred to as the *subject*, while the other core argument in a transitive construction is referred to as the *non-subject argument* (i.e., P in AV clauses, A in PV clauses; see Arka 2003; Riesberg 2014; Chen & McDonnell 2019). The non-subject argument usually forms a constituent with the verb, referred to here as the *predicate complex*.

Lampung Api represents a fairly typical symmetrical voice system, as in (1). In AV, the verb is prefixed with the homorganic nasal, while in PV, it is prefixed with *di*-. The subject precedes the predicate complex, while the non-subject argument follows the verb.

(1) Lampung Api, Symmetrical voice

a. *Dukun* *hina* *nga-batatuk* *ubat*
healer that AV-bring medicine
'The healer brought medicine.'

b. *Jexuan* *hinji* *di-suah* *tian*
firewood this PV-burn 3PL
'The firewood is burned by them.'
(Walker 1976: 20)

In a number of languages of Sumatra, PV constructions may occur without the prefix. In such a construction in Lampung Api, the non-subject argument is realized as a proclitic on the verb (2). Such a construction has been referred to by a number of names, most recently Object Voice (OV).

(2) Lampung Api, Object Voice
Jexuan *hinji* *tian=suah*
firewood this 3PL=[PV]burn
'The firewood is burned by them.'
(Walker 1976: 20)

In several languages, there is an asymmetry in whether the non-subject argument is realized as a proclitic or enclitic. In Gayo and Toba Batak, first person non-subject arguments in PV are proclitic while second and third person are enclitic (Eades 2005: 65; van der Tuuk [1867] 1971: 155-156). In Besemah, as in many other Malay varieties, first and second person are proclitics, while third person is enclitic. In a number of languages, there is an additional passive construction that utilizes the same prefix as PV constructions. However, in the passive, A is either unrealized or appears in an apparent by-phrase. Teasing apart the differences between PV and passive, however, is not straightforward (see Kroeger 2014; Kroeger & Riesberg 2024; McDonnell to appear).

The languages of Sumatra show many of the characteristics of symmetrical voice, although the nature of voice in Acehnese has been the subject of disagreement for decades. Contra earlier analyses by Durie (1985, 1987) that argue that Acehnese argument structure is based on semantic roles and *not* grammatical relations, Legate (2012, 2014) has argued for an analysis similar to one presented above for Lampung Api, although she refers to PV as a (non-canonical) passive (see also Legate 2021). The symmetrical voice systems in other languages of Sumatra have undergone different degrees of erosion. In a number of SBI languages, especially the Malay varieties of the eastern lowlands, the voice system is breaking down: PV constructions must occur with a PV prefix (i.e., the loss of PV with a proclitic), and AV prefixes are optionally realized in AV constructions (Cole, Hermon, & Yanti 2008). It is worth noting that a number of studies have described languages of Sumatra, especially the Malay varieties, as possessing a system with a three-way distinction between OV, an active voice analogous to AV, and a passive voice analogous to PV constructions with a dedicated prefix (see, e.g., the sketch of Jambi Malay in Section 5).

Voice in the Barrier Islands languages is less clear; Simeulue, Sigulai and Mentawai appear to possess some of the characteristics of symmetrical voice (see, e.g., Mentawai in Section 7). In Nias, symmetrical voice has eroded to the point that it is only present in dependent clauses (Brown 2001), while Enggano shows no evidence of symmetrical voice.

Word order in symmetrical voice languages is typically variable between subject-initial, as in the Lampung Api examples above, and subject-final where the subject follows the predicate complex. In Besemah, both orders are possible. In some cases, there is an asymmetry in word order based on voice, such that PV may restrict or even prohibit subject-initial word order (see examples from Rejang in Section 6), while AV may restrict or prohibit subject-final word order (e.g., Gayo; Eades 2005: 174). In Nias, the predicate complex typically occurs before the subject, and while Enggano is not a symmetrical voice language, word order alternates based on the presence of a verbal prefix: VS/VAP is the typical word order with *bu-* marked verbs whereas SV/AVP is the typical word order with *ki-* marked verbs (Edwards 2015: 61).[7]

Case marking and obligatory argument indexing are found in many Austronesian languages (e.g., Philippine-type languages; see McDonnell & Chen 2022) but are limited in SBI languages. Enggano has limited case-marking and indexing (see Edwards 2015: 60–62), but Nias

[7] Except when the predicate is stative, in which case the order is VS (Crowley n.d.: 32).

is an exemplary case for both. Consider the examples in (3). As mentioned in the previous section, consonant mutation serves as a type of case marking and typically occurs with S (3a) and P arguments (3b) as well as arguments of prepositions (3c). A arguments are unmarked, as in (3b). Person/number agreement is robust and obligatorily marked on the verb for A in realis mood (3b–c) and S (3d) and A (3e) in irrealis mood.

(3) Nias, case marking and argument indexing

a. *a-böu* *dana-gu.*
STAT-smell.bad hand\MUT-1SG
'my hand smells bad.'
(Brown 2001: 119)

b. *la-a* *gö-ra* *samösana*
3PL.REAL-eat food\MUT-3PL.POSS each
'they each ate their food.'
(Brown 2001: 138)

c. *i-be* *khö* *vo'omo-nia* *nukha* *s=oyo.*
3SG.REAL-give DAT spouse\MUT-3SG.POSS cloth REL=red
'she gave her husband the red cloth.'
(Brown 2001: 349)

d. *gu-möi* *ba* *fasa* *mahemolu*
1SG.IRR loc market tomorrow
'I want to go to the market tomorrow.'
(Brown 2011: 153)

e. *ya-mbu'a* *gömö-nia*
3SG.IRR-repay:IRR debt\MUT-3SG.POSS
'he might repay his debt.'
(Brown 2011: 523)

Enggano has two sets of person/number agreement markers, the first primarily limited to subordinate and negation clauses and the second found commonly in discourse attached to verbs affixed with *bu-*. In other SBI languages, pronominal clitics commonly attach to verbal predicates and may be coreferential with an argument NP. According to Durie (1987), Acehnese has compulsory coreference of Actor arguments on both intransitive and transitive verbs whereas Undergoers are optionally indexed.[8] In a number of Malay varieties, such as Besemah, pronominal A arguments in PV are coreferential with a prepositional phrase. In Mentawai, first- and second-person plural A arguments in PV must be indexed with the first and second person proclitics on the verb (see Section 7.2). This apparently occurs in Toba Batak as well (see van der Tuuk 1971: 156).

Finally, the final rime alternation in Kerinci may also be determined by its position in the phrase: the absolute form occurs in phrase-final position and the oblique form occurs in phrase-medial position. Consider the NPs in (4).

(4) TPM Kerinci, Positional final rime alternation

a. *batɨ*
tree\ABSL
'tree'

b. *batʌŋ* *pina*
tree\OBL areca.nut\ABSL

[8] See Legate (2012, 2014) for an alternative perspective.

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

Intransitive verbs, even when followed by an adjunct, are marked absolute (5a). Transitive AV marked verbs with a P argument are marked oblique (5b), but when they are used intransitively with an unspecified P, they are marked absolute (5c).

(5) TPM Kerinci, Syntactic final rime alternation

a. *ɲo* *tidə͡o* *lamo*
3 sleep\ABSL long\ABSL
‘He slept for a long time.’
(McKinnon, Cole & Hermon 2011: 724)

b. *ka* *ndə͡oʔ* *makən* *nasa͡e* *minin* *lah*
1SG FUT eat\OBL rice\ABSL now just
‘I am going to eat rice right now.’
(McKinnon, Cole & Hermon 2011: 721)

c. *ka* *ndə͡oʔ* *maka* *minin* *lah*
1SG FUT eat\ABSL no just
‘I am going to eat right now.’ (meaning ‘eat something unspecified’)
(McKinnon, Cole & Hermon 2011: 724)

The development of this system arises from several processes, and McKinnon et al. (2015) and Yanti el al. (2018) show that these patterns are not only found in Kerinci but also other Malay varieties in the region.

5 Sketch of Jambi Malay

Jambi Malay (glottocode: jamb1236, ISO 639-3: jax) comprises several Malay dialects spoken in Jambi province, Sumatra. It is spoken by around 1.6 million people (Eberhard et al. 2025). This sketch describes the Tanjung Raden variety of Jambi Malay, spoken in Tanjung Raden village, which is located across from the center of Jambi City, separated by the Batanghari River. Jambi people refer to the Malay variety spoken in Tanjung Raden and its neighboring villages as *Baso Seberang* ‘the language from across.’ It is based on Yanti (2010) and Yanti, Cole, & Hermon (to appear).

Data in this section are marked with *Elicited* if they were collected through elicitation. Naturalistic data are marked with the recording session name and the utterance ID. The collection of Tanjung Raden data is archived at The Language Archive (Yanti et al. 2002-2008).

5.1 Phonology

Jambi Malay exhibits 24 consonant phonemes, as shown in Table 5. The phoneme /z/, shown in brackets in Table 5, is only found in loanwords from Arabic. In loanwords that contain /f/, /p/ is used instead.

Table 5. Consonant phonemes of Jambi Malay.

		Labial	**Dental~alveolar**	**Palatal**	**Velar**	**Glottal**
Stops	**Voiceless**	p	t	c	k	ʔ
	Voiced	b	d	ɟ <j>	g <g>	

Fricatives	**Voiceless**		s			h
	Voiced		(z)			
Nasals	**Simple**	m	n	ɲ	ŋ	
	Post-occluded	m^b	n^d	$ɲ^ʝ$ <$ɲ^j$>	$ŋ^g$ <$ŋ^g$>	
Liquids			l		ʁ <r>	
Glides		w		j <y>		

The Tanjung Raden variety of Jambi Malay exhibits a series of post-occluded nasals /m^b, n^d, $ɲ^j$, $ŋ^g$/. They primarily occur intervocalically, as in *sam^{b}al* [sam^{b}al] 'chili sauce', *man^{d}i* [mãn^{d}i] 'bathe', *iɲ^{j}aʔ* [iɲ^{ʝ}aʔ] 'step on', and *miŋ^{g}u* [mĩŋ^{g}u] 'week or Sunday', but they are also found word initially, as in *m^{b}oʔ* [m^{b}oʔ] 'older sister', *n^{d}aʔ* [n^{d}aʔ] 'want', *ɲ^{j}ʊʔ* [ɲ^{j}ʊʔ] 'give, and *ŋ^{g}ɪ* [ŋ^{g}ɪ] 'yes (honorific)'. Palatals, voiced obstruents, glides, post occluded nasals, and the glottal fricative do not occur in word-final position. The voiceless velar stop /k/ and glottal stop /ʔ/ are contrastive word finally, *ketek* [ke.te^{ŋ}k] or [ke.tekˀ] 'boat' vs. *keteʔ* [ke.teʔ] 'armpit'.

Voiceless oral stops /p/, /t/, and /k/ in root-final positions are realized as prenasalized oral stops [^{m}p], [^{n}t], and [^{ŋ}k], respectively, and these prenasalized stops are in free variation with the unreleased stops [pˀ], [tˀ], and [kˀ]. Simple nasals [m], [n], and [ŋ] are in free variation with pre-occluded nasals [^{b}m], [^{d}n], and [^{g}ŋ] phrase-finally.

There are seven phonemic monophthong vowels in Jambi Malay, as shown in Table 6.

Table 6. Vowels phonemes of Jambi Malay.

	Front	**Central**	**Back**
High	i		u
Mid-high	ɪ		ʊ
Mid	e		o
Low		a	

Phonemic mid-high vowels /ɪ/ and /ʊ/ are only found in final syllables and never in penultimate syllables. Jambi Malay also exhibits sequences of a vowel and a glide, which are analyzed as tautosyllabic vowels, and the glide occupies the coda position. Tautosyllabic vowels are primarily found in loanwords and interjections and, thus, are rare. Vowel-glide sequences which are historically word-final are realized as monophthong-final in synchronic forms (e.g., *sampay* > *sampe* 'arrive', *suŋay* > *suŋe* 'river', *rantay* > *rante* 'chain').

