Proto-Mentawai

Emma Keith

australian national university

The term “Mentawai”, which has been haphazardly applied to the languages indigenous to the Mentawai island chain off the coast of West Sumatra, is not an endogenous identifier for any of the language groups which it covers. However, due in large part to a lack of adequate primary data, these languages have hitherto been treated as exemplifying a singular “Mentawai” language, accompanied by either the (baseless) insistence that the island chain is linguistically unified, or an addendum regretting the inadequacy of presuming a single “Mentawai” language, and stating that the true status of these languages requires further investigation. This paper aims to resolve this undesirable situation by reconstructing the phonology, phonotactics, and part of the lexicon of the shared proto-language of the Mentawai languages. In doing so, it confirms the status of the Mentawai languages as constituting a small subfamily of closely related but nonetheless distinct languages, comparable to the Batak subfamily on the Sumatran mainland. Sound changes and numerous irregular processes are posited to explain the evolution of Proto-Mentawai from Proto-Sumatran and into its various daughter languages. Finally, the status of the Mentawai (sub)family as a primary branch of the recently described Sumatran family is evaluated; evidence for subgrouping with Enggano and Central Barrier Islands is evaluated in turn, with a potential link to Enggano showing promise on account of shared idiosyncracies in pronominal indexing.

*Note: This is the version of the paper submitted to Oceanic Linguistics on 8/08/2025. It has not yet undergone peer review. I am interested in comments in the meantime, which can be directed to* *emma.keith@anu.edu.au**. To cite this paper in its current form treat it as an unpublished manuscript.*

**Keywords:** Mentawai, Sumatran, reconstruction, proto-language

Introduction.[[1]](#footnote-1)

 The name “Mentawai” has long been observed to be a term of convenience. It groups together a series of geographically and culturally linked language groups who, though unified by a shared identity in a modern context, acted as separate and distinct groups prior to the colonial era (Reeves 1999). Accordingly, the languages of these groups differ to about the extent that we would expect of a small, isolated subfamily: there is clear evidence of relatedness and a shared proto-language, but one which is reflected in its various daughter languages through numerous sound changes and lexical innovations. This paper attempts a reconstruction of that proto-language, including its phonology, phonotactics, and a substantial lexicon, on the basis of the modern-day Mentawai languages.

Section 2 begins with a summary of the languages used in the reconstruction: their prior attestations, phoneme inventories, and geographic extents. Section 3 then describes how the reconstruction of Proto-Mentawai (PMEN) is created and evidenced on the basis of regular and irregular sound change, and what the application of these changes tells us about how the Mentawai languages split from PMEN diachronically. The pattern of splitting revealed in this way suggests settlement of the Mentawais first in North or Northwest Siberut, where the greatest linguistic diversity lies, then rapid and recent expansion Southward to cover all other inhabited areas of the archipelago. Section 4 discusses the status of the Mentawai subfamily as a primary branch of the Sumatran family, in dialogue with the recent clarification of that family’s membership and diagnostic sound changes in Billings and McDonnell (2024). Numerous sound changes, regular and irregular, are shown to occur between Proto-Sumatran (PSUM) and PMEN. Some of these are shared with Enggano and Proto-Central Barrier Islands; accordingly the potential for subgrouping with each of these languages is explored, and pronominal evidence for Enggano tentatively tabled in favour of a connecting node below PSUM.

Mentawai as a language family.

 The name ‘Mentawai’, and its various derivative forms, gained traction firstly as a colonial exonym, then as a discursive tool which has enabled the various people of the Mentawai islands to gain recognition, particularly from the inhabitants of mainland Sumatra, as a distinct and meaningful group (Reeves 1999). The term