Schwa is not distinctive, and its occurrences can be explained as the result of either vowel reduction (e.g., [səpatu] from /sapatu/ 'shoe') or vowel insertion (e.g., [təloʔ] from /tloʔ/ 'egg'). Vowel harmony is one characteristic of the Jambi Malay spoken in most areas in Tanjung Raden, Mudung Darat, and Tanjung Pasir villages. Vowel harmony occurs when a penultimate vowel agrees in vowel height with the final vowel within a morpheme. It applies to /i/ and /ɪ/ and /u/ and /ʊ/ (e.g., [bulu] 'body hair', [gigi] 'tooth', [pɪkɪʁ] 'think', [ɪsʊʔ] 'tomorrow'). Vowels are nasalized when following a simple nasal (e.g., /teŋoʔ/ [teŋõʔ] 'look', /lamo/ [lamõ] 'long time') but not a post-occluded nasal (e.g., /m^{b}oʔ/ [m^{b}oʔ] 'older sister', /mãn^{d}i [mãn^{d}i] 'bathe').

The basic syllable shapes of Jambi Malay are V, CV, VC, and CVC. Other types of syllables are C (syllabic), CCV, and CCVC. Syllabic consonants are restricted to nasals. Two or

more consecutive consonants do not follow V. Table 7 displays examples of different syllable types in Jambi Malay.

Table 7. Syllable structures of Jambi Malay.

Syllable type	Example
V	**u**.mõ 'paddy'
CV	**gʊ**.gʊʔ 'area'
VC	**un**.tal 'roll'
CVC	**jan**.tan 'male'
C (syllabic)	**m̩**.pa^{n}t 'four'
CCV	**sʁo**.al 'pants'
CCVC	**bʁas** 'uncooked rice'

Words are mostly disyllabic (e.g., *ma.saʔ* 'cook', *a.ɲʊt* 'washed away'), but single-syllable words are also found (e.g., *lang* 'eagle', *m^{b}oʔ* 'older sister').

5.2 Morphology

Word formation processes in Jambi Malay include affixation, cliticization, reduplication, truncation, and compounding. The first three of these processes are discussed in this section. The primary affixes and their main functions are presented below. Borrowed affixes have been excluded.

- *ŋ-*: forms active voice verbs (alternatively A-Voice, see Section 4.3), as in *ŋ-am^{b}ɪʔ* 'ACT-to.take' from *ambɪʔ* 'to take' and *m-aco* 'ACT-to.read' from *baco* 'to read'
- *di-*: forms passive voice verbs (alternatively P-Voice, see Section 4.3), as in *di-ajaʔ* 'PASS-to.invite' from *ajaʔ* 'to invite' and *di-jmpʊt* 'PASS-to.pick.up' from *jmpʊt* 'to pick up'
- *ba-*: forms intransitive verbs, as in *ba-tam^{b}a* 'to increase' from *tam^{b}a* 'to add' and *ba-bua* 'to bear fruit' from *bua* 'fruit'
- *paŋ-*: forms nouns indicating a tool/instrument used to perform an action or a person who performs an action, as in *pa-makan* 'eater' from *makan* 'to eat' and *pa-rkam* 'recorder' from *rkam* 'to record'
- *sa-*: forms comparative adjectives, as in *sa-bsaʔ* 'as big as' from *bsaʔ* 'big'
- *ka-*: forms ordinal numbers, as in *ka-limo* 'fifth' from *limo* 'five'
- *-an*: mainly forms nouns, as in *gantʊŋ-an* 'hanger' from *gantʊŋ* 'to hang' and *kotor-an* 'dirt' from *kotor* 'dirty'
- *-kan*: forms causative and applicative constructions, as in *pca-kan* 'to break' from *pca* 'broken' and *naeʔ-kan* 'to raise' from *naeʔ* 'to go up'
- *-i*: forms applicatives, as in *garam-i* 'to add salt into' from *garam* 'salt' and *masaŋ-i* 'to set something for someone' from *masaŋ* 'to set'
- *paŋ-…-an*: forms abstract nouns, as in *pa-watas-an* 'border' from *watas* 'limit'
- *sa-…-e/sa-…-ɲo*: forms adverbs, as in *sa-bnar-e* 'actually' from *bnar* 'right' and *sa-kuat~kuat-ɲo* 'as strong as possible' from *kuat* 'strong'

5.2.1 Personal pronouns

Jambi Malay has both free and clitic pronouns (Table 8). The range of independent first-person pronouns cover a variety of different social contexts, with *aku* '1SG' being primarily used among peers or in conversation with those of lower social status, *sayo* '1SG.POL' with those of higher social status, and *kulo* '1SG.POL' used in the higher speech level of Tandjung Raden and surrounding villages but not in other varieties of Jambi Malay. The first-person plural exclusive pronoun *kami* '1PL.EXCL' can also be used as a polite first-person singular pronoun. Similarly, the first-, second-, and both third-person pronouns *kito* '1', *kamu* '2', *dioʔ* '3', and *ɲo* '3', respectively, can all be used for both singular and plural referents, though with no difference in formality. A final pronoun *awaʔ* (not shown in Table 8) is a generalized pronoun used for referring to any combination of number and person except third-person plural where the reduplicated form *awaʔ-awaʔ* '3PL' is used instead.

Clitics in Jambi Malay include first and second-person proclitics *ku=* and *kau=* (e.g., *ku=tanam* 'I plant'), first person enclitic *=ku* (e.g., *mato=ku* 'my eyes'), and third person enclitic pronouns *=e* and *=ɲo* (a newer form). The first and second person proclitics *ku=* and *kau=* are both used to mark the A arguments in Object Voice (see Section 5.3.3), whereas the first-person enclitic *=ku* marks nominal possession (see Section 5.3.7). The multifunctional third person enclitics *=e* and *=ɲo* fulfill both of these roles, marking an A argument in passive voice (e.g., *di-tiŋᵍal=e* 'left behind by him/her') as well as possession (e.g., *maʔ=e/=ɲo* 'his/her mother'). Other functions include nominalization (e.g., *balɪʔ=e/=ɲo* 'the return'), marking for definiteness (e.g., *batu=e/=ɲo* 'the stone'), and indicating direct objects in active voice (e.g., *mᵇuat=e/=ɲo* 'to make it').

Table 8. Jambi Malay pronominal forms.

	Singular	**Plural**	
		EXCL	**INCL**
1	*aku* *sayo* *kami* *kulo** *kito* *ku=* *=ku*	*kito* *kami*	*kito*
2	*kau* *kau=* *kamu*	*kamu*	
3	*dio(ʔ)* *ɲo* *=ɲo* *=e*	*dio(ʔ)* *ɲo* *=ɲo* *=e*	

5.2.2 Reduplication

Nouns, pronouns, verbs, adjectives, and numerals can be reduplicated. There are three main reduplication types: full reduplication (e.g., *pukat~pukat* 'dragnets'), partial reduplication (e.g., *ka~kawan* 'friends'), and imitative reduplication (e.g., *bolaʔ~balɪʔ* 'back and forth'). Full and partial reduplication can be further distinguished into full reduplication with no final

consonant (e.g., *sopa~sopan* 'very polite', full reduplication with a glottal stop (e.g., *dikiʔ~dikit* 'little by litte', CV reduplication (e.g., *ki~kiro* 'average'), CV*ʔ* reduplication (e.g., *buʔ~buat* 'to make something up'), C*a* reduplication (e.g., *pa~potoŋ* 'cut repeatedly'), C*ə* reduplication (e.g., *sə~sopan* '(more) polite'), C*aʔ* reduplication (e.g., paʔ~pijat 'to give massage repeatedly'), and C*əʔ* reduplication (e.g., *kəʔ~kilat* 'very shiny').

5.2.3 Morphophonology

The nasal-substitution active voice prefix *ŋ-* appears as different forms depending on the first segment of the base it attaches to. When the prefix attaches to vowel-initial bases, it appears as *ŋ-* (e.g., *ŋ-* + /ɪkʊt/ → [ŋ̃ĩkʊⁿt] 'to follow'). When the base begins with a voiceless stop (/p/, /t/, /c/, /k/), the prefix is realized as a nasal with the same place of articulation that *replaces* the voiceless stop (e.g., *ŋ-* + /tanam/ → [nãnãm]). When it attaches to a base that begins with a voiced stop (/b/, /d/, /j/, /g/), the nasal prefix and the voiced stop form post-occluded nasals (e.g., *ŋ-* + /buaʔ/ → [mᵇuaⁿt] 'to make'). In bases that begin with a nasal, the nasal prefix is not realized (e.g., *ŋ-* + /ɲaɲi/ → [ɲãɲĩ] 'to sing'). When it attaches to a glide or a liquid, the nasal prefix surfaces as [m̩] or [mə] (e.g., *ŋ-* + loloi → [m̩loloi] 'to trick' and *ŋ-* + yakın → [mẽyakınkan]). Finally, when the prefix attaches to bases that begin with /s/, it surfaces as a palatal nasal (e.g., *ŋ-* + /sbʊt/ → [ɲ̃ãbʊⁿt]).

Similar nasal-subsitutation patterns occur with the prefix *paŋ-*, which again appears in different forms depending on the first segment of the base it attaches to.[9] When it attaches to vowel-initial base, it appears as *paŋ-* (e.g., *paŋ-* + /iŋat/ → 'a reminder; someone who remembers well', *paŋ-* + /kbat/ → [paŋbat] 'something for tying'). When *paŋ-* attaches to a base that begins with a voiceless obstruent (/p/, /t/, /c/, /s/, /k/) the last segment of the prefix is realized as a nasal with the same place of articulation (e.g., *paŋ-* + /kbat/ → [paŋbat]/ 'something for tying'). When it attaches to a base with a voiced-initial stop (/b/, /d/, /j/ and /g/), the final segment of *paŋ-* is realized as post-occluded nasal (e.g., *paŋ-* + /bʊŋkʊs/ → [pamᵇʊŋkʊs] 'wrapper'). When it attaches to a base that begins with a liquid, a glide or nasal, the prefix appears as *pa-* (e.g., *paŋ-* + /makan/ → [pamãkãn]/ 'eater').

5.3 Syntax

Jambi Malay is head-initial with prepositions preceding the noun phrase, adjectives follow nouns, and aspect markers, modal auxiliaries, and negators precede the predicate. Verbs in Jambi Malay are often marked by a verbal prefix, such as *ŋ-* 'ACT', *di-* 'PASS', and *ba-* 'INTR'. As a head-initial language, if an argument is present, the verb in a verb phrase (analogous to the predicate complex, see Section 4.3) precedes the complement. The canonical word order for intransitive clauses is SV (6), while VS (7) is also allowed.

(6) Jambi Malay, Intransitive clause-SV

yayaŋ	*naŋis*
Yayang	ACT:cry

'Yayang is crying.'

[9] The same patterns are true of the prefixal element of the abstract nominalizing circumfix *paŋ- -an* 'NMLZ'.

(TJR-180106_a;
482603111031170406)

(7) Jambi Malay, Intransitive clause-VS

balɪʔ	*la*	*budaʔ*	*ko*	*tadi*
return	EMPH	kid	DEM.PRO X	earlier

'They went home then.'
(TJR-050805_a; 217400153710201105)

The unmarked word order for transitive clauses is AVP (8), although the word order is extremely flexible. For example, VPA (9), PAV (10), and VAP (11) orders are also found. Note that the active prefix *ŋ-* 'ACT' is optionally realized (see below).

(8) Jambi Malay, Transitive clause-AVP

awaʔ	*mgaŋ*	*tip*	*di*	*pon^{d}oʔ*
1/2/3	ACT:hold	tape	LOC	hut

'We use a tape in the hut.'
[Lit. 'We hold a tape in the hut.']
(TJR-100805; 925905134053300106)

(9) Jambi Malay, Transitive VPA

minum	*te*	*kito*	*malam*	*ko*
ACT:drink	tea	1	night	DEM.PROX

'We drink tea tonight.'
(TJR-121105;
160395101747020206)

(10) Jambi Malay, Transitive PAV

bulɔ	*tu*	*maʔ*	*aku*	*lagi*	*krat*
bamboo	DEM.DIST	mother	1SG	PROG	ACT.cut

'My mother has cut the bamboo.'
(Elicited)

(11) Jambi Malay, Transitive VAP

iyo,	*ɪkʊt*	*aku*	*pestipal*
yes	ACT.follow	1SG	festival

'Then, I joined in the festival.'
(TJR-050805_b; 418671192313211105)

In active ditransitive constructions, as in (12), the objects are unmarked and the recipient precedes the theme.

(12) Jambi Malay, Ditransitive AVRT

kito	*biaso*	*di*	*jam^{b}an*	*m^{b}agi*	*antu*	*aeʔ*	*makan*
1	usual	LOC	toilet	ACT:give	ghost	water	eat

'We used to give the water ghost food on the toilet.'
(TJR-190106_a; 711683100356270506)

5.3.1 Voice

Jambi Malay distinguishes three voices: active voice (13), passive voice (14), and object voice (15). As discussed in Section 4.3, these voices are analogous to AV, the prefixal PV, and the bare PV, respectively. The active voice is optionally marked by the nasal prefix *ŋ-* 'ACT' and its phonological variants. In an active transitive clause, A appears as the subject and P surfaces as the direct object of the verb. The passive voice is marked by the prefix *di-,* the patient/theme is realized as the subject, whereas A is optionally realized as an adjunct and optionally marked by the preposition *buat/samo* 'by'. In OV, the verb does not take a prefix, P is realized as the subject, and the presence of A adjacent to the verb is considered obligatory.

(13) Jambi Malay, Active voice

budi	*ɲ-jolaʔ*	*siti*	*pagi*	*tadi*
Budi	ACT-push	Siti	morning	earlier

'Budi pushed Siti this morning.'
(Elicited)

(14) Jambi Malay, Passive voice

siti	*di-jolaʔ*	*((buat)*	*budi)*	*pagi*	*tadi*
siti	PASS-push	by	Budi	morning	earlier

'Siti was pushed by Budi this morning.'
(Elicited)

(15) Jambi Malay, Object Voice

budi	*daʔ*	*siti*	*jolaʔ*
Budi	NEG	Siti	push

'Budi wasn't pushed by Siti.'
(Elicited)

5.3.2 Aspect markers, modal auxiliaries, and negators

Jambi Malay does not exhibit tense morphology. Aspect markers include *lagi/sdaŋ* 'PROG', *masɪ* 'still', *la* 'PFV', *baru* 'just', *prna* 'ever', *sampʊn* 'PFV (honorific)', *blʊm/lʊm* 'not yet', and $n^{d}aʔ$ 'want'. Modal auxiliaries are *biso, tukaŋ*, *tlap* 'can', *məsti*, *harus* 'must', and *bole* 'may'. Negators in Jambi Malay are *kɲoʔ* and *bukan*, which can only negate nouns, and *daʔ*, which negates clauses.

Aspect markers, modal auxiliaries, and negators appear before the predicate in a clause. If an aspect marker or a modal auxiliary and a negator are present, the negator precedes the aspect marker or the modal auxiliary, as in (16) and (17).

(16) Jambi Malay, Negation and modal auxiliary

datʊʔ	*kau*	*daʔ*	*tlap*	*makan*	*baŋaʔ*	*daʔ*
grandfather	2SG	NEG	can	eat	a.lot	NEG

'Your grandfather cannot eat it much.'
(TJR-231205_b; 143922080901300306)

(17) Jambi Malay, Negation and aspect

kami	*daʔ*	*lagi*	*mancɪŋ*
1	NEG	PROG	ACT:fishing.rod

'We are not fishing.'
(Elicited)

5.3.3 Existential and possessive constructions

In addition to clauses discussed above, Jambi Malay also has existential (18) and possessive (19) constructions using *ado* 'exist'.

(18) Jambi Malay, Existential construction

ado	*kotaʔ*	*meʔ*	*ko*	*buat=e,*	*ha*
exist	box	like	DEM.PROX	make=3	EXCLAM

'There is a box, he made it like this.'
(TJR-270605_b; 164351094013171005)

(19) Jambi Malay, Possessive construction

dioʔ	*ado*	*prau*	*beŋen*	*tu*
3	exist	canoe	old.time	DEM.DIST

'He had a canoe in the past.'
(Elicited)

Possession can also be expressed using *puŋo* 'have, own', as in (20).

(20) Jambi Malay, Possessive construction

ay,	*ba-pɪkɪr*	*pulaʔ*	*awaʔ*	*puŋo*	*laki,*	*awaʔ*	*lagi*	*mudo*
EXCLAM	INTR-think	PART	1/2/3	possess	husband	1/2/3	PROG	young

'Well, I was thinking that I already had a husband and I was still young.'
(TJR-180106_b; 894003081611160506)

5.3.4 Non-verbal predicates

Depending on the phrase type of the predicate, non-verbal clauses can be further classified into three types: nominal clauses (21), adjectival clauses (22), and prepositional clauses (23).

(21) Jambi Malay, Nominal clause

ha,	*iko*	*mato=ŋo*
EXCLAM	DEM.PROX	eye=3

'Well, this is the hook.'
[Lit. 'Well, this is the eye.']
(TJR-051002; 917601112351311002)

(22) Jambi Malay, Adjectival clause

badan=e	*bsaʔ*	*bnar*	*la*
body=3	big	right	EMPH

'Her body is really big.'
(TJR-121105; 454181101740020206)

(23) Jambi Malay, Prepositional clause

bɪbɪt	*di*	*memaŋ*	*dari*	*blanᵈo*
seed	earlier	indeed	from	Netherlands

'The seeds were indeed from Netherlands.'
(TJR-010507; 994212141505120507)

5.3.5 Noun phrase structure

Nouns in Jambi Malay are not overtly marked. Modifiers in NPs, such as demonstratives (24a), adjectives (24b), attributive nouns (24c), and possessors (24d) follow the head noun.

(24) Jambi Malay, NPs

a. Demonstratives

umo	*ko*
paddy	DEM.PROX

'this paddy'

b. Adjectives

baju	*merah*
garment	red

'red clothes'

c. Attributive nouns

bataŋ	*kmaŋ*
tree	kind.of.fruit

'a *kemang* tree'

d. Possessors

motor	*aku*
motorcycle	1SG

'my motorcycle'

When a numeral and a classifier are present in a noun phrase, the unmarked order is Numeral-Classifier-Noun, as in (25), although it is possible for the numeral and classifier to follow the noun.

(25) Jambi Malay, Numeral-Classifier-Noun

tigo	*ekoʔ*	*burʊŋ*
three	CLF	bird

'three birds'

Likewise, relative clauses, headed by a relativizer *yaŋ* or *naŋ*, modify NPs. The relativizer *naŋ* is only used in Tanjung Raden and the surrounding villages, but *yaŋ* is used also used in Jambi City. Relative clauses always follow the head noun and may be headed, as in (26), or headless, as in (27).

(26) Jambi Malay, Relative clause

dataŋ	*tu*	*num^{b}or*	*jarɪŋ*	*yaŋ*	*kito*	*bntaŋ*	*tadi*	*ko*
come	DEM.DIST	ACT:hit	net	REL	1	spread.over	earlier	DEM.PROX

'[It] comes and hits the net that we set before.'
(TJR-100805; 223202134049300106)

(27) Jambi Malay, Headless relative clause

naŋ	*kasar*	*dapat*	*rgo*	*mahal*
REL	big	get	price	expensive

'The ones that are big are expensive.'
[Lit. 'The ones that are big get an expensive price.']
(TJR-170605; 238471171059090705)

6 Sketch of Rejang

Rejang (glottocode: reja1240, ISO 639-3: rej) is spoken by approximately 350,000 people in the highlands and along the coast of southwest Sumatra with five major dialects: Lebong, Kepahiang (Kebanangung), Pasisir, Musi, and Rawas (Eberhard et al. 2025). Of the five dialects, Rawas is the most divergent and may not be mutually intelligible with the other four (McGinn 2005). This sketch describes the Musi variety of Rejang (glottocode: musi1242, ISO 639-3: rej) as it is spoken in Kampung Sambé, Curup, Bengkulu province in southwest Sumatra. It is based entirely upon Richard McGinn's unpublished and partially completed grammar of Rejang. Data sources comprise examples from conversations, traditional narratives, a set list of over 400 elicited words and 200 sentences, and targeted elicitation of phonological and morphosyntactic phenomena.

6.1 Phonology

Rejang contrasts 21 consonant phonemes as shown in Table 9. The phonemes /ɾ/ and /h/ only occur in recent loanwords.

Table 9. Consonant inventory of Musi dialect.

		Labial	Dental~Alveolar	Palatal	Velar	Glottal
Stops	**Voiceless**	p	t		k	ʔ
	Voiced	b	d		g <g>	
Affricates	**Voiceless**			t͡ʃ <c>		
	Voiced			d͡ʒ <j>		
Fricatives			s			(h)
Nasals	**Plain**	m	n	ɲ	ŋ	
	Barred	m^b <m̄>	n^d <n̄>	$ɲ^{d͡ʒ}$ <ɲ̄>	$ŋ^g$ <ŋ̄>	
Liquids			l			
Taps			(ɾ) <r>			
Glides		w		j <y>		

The most striking feature of the Musi phoneme inventory is the series of postploded nasals /m^b, n^d, $ɲ^{d͡ʒ}$, $ŋ^g$/, which have also been referred to as *barred* or *funny* nasals (Coady & McGinn 1982). Post-occluded nasals are limited to intervocalic positions, which is a result of their historical development from nasal-stop clusters which only occur intervocalically. One salient difference between post-occluded and plain nasals is the presence of nasality on the following vowel: the vowel following a plain nasal is subject to progressive nasal spreading (e.g., *jaməw* [ja.mə̃w̃] 'banquet'), while the vowel following a post-occluded nasal is not (e.g., *jam̄əw* [ja.m^bəw] 'guava'). The phonemes /t, n, l, ɾ/ are produced with a dental place of articulation [t̪, n̪, l̪, ɾ̪], while /d, n^d/ are produced with alveolar place of articulation [d, n^d].

There are a number of restrictions on consonants occurring in root/word-final position: (i) palatal consonants may not occur in this position, and (ii) only voiceless stops may occur in this position. Furthermore, /h/ does not occur in word-final position, even in recent loanwords, and /l/ only occasionally occurs word-finally in a handful of loanwords. While /k/ and /ʔ/ contrast in word-final position, there are very few roots that actually end in /k/. Finally, there is allophonic

variation in plain nasals word-finally, which may be realized with preplosion (e.g., *tun* [tun]~[tu^{d}n] ‘person’). Word-medial codas are restricted to /l/ in native vocabulary (e.g. *pəlpəaʔ* ‘palm frond’) with other possibilities in loanwords (e.g., *sərgap* ‘attack’, *waktəw* ‘time’).

Musi has six monophthong and seven diphthong vowels, shown in Tables 9 and 10, respectively. Prepenultimate syllables are limited to schwa in both roots and morphologically derived words (e.g., *bənataŋ* ‘animal’, *t<ən>imaʔ* ‘shoot’, *mə-liləy* ‘run away’). All six monophthongs occur in all other positions, although diphthongs are much more common word-finally. Word final monophthongs are limited to function words, loanwords, and a few instances of words with an unknown origin. A non-phonemic glottal stop occurs in vowel initial words (e.g., /anaʔ/ → [ʔanaʔ] ‘child’). There is an invariant vowel harmony in roots; when the penultimate syllable is a peripheral mid-vowel /e, o/, the final syllable is identical (e.g., *leŋet* ‘sky’, *monoʔ* ‘chicken’).

Diphthongs have a more limited distribution, only occurring in the ultima. With the exception of mid-low vowel diphthongs that may be followed by a glottal stop, all other diphthongs cannot be followed by a coda. In addition to these diphthongs, there are also vowel-glide sequences /aj, aw, əj, əw/ that occur in word-final position. Finally, high vowels optionally undergo vowel breaking before word-final /t, s/ (e.g., *sakit* [sa.ˈkit] ~ [sa.ˈkij̯ət] ‘sick’, *lut* [lut] ~ [luw̯ət] ‘very’).

Table 10. Rejang vowel inventory.

	Front	**Central**	**Back**
High	i		u
Mid	e	ə	o
Low		a	

Table 11. Rejang diphthong inventory.

	Front	**Central**	**Back**
High-Mid	ie̯ (~əe̯)		uo̯ (~əo̯)
Mid-Mid	eo̯		oe̯
Mid-Low	ea̯	əa̯	oa̯

Syllables are minimally a nucleus (V), but typically have an onset (CV), coda (VC), or both (CVC). Words may comprise a single (open or closed) syllable (e.g., *bo* ‘land cleared’, *waʔ* ‘aunt’), although most words are disyllabic (e.g., *ti.up* ‘blow’, *o.toʔ* ‘brain’). Stress falls on the final syllable.

6.2 Morphology

Rejang has only prefixes and infixes. Some of the prefixal forms are polyfunctional (e.g. *kə-*), and a few of them have both nominal and verbal functions. At the same time, some functions are expressed by more than one prefixal form (e.g., *de-*, *se-* with numeral, simulative, and comitative meanings).

- With few exceptions, transitive verbs are infixed with the A-Voice (AV) <*əm*> or P-Voice (PV) <*ən*> infix or their prefixal allomorphs (e.g., *t<əm>imaʔ*, *t<ən>imaʔ* 'shoot'). These infixes may also derive transitive verbs from intransitive bases (e.g., *təko* 'come' → *t<əm>ko*, *t<ən>ko* 'import').
- The primary intransitive verbal prefixes are *mə(N)-* and *bə-*:
    - *mə(N)-* derives intransitive verbs from various bases, many of which take on a meaning of locomotion (e.g., *mə-liləy* 'run, run away', cf. *liləy* 'run'; *mən-əbo* 'elope' lit. 'go to the mountain', cf. *təbo* 'mountain').
    - *bə-* may attach to nominal bases and express possession 'have NOUN' (e.g., *besawəaʔ* 'have/own wet-rice fields', cf. *sawəaʔ* 'rice field') or verbal bases and express locomotion meanings, which contrast with *me(N)-* (e.g., *bə-panuo* 'have/take a walk', cf. *panuo* 'walk'; *b-eleʔ* 'go home', cf. *-eleʔ* 'return s.t.').
- The prefix *tə-* marks non-volitionality, including states (*tə-kəjir* 'surprised'), psychological states (e.g., *t-iŋət* 'remember'), and inadvertent actions (e.g. *tə-pəŋ̄aŋ* 'cry out (in pain)').
- The prefixal form *kə-* has at least four different functions, which appear to be quite disparate and may, in fact, represent four homophonous prefixes. Two of these have verbal functions with the other two having non-verbal functions:
    - *kə-* marks imperatives as a transitive verbal prefix (e.g., *k-əmūʔ ba!* 'eat!'), and when attached to other stative verbs it has an intensive function (e.g., *kə-oaʔ* 'very far').
    - *kə-* marks an adversative meaning in a small set of verbs (e.g., *k-ujən* 'get rained on'), similar to the adversative circumfix (e.g. Malay *ke- -an*) found in many other Malayo-Polynesian languages of the region.
    - *kə-* occurs on ordinal numbers (e.g., *duəy* 'two' → *kə-duəy* 'second') .
    - *kə-* forms abstract nouns from various bases (e.g., *namən* 'know' → *kə-namən* 'knowledge', *ləy* 'big' → *kə-ləy* 'size, bigness').
- The prefixes *sə-* and *də-* appear to have the same range of functions. The prior is likely a borrowing from the Malay numeral prefix *se-* 'one', while the latter is likely a reduced form of the Rejang numeral *do* 'one'. As a numeral, these prefixes typically attach to numeral classifiers (e.g., *də-tudaʔ pisaŋ* 'a bunch of bananas'). They also express a timespan with the meaning 'a complete unit of time' (e.g., *də-maləm* 'all night', cf. *maləm* 'night'; *də-biləy* 'all day', cf. *biləy* 'day') and form predicates with the meaning 'same as' or 'together with'.
- The agent nominalizer *pə-* forms nouns that typically evoke notions of power and authority (e.g., *tuluŋ* 'help' → *pə-tuluŋ* 'delegated (top) assistant'), but there are a number of cases that do not involve an agent (e.g., *toaʔ* 'bribe' → *pə-toaʔ* 'a bribe').
- The instrumental nominalizer *pə(N)-* forms nouns that designate the utility of the root (e.g., *peʔeaʔ* 'sour' → *pəm-eʔeaʔ* 'spice for souring') but may form nouns that indicate habitual or characteristic actions (e.g., *ujən* 'rain' → *musim pəŋ-ujən* 'rainy season'; *səbit* 'complain' → *tun pəŋ-əbit* 'complainer; person who complains').

6.2.1 Person Pronouns

Person pronouns come in both free and dependent (clitic) forms. They distinguish singular and plural in first and second person, but, with third person, the independent *si* and clitic

=nə forms have no such distinction. There is an apparent third person plural form *tobo o* (lit. 'that group'). First person plural pronouns also demonstrate an inclusivity distinction. Pronominal enclitic forms occur as the possessor of a noun (phrase) or A in a PV construction in first and second person singular forms as well as the third person singular and plural. They may occur as the object of a preposition but independent pronouns are the preferred.

Table 12. Rejang pronominal forms.

	Independent		**PV A, Possessor**	
	Singular	**Plural**	**Singular**	**Plural**
1.EXCL	*uku*	*kami*	*=ku*	*kami*
1.INCL		*itə*		*=tə*
2	*ko* *kumu* (honorific)	*udi* *kumu kumu* (honorific)	*=nu* *kumu* (honorific)	*udi* *kumu kumu* (honorific)
3	*si*	*tobo o*	*=nə*	*tobo o*

6.2.2 Reduplication

Full reduplication is productive with both nouns and verbs in Rejang. With verbs, it typically has an aspectual meaning (e.g., *məto~məto* 'walk around', cf. *meto* 'walk'), and with nouns, it marks plurality (e.g., *kuat~kuat* 'friends', cf. *kuat* 'friend') or derives temporal adverbs (e.g., *biləy~biləy* 'daily', cf. *biləy* 'day'). Other types of reduplication are not productive, including reduplication of interrogative pronouns (e.g., *jano~jano* 'whatever', cf. *jano* 'what') and some adverbs (e.g., *lut~lut* 'very much', cf. *lut* 'much, many'). There are several apparent patterns in morphologically complex words. Roots are reduplicated with prefixes *bə-* and *tə-* (e.g., *tə-pəŋᵍaŋ~pəŋᵍaŋ* 'cry out repeatedly' based on *tə-pəŋᵍaŋ* 'cry out (unintentionally)'). With the verbal prefix *meN-,* the root with the substituted nasal is reduplicated (e.g., *mənokoa~nokoa* 'go shopping', cf. *tokoa* 'buy'). In various other instances, it appears that full words are reduplicated with affixes (e.g., *də-maləm~də-maləm* 'nights; at night', cf. *maləm* 'night').

6.2.3 Morphophonology

Affixes are limited to prefixal (e.g., *bə-*, *mə(N)-*) and infixal forms (e.g., <*əm*>, <*ən*>). All affixal forms contain a single reduced vowel /ə/, which is optionally realized when it is a prefix or a prefixal allomorph. Typically, schwa occurs in the prefix when the base begins in a consonant (e.g., *tə-pəŋᵍaŋ* 'cry out') but is absent when it attaches to a vowel-initial base (e.g., *t-iŋət* 'reminisce').

The prefixes *mə(N)-* and *pə(N)-* contain an underspecified nasal prefix (represented by *N*) that triggers nasal substitution. In bases with an initial voiceless oral stop /p, t, k/, the nasal takes the place of articulation of the initial consonant and replaces it in the affixed form (e.g., *peʔeaʔ* 'sour' → *pəmeʔeaʔ* 'spice for souring', *təbo* 'hill' → *mənəbo* 'escape to the hills', *kəleaʔ* 'see' → *məŋəleaʔ* 'stare'). The only other consonant-initial root to which nasal substitution applies is /s/, which surfaces as a palatal nasal [ɲ] (e.g., *suʔət* 'letter' → *məɲuʔət* 'exchange letters'). In vowel initial roots, *N* is realized as a velar nasal [ŋ] (e.g., *aloa* 'flow' → *məŋaloa* 'flow toward'). In all

other roots, the prefix is realized without the underspecified nasal, which is the reason *N* is placed in parentheses in *mə(N)- and pə(N)-* (e.g., *dənoŋ* 'stare' → *mədənoŋ* 'stare', *cəliʔ* 'appearance' → *məcəliʔ* 'appear (as in the sun)', *liləy* 'run' → *məliley* 'run away').

The infixes <*əm*> and <*ən*> also have prefixal allomorphs *m-* or *mə-* and *n-* or *nə-*, respectively. The shape of this affix is determined by two phonological properties of the root: (i) whether or not the root is a single syllable or not, and (ii) the initial phoneme of the root.

1) The infixal form occurs in polysyllabic bases that begin with a non-labial obstruent (e.g., *təko* 'come' → *təməko* 'import'; *kəleaʔ* 'see' → *kəməleaʔ* 'see').
2) In polysyllabic bases that begin with /p/, the initial consonant is replaced by the prefixal form *m-* or *n-* (e.g., *peleaʔ* 'choice' → *meleaʔ*, *neleaʔ* 'choose').
3) In bases beginning in a vowel, *m-* or *n-* is prefixed to the base (e.g., *əm̄uʔ* 'eat' → *məm̄uʔ* 'eat').
4) The prefixal form *me-* or *ne-* occurs in monosyllabic bases (e.g., *lie* 'give' → *məlie* 'give') or polysyllabic bases beginning with /b/ or a sonorant consonant /n, ɲ, ŋ, l, y, w/ (e.g., *liləy* 'run' → *mə-liləy* 'abduct').

6.3 Syntax

Rejang is largely head-initial: prepositions precede the head noun, TAM markers and negators precede the predicate, and (non-numeral) modifiers follow the head noun (see below for examples). Arguments of the predicate are often unrealized in discourse.

Intransitive verbal predicates may be unaffixed, as in (28a), or appear with one of a number of prefixes outlined in the previous section, as in (28b). As in many Austronesian languages there is not a clear distinction between adjectives and stative verbs, both of which can modify a noun or act as a predicate.

(28) Rejang, Intransitive verbal predicates[10]

a.	*si*	*təmot*	*paʔaʔ*	*ku*
	3	sit	near	1SG
	'He sat near me.'			
	(Elicited, 200 sentences)			
b.	*əgo*	*tənoa*	*mə-neʔ*	
	price	egg	INTR-rise	
	'The price of eggs rose.'			
	(Elicited, 200 sentences)			

6.3.1 Voice

With few exceptions, transitive verbal predicates are marked by the AV or PV infix, as in (29).

(29) Rejang, Transitive alternations

a.	*si*	*k<əm>leaʔ*	*imuo*
	3	<AV>see	tiger

[10] In the examples in McGinn's unpublished manuscript, clitic pronouns are written as separate words, and we have chosen to follow this convention. However, it is clear from the form of the pronoun whether it should be considered a free or clitic pronoun (see Section 6.2.1).

‘He saw tigers.’
(Elicited)

b. *imuo* *k<ən>leaʔ* *nə*
tiger <PV>see 3
‘He saw tigers.’
(Elicited)

Intransitive and PV constructions have variable word order wherein the predicate complex may appear before or after the subject. Compare examples in (29) to those in (30). In AV constructions, the subject precedes the predicate complex. It is apparently not possible for the subject to follow the predicate or the predicate complex.

(30) Rejang, Variable word order

a. *n-əm̄in* *nə* *bəlas* *ŋən* *siləy*
PV-bring 3 uncooked.rice with salt
‘He brought uncooked rice and salt.
(Narrative *Lalan* ‘Lalan’)

b. *tə-kəjir* *si*
NVOL-surprise 3
‘He was amazed.’
(Narrative *Lalan* ‘Lalan’)

In ditransitive constructions, the R(ecipient) and T(heme) are unmarked and, in AV constructions (31), follow the verb in that order. In PV constructions, it appears that either R or T may occur as the subject, as in (32) and (33). It is worth noting that when a pronoun is R, it occurs in the non-clitic form.

(31) Rejang, Ditransitive in AV

uku *galaʔ* *t<əm>uɲ̄uʔ* *si* *baso* *Ejang* *o.*
1SG always AV-show 3 language Rejang that
‘I always show him the Rejang language.’
(Conversation in *Dusun Duku Ilir*)

(32) Rejang, Ditransitive in PV

jərusan *jano* *n-ajəa* *nə* *ko,* *Bi?*
subject what PV-teach 3 2SG Bi
‘What subjects does he teach you, Bi?’
(Conversation in *Dusun Duku Ilir*)

(33) Rejang, Ditransitive in PV

si *galaʔ* *t<ən>uɲ̄uʔ* *ku* *baso* *əjaŋ.*
3 enjoy PV-show 1SG language Rejang
‘I enjoy showing him the Rejang language.’
(Elicited)

6.3.2 TAM markers and negators

Rejang has a number of TAM markers as well as negation expressed by preverbal elements, such as tense markers (e.g., *ati* ‘not yet’), aspect markers (e.g., *bi* ‘already’, *masih, gi* ‘still’), and modals (e.g., *buleaʔ* ‘may’, *nam* ‘can’). The examples in (34) and (35) exemplify aspectual markers and modals, respectively.

(34) Rejang, Aspect
si *bi* *beleʔ.*
3 already return
'He went home.'
(Elicited)

(35) Rejang, Modal
uku *bulеaʔ* *t<əm>aboaʔ* *nəʔ*
1SG may <ACT>beat it
'May I beat it?'
(Narrative 'Mousedeer the Trickster')

Verbal negation is expressed by *coa* 'not' (36), while nominal negation is expressed by *iso* 'not' (37).

(36) Rejang, Verbal negation
si *coa* *tə-kəjir.*
he not STAT-surprised
'He was not surprised.'
(Elicited)

(37) Rejang, Nominal negation
Alui *iso* *mahasiswa.*
Alui not student
'Alui is not a student.'
(Elicited)

6.3.3 Existential and possessive constructions

The verb *ade* 'exist, be, have' is used in existential and possessive constructions, as in (38) and (39), respectively.

(38) Rejang, Existential construction
ade *tun* *bujaŋ* *di* *diəm* *naʔ* *də-buəaʔ* *sadie.*
exist person unmarried REL live in one-CLF village
'There was a young man who lived in a village.'
(Elicited)

(39) Rejang, Possessive construction
Arma *ade* *tələw* *likup* *anaʔ.*
Arma has three child CLF(tail)
'Arma has three children.'
(Elicited)

In non-verbal predicate constructions, it is possible to use *ade* followed by the particle *ba* as a copula, as in (40). Common non-verbal predicates are NPs or PPs.

(40) Rejang, Predicate NP
CuʔupCurup *ade ba* *kota* *səŋaʔ*
Curup COP city cold
'Curup is a cold city.'
(Narrative "My Village")

6.3.4 Nonverbal predicates

Non-verbal predicates may also occur without any copula, being expressed as the juxtaposition of the subject and the predicate, as in (41).

(41) Rejang, Predicate NP

gen	*nə*	*Bujaŋ*	*Kuruŋ,*
name	his	Bujang	Kurung

'His name was Bujang Kurung.'
(Narrative *Lalan* 'Lalan')

6.3.5 Noun phrase structure

NPs have the following structure: NUM CLASS N ADJ NP$_{POSS}$ DEM, as in (42) and (43). However, numerals and classifiers may also follow the head noun.

(42) Rejang, NP structure

a.

duəy	*buəaʔ*	*maɲ̄is*
two	CLF(fruit)	mangosteen

'two mangosteens'
(Elicited)

b.

maɲ̄is	*duəy*	*buəaʔ*
mangosteen	two	CLF(fruit)

'two mangosteens'
(Elicited)

(43) Rejang, NP structure

Tun	*tuəy*	*o*	*kə-mojoa*	*kulo*	*co-ade*	*anaʔ.*
person	old	that	DEF-lucky	also	not-have	child

'The old man was very fortunate (because he) had no children.'
(Elicited)

Relative clauses follow the head noun and are introduced by the relativizer *di*, as in (44). Headless relative clauses with *di* are also common, as in (45).

(44) Rejang, Relative clause

tun	[*di*	*nə-liə*	*səlawie*	*o*	*caci*]	*o…*
person	REL	PV-give	woman	that	money	the

'The person that was given money by that woman…'
(Elicited)

(45) Rejang, Headless relative clause

di	*nə-lie*	*Alui*	*magəa*	*Arma*	*(ade ba)*	*buŋəy.*
REL	PV-give	Alui	to	Arma	COP	flower

'What was given by Alui to Arma was flowers.'
(Elicited)

7 Sketch of Mentawai

Mentawai (glottocode: ment1249, ISO 639-3: mwv) is a Sumatran language spoken by around 60,000 people in the Mentawai Archipelago (Eberhard et al. 2025), an island chain located to the west of Sumatra and consisting primarily of four large islands. While there is some variation in Mentawai spoken on the Pagai Islands, Sipora, and southern Siberut, the greatest amount of dialect diversity, both in terms of lexicon and grammar, is found in the northern half

of the island of Siberut (Pampus 1989), the proposed homeland of the Mentawai people (Tulius 2012). Unless otherwise noted, this sketch describes the variety of Mentawai spoken on the island of Sipora as analyzed by Morris (1900), Adriani (1928), and Willms (1956).

7.1 Phonology

The consonant inventory for Mentawai is as laid out in Table 13 below. In varieties of Mentawai spoken in the Nassau islands, there is variation in the phonetic realization of some of the consonant phonemes, namely /g/ freely varies between [g] and [ɣ] as does /p/ between [p] and [f] (Morris 1900: 2). Notably, the Mentawai phoneme inventory lacks the semivowels /j/ and /w/ outside of the phonemic diphthongs, having historically merged with /d͡ʒ/ and /b/, respectively. A phoneme /t͡ʃ/ <c> is a recent import found almost exclusively in loans from Malay.

Table 13. Mentawai consonant inventory.

		Bilabial	Alveolar	Palatal	Velar
Stop	**Voiced**	b	d		g <g>
	Voiceless	p	t		k
Affricate				d͡ʒ <j>	
Fricative			s		
Nasal		m	n	ɲ <ny>	ŋ <ng>
Trill			r		
Lateral			l		

The vowel inventory (Table 14) is a canonical five-vowel system and does not contain the /ə/ commonly encountered elsewhere in SBI languages. The inventory of diphthongs (Table 15), on the other hand, is very rich. Analysis as diphthongs and not vowel-glide sequences is justified by stress placement rules where a diphthong is counted as a single syllable nucleus. Many of these have arisen as a result of palatalization adjacent to the alveolar phonemes /s/, /n/, /r/, /l/ (e.g., /asu/~/aɨsu/ ‘to fetch water’). Others can be traced from inherited diphthongs, as in PMP *sapaw ‘field hut’ > Mentawai /sapou̯/ ‘field hut’, or to a shift of original final PMP *g, *j, *k > Mentawai /u̯/ (e.g., PMP *pusəj ‘navel’ > Mentawai /pusou̯/ ‘navel’).

Table 14. Mentawai vowel inventory.

	Front	Central	Back
High	i		u
Mid	e		o
Low		a	

Table 15. Mentawai diphthong inventory.

	Front	Central	Back
High-High	uɨ̯		iu̯

Mid-High	e̯i̯	o̯u̯
Low-High	a̯i̯	a̯u̯

The basic syllable structure of Mentawai is (C)V(C). Minimally, then, a syllable can consist simply of a single vowel or diphthong. The majority of syllables, however, contain at least an onset. Any of the consonants can fill the onset position. The coda is much more restricted; word-medially, it can only be occupied by a nasal homorganic with the following consonant. In the variety of Mentawai spoken on Sipora, these nasal-stop sequences are realized as geminate stops. Word-finally, only the voiceless stops and non-palatal nasals are found in coda position.

The restriction on /r/ and /l/ appearing word-finally is a less common constraint. In avoidance of word-final /r/ and /l/, an echo of the preceding vowel is appended after the liquid, thus creating trisyllabic roots from disyllabic roots. As a consequence, whereas stress is regularly penultimate, in forms that have arisen in such a manner, stress is retained on the new antepenult (PMP *bakul 'basket' > Mentawai *bakulu* [ˈbakulu] 'container'). Whether these words should in actuality be analyzed as underlyingly di- or tri-syllabic remains unclear.

Roots are typically disyllabic, most being of the structure CV(N)CVC. Complex forms, whether reduplicated or affixed, do not necessarily adhere to the consonant cluster constraints detailed above, as in *buk-buk* 'to burn' and *lalep=ku* 'house=1SG'. Stress in Mentawai is not phonemic, can fall equally on a simple vowel and on a diphthong, and is regularly penultimate.

7.2 Morphology

As is common in the languages of Sumatra, Mentawai is predominantly prefixing. Affixes are too numerous to discuss each in depth, so a general overview follows:

- *ma-*: Forms attributive or stative verbs, as in *ma-sulu* 'sunny' from *sulu* 'sun' and *ma-batu* 'rocky' from *batu* 'stone'.
- *maN-* and *masi-*: Both form A-voice (AV) verbs, as in *man-ebek* 'AV-pierce' from *tebek* 'to pierce' and *masi-ala* 'AV-fetch' from *ala* 'to fetch'; the distinction between these two prefixes is unclear in available sources.
- *mu-*: Forms intransitive verbs, as in *mu-luji* 'INTR-spit' from *luji* 'to spit' and *mu-urai* 'INTR-sing' from *urai* 'song'.
- *pa-*: Forms causative or reciprocal verbs, as in *pa-kom* 'to feed' from *kom* 'to eat' and *pa-siguk* 'to kiss each other' from *siguk* 'to kiss'.
- *paN-*: Forms nouns, frequently either the process of a verb or an instrument, as in *pan-úgulu* 'the felling of trees' from *túgulu* 'to fell a tree' and *pan-úsuru* 'knife for scaling fish' from *súsuru* 'fish scales'.
- *pu-*: Nominalizes agents, as in *pu-luga* 'rower' from *luga* 'to row' and *pu-toro* 'one that shines' (i.e., sun, moon, stars, etc.) from *toro* 'to shine'.
- *i-*[11]: Forms P-voice (PV) verbs, as in *i-seu* 'PV-cook' from *seu* 'to cook' and *i-ala* 'PV-take' from *ala* 'to take'.
- *tu-*: Forms non-volitional verbs with an unspecified agent as in *tu-betek* 'made tense' from *betek* 'tense' and *tu-ale* 'became better' from *ale* 'to heal'.

[11] A morpheme of the same phonological form is both analyzed as the PV prefix and the 3S proclitic. It is unclear from available descriptions what difference, if any, there is between the two.

- *-i*: Acts as an applicative suffix (marking locations, goals, and other arguments in transitive constructions), as in *enung-i* 'to enter' from *enung* 'to walk' and *arep-i* 'to listen to' from *arep* 'to hear'.
- *-ake*: Acts as a causative or instrumental applicative, as in *a-reu-ake* 'to remove' from *a-reu* 'far' and *a(i)su-ake* 'to fetch water with' from *a(i)su* 'to fetch water'.
- *-en*: Forms patient nominalizations, as in *urem-en* 'a plant' from *urep* 'to plant' and *rep-dem-en* 'gift of remembrance' from *rep-dem* 'to remember'.
- *-an*: Forms locative nominalizations, as in *enung-an* 'road' from *enung* 'to walk' and *pileng-an* 'rudder handle' from *pileng* 'to rotate'.

Nearly all of these affixes are fairly transparent inheritances from PMP. The PMP infixes *<um> and *<in> are no longer productive in Mentawai but can be identified in some fossilized forms, as in *t<in>ibo* 'smoked meat' from *tibo* 'smoke'. A great number of combinations of affixes with relatively transparent semantics are permissible in Mentawai.

7.2.1 Personal pronouns

Mentawai has three sets of pronominal forms, as shown in Table 16. The first of these is the set of independent forms. The second set is used to denote A for AV or unmarked verbs. When *kai* '1PL.EXCL' or *kam* '2PL' is A, it immediately follows the verb and must be indexed on the verb with the singular form of the proclitic *ku=* '1SG' or *nu=* '2SG', respectively (e.g., *nu-galai kam* 'we make'; compare also the examples in (47a), (67) below). Pronominal forms of the last set are used either attached to a verb, where they denote the P argument in AV or the A argument in PV, or attached to a noun, where they denote possession.

Table 16. Mentawai pronominal forms.

	Independent		**AV A**		**AV P, PV A, Possessor**	
	Singular	**Plural**	**Singular**	**Plural**	**Singular**	**Plural**
1.EXCL	*aku*	*kai*	*ku=*	*ku=... kai*	*=ku*	*=mai*
1.INCL		*sita*		*ta=*		*=ta*
2	*ekeu*	*kam*	*nu=*	*nu=... kam*	*=nu/=m*	*=mui*
3	*nia*	*sia*	*i=*	*da=/ra=*	*=nia*	*=da/=ra*

7.2.2 Reduplication

In Mentawai, nouns, verbs, and numerals can be reduplicated. This typically takes one of two forms: either full reduplication (e.g., *bit-bit* 'to fan flames (repeatedly)' from *bit* 'to fan flames') or initial-syllable reduplication (e.g., *ka-kabili* 'fishhook' from *kabili* 'to fish'). In full reduplication of a trisyllabic word that ends in a liquid followed by an echo vowel, typically only the first two syllables are reduplicated (e.g., *pute-putere* 'to twist' from *putere* 'to spin'). Both full and initial-syllable reduplication can be used to express repetition or intensive action for verbs (as in *bit-bit* above) or diverse plurality for nouns (e.g., *bi-bilet* 'household goods' from *bilet* 'a thing'). Initial-syllable reduplication is used to derive instrumental nouns from verbs (as in *ka-kabili* above) and to derive distributives from numerals in combination with the prefix *pa-* (e.g., *pa-te-telu* 'each of three' from *telu* 'three'). Reduplication can co-occur with other affixes

(e.g., *ti-tibo-an* ‘all sorts of chatter’ from *tibo* ‘to talk’) and even operate over these other affixes (e.g., *ma-na-naro* ‘AV-bring (multiple)’ from *taro* ‘to bring’ with the AV prefix *maN-*).

7.2.3 Morphophonology

The Mentawai nasal substitution prefixes *maN-* and *paN-* alter form based on the root they combine with. When attached to vowel-initial roots, these take the form of *maŋ-* and *paŋ-* (e.g., *maŋ-urep* ‘AV-plant’ from *urep* ‘to plant’). In the case of the voiceless stops /p, t, k/, the fricative /s/, and the voiced palatal affricate /j/, assimilation similarly occurs and the original consonant is replaced entirely (e.g., *man-akou* ‘AV-steal’ from *takou* ‘to steal’, *man-uba* ‘AV-fish’ from *suba* ‘fishnet’, *pan-a~ja* ‘cutting tool’ from *ja~ja* ‘to cut’). Before nasals, the nasal element of the prefix does not surface as a separate element (e.g., *pa-min* ‘cord for fastening’ from *min* ‘to fasten’). Neither Morris (1900), Adriani (1928), nor Willms (1956) explicitly discuss nasal substitution on roots beginning with a voiced stop /b, d, g/ or liquid /l, r/. Regardless, the available data for such roots seems to indicate overwhelming use of *masi-* for AV forms (e.g., *masi-buka* ‘AV-open’ from *buka* ‘to open’) and *pa-* without a nasal element for derived nominals (e.g., *pa-luga* ‘oar’ from *luga* ‘to row; rudder’, the former noted in Cambielli n.d.: 92), although the nominalizing *pa-* without a nasal element is found with other roots as well (e.g., *pa-sogi* ‘hair comb’ from *sogi* ‘to make fringes’).

The genitive marker *=n*[12] ‘GEN’, changes form based on the preceding coda and the following onset. According to Adriani (1928), if the following word does not have an onset, the genitive marker surfaces as *=n* (*kabei=n ukui* ‘the hand of the father’). If the following word does have an onset but the preceding word neither ends in a consonant nor a diphthong, the marker surfaces as *=t* (*bulu=t ŋuŋu* ‘beard’, lit. ‘hair GEN mouth’). In the case that the genitive marker is bound both by a preceding consonant or diphthong and a following consonant, the marker does not surface at all (*alai kapala* ‘head hair’). Given the variability of the realization of final nasals and voiceless stops in varieties of Mentawai, this analysis probably differs by dialect.

Other affixes and markers typically do not alter form based on the root they attach to. Even when the vowel of an affix comes into contact with the vowel of a root, both vowels are often retained. In the case where the applicative *-i* is attached to a root with a final diphthong ending in /i/, the suffix may surface as *-ji* (e.g., *teitei-ji* ‘to go over a bridge’ from *teitei* ‘bridge), presumably through strengthening of an epenthetic glide.

7.3 Syntax

The unmarked word order for Mentawai is VS for intransitive constructions (46a) and VAP for transitive (46c). Word order, however, is not strict, and intransitive subjects are frequently fronted for prominence (46b).

(46) Mentawai, Clausal word order

a. Intransitive VS

a=mu-macep	*aku*
PFV=INTR-morning	1SG

‘I was busy this morning.’
(Morris 1900: 274; Adriani 1928: 50)

b. Intransitive SV

[12] Although this marker is variously described as an enclitic and a proclitic, it is perhaps best analyzed as an enclitic based on the fact that it can further be followed by the enclitic *=nia* ‘3SG’.

sita *kam* *pa-ombun* *ka* *uggu.*
1PL.INCL 2PL RECIP-play on floor
'Let's play together on the floor.'
(Adriani 1928: 58)

c. Transitive VSO
galai *si rimanua* *kinapat.*
make person boat
'The person made a boat.'
(Morris 1900: 54)

Mentawai is primarily head-initial – TAM marking precedes verbs (47a), NP modifiers (apart from numerals) and demonstratives follow their head noun (47b), and prepositions rather than postpositions are used (47c). This order also holds true for genitive constructions as described below.

(47) Mentawai, Head-initial word order

a. TAM > Verb
alepaan *a=ku=galai* *kai* *enungan*
already PFV=1SG=make 1PL.EXCL path
'We have already finished the path.'
(Adriani 1928: 97)

b. Noun > Demonstrative
si rimanua *ne*
person DEM.PROX
'this person'
(Morris 1900: 35)

c. Head Noun > PP; Preposition > Noun
mei *aku* *ka* *laggai*
go 1SG to village
'I am going to the village.'
(Adriani 1928: 99)

7.3.1 Voice

Verbs can be marked for one of two transitive voice constructions: A-Voice (AV) or P-Voice (PV). In AV constructions, use of the prefix *maN-* is optional; compare (48) and (49). In PV constructions (50), the A argument, if present, must immediately follow the verb, whether in the form of a pronominal enclitic or an independent form. The P argument in PV constructions is most frequently found preceding the verb but can also occur after the predicate complex.

(48) Mentawai, AV
aku *mang-akkak* *tobat* *sapou*
1SG AV-sew roof hut
'I sewed the roof of the field hut.'
(Morris 1900: 162)

(49) Mentawai, Unmarked AV
sapo *aku* *guluk=nu*
infect 1SG skin.disease=2SG
'Your skin disease infected me.'
(Morris 1900: 176)

(50) Mentawai, PV

tak	*i-oba=nu*	*kai*
NEG	PV-want=2SG	1PL.EXCL

'You do not want us.'
[lit. 'We are not wanted by you.']
(Willms 1956: 52)

In ditransitive constructions, as in (51), the recipient argument immediately follows the theme argument and is unmarked.

(51) Mentawai, AV ditransitive

ei	*sita*	*ta=kau*	*iba=da*	*sa*	*kalabai=nu*
go	1PL.INCL	1PL.INCL=give	food=3SG	ART	aunt=2SG

'We want to bring food to your aunt.'
(Morris 1900: 120)

As exemplified by the AV ditransitive in (52) below where both the theme and recipient are implied, arguments are frequently unrealized, presumably if they are understood from discourse.

(52) Mentawai, AV ditransitive (implicit theme and recipient)

tidou	*pei*	*gou-gou*	*rua*	*ele*	*telu*	*ngamuneng*	*ka*	*tubu=mai,*	*buleu*	*i=ake*
request	just	chicken	two	or	three	CLF(animal)	for	body=1PL.EXCL	maybe	3SG=give

'Just ask for two or three chickens for us, if maybe he will give (us them).'
(Adriani 1928: 83)

7.3.2 TAM markers and negators

Apart from the perfective proclitic *a=* 'PFV', most TAM marking is expressed via preverbal auxiliaries that mark tense *geti* 'will' (53), aspect *alepaan* 'already' (47a, above), and modal meanings *ai leu* 'must' and *ai pei* 'must' (67, below). Negation is expressed via the marker *tak* 'NEG' (54), with the marker *ba* 'VET' as a vetative (55).

(53) Mentawai, Tense

ma-uneng	*geti*	*aku.*
STAT-sleep	will	1SG

'I will sleep.'
(Morris 1900: 18)

(54) Mentawai, Negation

tak	*ku=agai.*
NEG	1SG=know

'I don't know.'
(Morris 1900: 38)

(55) Mentawai, Vetative

ba	*ei*	*kam*
VET	go	2PL

'I will sleep.'
(Morris 1900: 38)

7.3.3 Existential and possessive constructions

Existentials in Mentawai are often expressed using the existential verb *bara* 'to exist', as in (56) and (57), or, in the negative, using *tak u* 'to not exist', as in (58). The verb *bara* may also function as a copula.

(56) Mentawai, Existential
bara sara uma.
exist one house
'There is one house.'
(Morris 1900: 148)

(57) Mentawai, Existential
bara ukui ka sila.
exist father at side
'Father is on the other side.'
(Morris 1900: 112)

(58) Mentawai, Negative existential
tak u si agai=nia.
not exist REL know=3SG
'There is no one who knows.'
(Morris 1900: 36)

Possession in Mentawai is indicated either through a pronominal enclitic or through a separate NP connected by the genitive marker *=n* 'GEN', although this marker may or not have a surface realization depending on the surrounding phonological context, as in (59) and (60).

(59) Mentawai, Genitive with enclitic
iba=m
fish=2SG
'your fish'
(Morris 1900: 34)

(60) Mentawai, Genitive with *=n*
sese=n uba neda
ash=GEN tobacco DEM.DIST
'that tobacco ash'
(adapted from Morris 1900: 16)

7.3.4 Nonverbal predicates

As is the case for many other SBI languages, Mentawai permits clauses with nonverbal predicates, most frequently either an NP (61) or a PP (62).

(61) Mentawai, Nonverbal predicate: NP
piga usu si rimanua?
how.many rib person
'How many ribs does a person have?'
(Morris 1900: 172)

(62) Mentawai, Nonverbal predicate: PP
koilip uma sarat toitet.

around house only coconut
'There are only coconut trees around the house.'
(Morris 1900: 22)

7.3.5 Noun phrase structure

Noun phrases have the general structure laid out in (63) below.

(63) Mentawai, NP structure
NUM LIG-CLF N REL DEM

Examples of NPs with demonstratives have already been given in (47b) and (60); examples for NPs containing numerals and classifiers are given below. When a numeral is used, classifiers are frequently found but are not obligatory. Whenever a classifier is used, the ligature *nga* is obligatorily inserted between it and the preceding numeral.

(64) Mentawai, NP
lima *nga-muneng* *jo'jo'*
five LIG-CLF(animal) dog
'five dogs'
(Morris 1900: 23)

(65) Mentawai, NP
balu *gougou*
eight chicken
'eight chickens'
(Adriani 1928: 92)

Additionally, the numeral-classifier complex may also follow the noun it modifies.

(66) Mentawai, NP
atelu *lima* *ng-ambe*
egg five LIG-CLF(seeds/eggs)
'five eggs'
(Adriani 1928: 91)

Both possessors and relative clauses follow the noun they modify. Relative clauses are formed with the relativizer *si* and can be headed (67) or headless (68). Stative verbs cannot directly modify a noun without the relativizer; when a noun and stative verb not marked with *si* are adjacent, the stative verb typically functions as a predicate (69).

(67) Mentawai, Headed relative with stative
bago *si* *ma-nanam*
banana REL ATR-tasty
'a tasty banana'
(Morris 1900: 36)

(68) Mentawai, Headless relative with stative
si *ma-peset*
REL ATR-blind
'the blind (one)'
(Adriani 1928: 97)

(69) Mentawai, Stative predicate
ma-nanam *bago.*
ATR-tasty banana
'Bananas are tasty.'
(Morris 1900: 36)

Only the A argument in AV and the P argument in PV can be extracted from the relative clause, as shown in (70) for AV and (71) for PV. It is important to note that when the relativizer *si* immediately precedes a PV-marked verb, the verb cannot be marked with *i-*, as in (71). This can make it difficult to discern between unmarked AV verbs and PV verbs in relative clauses. The example in (71) can be identified as PV by the PV agent enclitic *=mui*.

(70) Mentawai, Extracted A in AV

ai leu	*si rimanua*	*si*	*agai=nia*	*masi-ale-ake*	*ekeu.*
must	man	REL	can=3SG	AV-heal-CAUS	2SG

'There must be someone who can heal you.'
(Adriani 1928: 64)

(71) Mentawai, Extracted P in PV

tak	*ma-eru*	*enungan*	*si*	*galai=mui.*
NEG	ATR-good	path	REL	make=2PL

'The road that you all have made is not good.'
(Adriani 1928: 87)

8 Conclusion

This chapter provides an overview of the sociolinguistic situations, phylogenetics, and typology of SBI languages, alongside short sketch grammars of three representative languages: Jambi Malay, Rejang, and Mentawai. SBI languages are highly diverse with complex histories of migration and patterns of language contact and multilingualism. Building from earlier proposals (Nothofer 1986), the Sumatran subgroup has been more firmly established in recent years (Smith 2017; Billings & McDonnell 2024), while the Malayic subgroup has long been recognized (Adelaar 1992). However, the place of Rejang and Lampungic languages within Malayo-Polynesian remains unclear, as does much of the internal structure of these subgroups.

SBI languages are difficult to characterize typologically as they vary at every level of linguistic structure. Consonant and vowel inventories range in size from medium (e.g., Batak languages) to very large (e.g., Acehnese). SBI languages show a range of prefixes that trigger nasal substitution, infixes, circumfixes, and reduplication, with varieties of Kerinci demonstrating a typologically marked derivational syllable rime alternation (see Ernanda 2017) and Nias exhibiting marked patterns of initial consonant mutation in nouns and pronouns (see Brown 2001). Many of the SBI languages have a simplified symmetrical voice system with a two-way distinction, although several languages of the Barrier Islands, such as Enggano and Nias, no longer have a symmetrical voice system in main clauses.

Over the past two decades, there have been concerted efforts to document and describe languages of Sumatra (see McDonnell 2017). These include surveys of Malay (e.g., Anderbeck 2008; McDowell & Anderbeck 2020) and Lampungic varieties (Anderbeck 2007a) in southern Sumatra as well as documentation and description of Jambi Malay (e.g., Yanti 2010; Yanti, Cole, and Hermon to appear), Kerinci (e.g., McKinnon 2011, Ernanda 2017), Besemah (e.g.

McDonnell 2016), Enggano (e.g., Hemmings et al. 2023; Hemmings & Arka 2023), and Nasal (e.g., Hakim 2024; Billings and McDonnell 2025; McDonnell to appear). Despite these efforts, there are still major gaps in the documentation and description of SBI languages. There is very little documentation and description of the Barrier Islands languages, especially Simeulue, Haloban, Leukon, Sigulai, and Mentawai, and languages of Sumatra, especially Lampungic languages, Rejang, and many of the Batak languages, such as Alas, Kluet, Dairi, Simalungun, Angkola, and Mandailing. This lack of documentation and description limits our ability to subgroup languages and characterize SBI languages typologically, and more research in all of these areas is needed.

Abbreviations

ABSL	Absolute	FUT	Future	PASS	Passive
ACT	Active	GEN	Genitive	PL	Plural
ART	Article	INCL	Inclusive	POL	Polite
ATR	Attributive	INTR	Intransitive	POSS	Possessive
AV	A-voice	IRR	Irrealis	PFV	Perfective
CAUS	Causative	LIG	Ligature	PROG	Progressive
CLF	Classifier	LOC	Locative	PROX	Proximal
COP	Copula	MUT	Mutated form	PV	P-voice
DAT	Dative	NEG	Negative	REAL	Realis
DEF	Definite	NMLZ	Nominalizer	RECIP	Reciprocal
DEM	Demonstrative	NUM	Numeral	REL	Relative
DIST	Distal	NVOL	Nonvolitional	SG	Singular
EMPH	Emphatic	OBL	Oblique	STAT	Stative
EXCL	Exclusive	OV	Object voice	VET	Vetative
EXCLAM	Exclamative	PART	Particle		

References

Adelaar, K. Alexander. 1981. Reconstruction of Proto-Batak phonology. In Blust, Robert A. (ed.), *Historical linguistics in Indonesia: Part 1*, 1–20. Jakarta: Universitas Katolik Indonesia Atma Jaya.

Adelaar, K. Alexander. 1992. *Proto-Malayic: A reconstruction of its phonology and parts of its lexicon and morphology*. Canberra: Pacific Linguistics.

Adelaar, K. Alexander. 1995. Toba Batak. In Darrell T. Tryon (ed.), *Comparative Austronesian dictionary: An introduction to Austronesian studies*, vol. 1, 421–432. Berlin ; New York: Mouton de Gruyter.

Adelaar, K. Alexander. 2005a. Structural Diversity in the Malayic Subgroup. In K. Alexander Adelaar & Nicholas P. Himmelmann (eds.), *The Austronesian languages of Asia and Madagascar* (Routledge Family Language Series), 202–226. New York: Routledge.

Adelaar, Alexander. 2005b. Malayo-Sumbawan. *Oceanic Linguistics* 44(2). 357–388.

Adelaar, Alexander. 2007. Book Review: *Issues in Austronesian historical phonology*, by John Lynch. *Bijdragen tot de Taal-, Land- en Volkenkunde* 163(1). 139–146.

Adelaar, K. Alexander & David J. Prentice. 1996. Malay: Its history, role and spread. In Stephen Adolphe Wurm, Peter Mühlhäusler & Darrell T. Tryon (eds.), *Atlas of Languages of Intercultural Communication in the Pacific, Asia and the Americas*, 6736–6793. Berlin: Mouton de Gruyter.

Adriani, N. 1928. Spraakkunstige schets van de taal der Mĕntawai-eilanden. *Bijdragen tot de taal-, land- en volkenkunde van Nederlandsche-Indie* 84. 1–117.

Anderbeck, Karl. 2007a. An Initial Reconstruction of Proto-Lampungic: Phonology and Basic Vocabulary. *Studies in Philippine Languages and Cultures* 16. 41–165.

Anderbeck, Karl. 2007b. Haji: one language from twelve. In Chong Shin, Karim Harun & Yabit Alas (eds.), *Reflections in Southeast Asian seas: Essays in honour of Professor James T. Collins*, vol. 2, 51–91. Pontianak: STAIN Pontianak Press.

Anderbeck, Karl. 2008. *Malay dialects of the Batanghari River Basin (Jambi, Sumatra)* (Sil E-Books 06). Dallas: SIL International. https://www.sil.org/resources/archives/9245.
Anderbeck, Karl. 2012. The Malayic-speaking; Orang Laut Dialects and directions for research. *Wacana, Journal of the Humanities of Indonesia* 14(2). https://doi.org/10.17510/wacana.v14i2.64.
Anderbeck, Karl R. & Herdian Aprilani. 2013. *The Improbable Language: Survey Report on the Nasal Language of Bengkulu, Sumatra*. SIL International.
Arka, I Wayan. 2003. *Balinese morphosyntax: a lexical-functional approach* (Pacific Linguistics 547). Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, Australian National University.
Aziz, Zulfadli A. & Robert Amery. 2016. A survey on the status of the local languages of Pulau Simeulue and Pulau Banyak and their use within the community. *Proceedings of the 1st English Education International Conference* 1(2). 487–490.
Billings, Blaine & Bradley McDonnell. 2024. Sumatran. *Oceanic Linguistics* 63(1), 112–174.
Billings, Blaine & Bradley McDonnell. 2025. Fillers in the gray: Indeterminacy in their use in everyday conversation. In Brigitte Pakendorf & Françoise Rose (eds.), *Fillers: hesitatives and placeholders* (Research on Comparative Grammar 5), 211–250. Berlin: Language Science Press. https://doi.org/10.5281/zenodo.15697585.
Blench, Roger. 2014. *The Enggano: Archaic foragers and their interactions with the Austronesian world*. Unpublished ms. http://www.rogerblench.info/Language/Austronesian/Enggano/Enggano%20and%20its%20history.pdf
Blust, Robert. 1981. The reconstruction of proto-Malayo-Javanic: an appreciation. *Bijdragen tot de taal-, land- en volkenkunde / Journal of the Humanities and Social Sciences of Southeast Asia* 137(4). 456–469. https://doi.org/10.1163/22134379-90003492.
Blust, Robert. 1984. On the History of the Rejang Vowels and Diphthongs. *Bijdragen tot de Taal-, Land- en Volkenkunde* 140(4). 422–450.
Blust, Robert. 1988. Malay historical linguistics: A progress report. In Mohd. Thani Ahmad & Zaini Mohamed Zain (eds.), *Rekonstruksi dan cabang-cabang Bahasa Melayu Induk* (Siri Monograf Sejarah Bahasa Melayu), 1–33. Kuala Lumpur: Dewan Bahasa dan Pustaka.
Blust, Robert. 1994. The Austronesian settlement of mainland Southeast Asia. In Adams, K. L. & T. J. Hudak (eds.), *Papers from the Second Annual Meeting of the Southeast Asian Linguistics Society*, 22–83. Tempe: Arizona State University.
Blust, Robert. 2010. The Greater North Borneo Hypothesis. *Oceanic Linguistics* 49(1). 44–118. https://doi.org/10.1353/ol.0.0060.
Blust, Robert. 2013. *The Austronesian languages* (Pacific Linguistics 602). Revised. Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, Australian National University.
Brown, Lea. 2001. *A Grammar of Nias Selatan*. Sydney: University of Sydney PhD Thesis.
Cambielli, Daniele. n.d. *Kamus Simalegi*. Unpublished ms.
Capell, Arthur. 1982. Bezirkssprachen im gebiet des UAN. In Rainer Carle, Martina Heinschke, Peter Pink, Christel Rost & Karen Stadtlander (eds.), *Gava': Studies in Austronesian languages and cultures dedicated to Hans Kähler*, 1–14. Berlin: Dietrich Reimer.
Chen, Victoria & Bradley McDonnell. 2019. Western Austronesian voice. *Annual Review of Linguistics* 5. 173–195.

Clynes, Adrian. 1997. On the Proto-Austronesian diphthongs. *Oceanic Linguistics* 36(2). 347–361.
Coady, James & Richard McGinn. 1982. On the so-called implosive nasals of Rejang. In Rainer Carle (ed.), *GAVA: Studies in Austronesian Languages and Cultures*, 437–449.
Cole, Peter, Gabriella Hermon, & Yanti. 2008. Voice in Malay/Indonesian. *Lingua* 118(10). 1500–1553. https://doi.org/10.1016/j.lingua.2007.08.008.
Crowley, Terry. n.d. *Grammatical sketch of Enggano*. Unpublished ms.
Durie, Mark. 1985. *A Grammar of Acehnese on the Basis of a Dialect of North Aceh*. Dordrecht: Foris Publications.
Durie, Mark. 1987. Grammatical Relations in Acehnese. *Studies in Language* 11(2). 365–399. https://doi.org/10.1075/sl.11.2.05dur.
Dyen, Isidore. 1965. *A Lexicostatistic Classification of the Austronesian Languages* (Supplement to International Journal of American Linguistics Memoir 20). Baltimore: Waverly Press.
Eades, Domenyk. 2005. *A grammar of Gayo: A language of Aceh, Sumatra* (Pacific Linguistics 567). Canberra: The Australian National University.
Eberhard, David M., Gary F. Simons, & Charles D. Fennig (eds.). 2025. *Ethonologue: Languages of the world*, 28th edn. Dallas, Texas: SIL International. Online version: https://www.ethnologue.com/
Edwards, Owen. 2015. The Position of Enggano within Austronesian. *Oceanic Linguistics* 54(1). 54–109. https://doi.org/10/gd6d95.
Ernanda. 2017. *Phrasal alternation in Kerinci*. Leiden: University of Leiden. (Doctoral dissertation.)
Gil, David. 2020. Dual heritage: The story of Riau Indonesian and its relatives. In David Gil & Antoinette Schapper (eds.), *Austronesian Undressed: How and why languages become isolating* (Typological Studies in Language), vol. 129, 119–212. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/tsl.129.03gil.
Hemmings, Charlotte & I Wayan Arka. 2023. Evidence of contact with Malay/Indonesian in the Enggano language. In Hiroki Nomoto & Asako Shiohara (eds.). *Language contact between Malay and indigenous languages. NUSA* 75. 19–51. https://doi.org/10.15026/0002000126.
Hemmings, Charlotte, I Wayan Arka, Engga Zakaria Sangian, Dendi Wijaya & Mary Dalrymple. 2023. Challenges in Enggano orthography development. *Language Documentation and Description.* 23(1): 4, 1–19. https://doi.org/10.25894/ldd.329.
Kähler, Hans. 1987. *Enggano-Deutsches Wörterbuch* (Veröffentlichungen des Seminars für Indonesische und Südseesprachen der Universität Hamburg 14). Berlin: Dietrich Reimer Verlag.
Katubi. 2007. Lampungic Languages: Looking for New Evidence of Language Shift in Lampung and the Question of Its Reversal. (Ed.) David Mead. *Studies in Philippine Languages and Cultures* 16. 1–10.
Kroeger, Paul. 2014. Passive agents in Malay: The binding properties and discourse functions of agentive *=nya*. In Siaw-Fong Chung & Hiroki Nomoto (eds.), *Current Trends in Malay Linguistics* (NUSA: Linguistic studies of languages in and around Indonesia 57), 5–29.
Kroeger, Paul & Sonja Riesberg. 2024. Voice and transitivity. In Alexander Adelaar & Antoinette Schapper (eds.), The Oxford Guide to the Malayo-Polynesian Languages of Southeast Asia, 795–824. Oxford: Oxford University Press. DOI: https://doi.org/10.1093/oso/9780198807353.003.0047

Lafeber, Abraham. 1922. *Vergelijkende klankleer van het Niasisch*. The Hague: Hadi Poestaka.
Legate, Julie Anne. 2012. Subjects in Acehnese and the nature of the passive. *Language* 88(3). 495–525.
Legate, Julie Anne. 2014. *Voice and v: Lessons from Acehnese*. Cambridge, MA: The MIT Press.
Legate, Julie Anne. 2021. Noncanonical passives: A typology of voices in an impoverished universal grammar. *Annual Review of Linguistics* 7. 157–176.
McDonnell, Bradley. 2016. *Symmetrical voice constructions in Besemah: A usage-based approach*. Santa Barbara: University of California Santa Barbara. (Doctoral dissertation.)
McDonnell, Bradley. 2017. Prospects of a Community Based Language Project in the Nasal Speech Community. In Nicholas Ostler, Vera Ferreira & Chris Moseley (eds.), *Communities in Control: Learning tools and strategies for multilingual endangered language communities*, 64–69. Foundation for Endangered Languages.
McDonnell, Bradley. To appear. What's a passive in Besemah? In Sonja Riesberg & Birgit Hellwig (eds.), *The Empirical Turn – From observable linguistic behaviour to typological generalizations.* Amsterdam: John Benjamins Publishing Company.
McDonnell, Bradley. To appear. Documenting Multilingualism in Southwest Sumatra. In Maya Ravindranath Abtahian, Abigail C Cohn, Thomas J Conners, Joseph J Errington, & Yanti (eds.), *Indonesian Languages and Linguistics: State of the Field*. Ithaca: SEAP Publications, Cornell University Press.
McDonnell, Bradley & Victoria Chen. 2022. The evolution of syntax in Western Austronesian. In Shei, Chris & Saihong Li (eds.), *The Routledge handbook of Asian linguistics*, 11–32. London: Routledge.
McDonnell, Bradley & Christina L Truong. 2024. Applicative constructions in languages of western Indonesia. In Zúñiga, Fernando & Denis Creissels (eds.), *Applicative Constructions in the World's Languages*, 971–1004. Berlin ; Boston: De Gruyter.
McDowell, Jonathan & Karl Anderbeck. 2020. *The Malay Lects of Southern Sumatra* (JSEALS Special Publication 7). Honolulu: University of Hawai'i Press. http://hdl.handle.net/10524/52466.
McGinn, Richard. 1982. *Outline of Rejang syntax* (NUSA). Vol. 14. Jakarta: Badan Penyelenggara Seri Nusa Universitas Katolik Atma Jaya Jakarta.
McGinn, Richard. 1997. Some Irregular Reflexes of Proto-Malayo-Polynesian Vowels in the Rejang Language of Sumatra. *Diachronica* 14(1). 67–107. https://doi.org/10/fhzcx6.
McGinn, Richard. 2000. Where Did the Rejangs Come From? In Marlys Macken (ed.), *Papers from the 10th Annual Meeting of the Southeast Asian Linguistic Society*, 247–262. Tempe, Arizona: Arizona State University.
McGinn, Richard. 2003. Raising of PMP *a in Bukar-Sadong Land Dayak and Rejang. In John Lynch (ed.), *Issues in Austronesian historical phonology* (Pacific Linguistics), vol. 550, 37–54. Canberra: Research School of Pacific and Asian Studies, Australian National University.
McGinn, Richard. 2005. What the Rawas Dialect Reveals about the Linguistic History of Rejang. *Oceanic Linguistics* 44(1). 12–64. https://doi.org/10/cqtjgs.
McGinn, Richard. 2009. Out-of-Borneo subgrouping hypothesis for Rejang: re-weighing the evidence. In K. Alexander Adelaar & Andrew Pawley (eds.), *Austronesian historical linguistics and culture history: a festschrift for Robert Blust* (Pacific Linguistics 601), 411–424. Canberra: Research School of Pacific and Asian Studies, Australian National University.

McGinn, Richard. n.d. A grammar of Rejang. Unpublished ms.
McKinnon, Timothy A. 2011. *The morphophonology and morphosyntax of Kerinci word-shape alternations*. Newark: University of Delaware. (Doctoral dissertation.)
McKinnon, Timothy, Peter Cole & Gabriella Hermon. 2011. Object Agreement and "Pro-Drop" in Kerinci Malay. *Language*. 87(4). 715–750.
McKinnon, Timothy A., Yanti, Peter Cole, and Gabriella Hermon. 2015. Infixation and apophony in Malay: Description and developmental stages. *Linguistik Indonesia* 33(1):1–19.
Morris, Max. 1900. *Die Mentawai-sprache*. Berlin: Conrad Skopnik.
Nothofer, Bernd. 1975. *The reconstruction of Proto-Malayo-Javanic* (Verhandelingen van het Koninklijk Instituut voor Taal-, Land- en Volkenkunde 73). The Hague: Martinus Nijhoff.
Nothofer, Bernd. 1985. The subgrouping of the languages of the Javo-Sumatra hesion: A reconsideration. *Bijdragen tot de Taal-, Land- en Volkenkunde* 141(2/3). 288–302.
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.
Nothofer, Bernd. 1988. A discussion of two Austronesian subgroups: Proto-Malay and Proto-Malayic. In *Rekonstruksi dan cabang-cabang Bahasa Melayu induk [Reconstruction and branching of Proto-Malay]*, 34–58. Kuala Lumpur: Dewan Bahasa dan Pustaka.
Nothofer, Bernd. 1994. The relationship between the languages of the Barrier Islands and the Sulawesi-Philippine languages. In Tom Dutton & Darrell T. Tryon (eds.), *Language contact and change in the Austronesian world* (Trends in Linguistics: Studies and Monographs), vol. 77, 389–409. Berlin: Mouton de Gruyter.
Pampus, Karl-Heinz. 1989. Zur dialektgeographischen Gliederung des Mentawai-Archipels. In Wagner, Wilfried (ed.), *Mentawai: Identität im Wandel auf indonesischen Außeninseln*. Bremen: Übersee-Museum.
Prentice, D. J. & A. Hakim Usman. 1978. Kerinci sound changes and phonotactics. In Wurm, S. A. & L. Carrington (eds.), *Second International Conference on Austronesian Linguistics: Proceedings*, 121–163. Canberra: The Australian National University.
Riesberg, Sonja. 2014. *Symmetrical voice and linking in Western Austronesian languages* (Pacific Linguistics 646). Berlin: De Gruyter Mouton.
Sidwell, Paul. 2005. Acehnese and the Aceh-Chamic Language Family. In Anthony Grant & Paul Sidwell (eds.), *Chamic and Beyond: Studies in Mainland Austronesian Languages* (Pacific Linguistics 569), 211–246. Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, Australian National University.
Sidwell, Paul. 2006. Dating the Separation of Acehnese and Chamic by Etymological Analysis of the Aceh-Chamic Lexicon. *Mon-Khmer Studies* 36. 187–206.
Smith, Alexander D. 2017. The Western Malayo-Polynesian Problem. *Oceanic Linguistics* 56(2). 435–490. https://doi.org/10/gcz48t.
Soravia, Giulio. 2007. *The Alas language (northern Sumatra)*. Bologna: Università di Bologna.
Tadmor, Uri. 2003. Final /a/ mutation: A borrowed areal feature in western Austronesian languages. In John Lynch (ed.), *Issues in Austronesian Historical Phonology*, 15–35. Canberra: Pacific Linguistics.

Tadmor, Uri. 2018. Malay-Indonesian. In Bernard Comrie (ed.), *The world's major languages*. 2nd edn. London: Routledge.
Thurgood, Graham. 1999. *From ancient Cham to modern dialects: two thousand years of language contact and change* (Oceanic Linguistics Special Publication 28). Honolulu: University of Hawai'i Press.
Tulius, Juniator. 2012. Family stories: Oral tradition, memories of the past, and contemporary conflicts over land in Mentawai – Indonesia. Leiden: Leiden University. (Doctoral dissertation.)
van der Tuuk, H. N. [1867] 1971. *A grammar of Toba Batak* (Scott-Kemball, Jeune, trans.) (Koninklijk Instituut voor Taal-, Land- en Volkenkunde Translation Series 13). Berlin: Springer-Science+Business Media.
Voorhoeve, Petrus. 1955. *Critical survey of studies on the languages of Sumatra* (Koninklijk Instituut Voor Taal-, Land- En Volkenkunde Bibliography Series). Vol. 1. The Hague, Netherlands: 's-Gravenhage - Martinus Nijhoff.
Walker, Dale F. 1976. *A grammar of the Lampung language: The Pesisir dialect of Way Lima*. Jakarta: Badan Penyelenggara Seri NUSA.
Weber, Tobias. 2019. Sundaland: Examining the areal typological profile through a diachronic lens. Paper presented at *29th Southeast Asian Linguistics Society (SEALS)*, Tokyo, Japan, 27–29 May.
Willms, Alfred. 1956. Lautliche und syntaktische Untersuchungen über die Mentawai-Sprache. *Afrika und Übersee* 40. 1–72.
Woollams, Geoff. 1996. *A grammar of Karo Batak, Sumatra* (Pacific Linguistics Series C No. 130). Canberra: The Australian National University.
Yanti. 2010. *A reference grammar of Jambi Malay*. Newark: University of Delaware (Doctoral dissertation)
Yanti, Peter Cole, and Gabriella Hermon. To appear. *A Grammar of Jambi Malay*. (Pacific Linguistics 665) Berlin: De Gruyter Mouton.
Yanti, Timothy Mckinnon, Peter Cole & Gabriella Hermon. 2018. The Phonological Basis of Syntactic Change in Kerinci. *Oceanic Linguistics* 57(2). 433–483. https://doi.org/10.1353/ol.2018.0018.
Yanti, Peter Cole, Gabriella Hermon, Uri Tadmor, and Bradley Taylor. 2002-2008 Item "Tajung Raden" in collection "MPI EVA corpora". The Language Archive. https://hdl.handle.net/1839/00-0000-0000-0022-6405-8 (Accessed 2025-10-05)
Yoder, Brendon E. 2014. *Phonological and phonetic aspects of Enggano vowels* (SIL eBook 62). Dallas: SIL International.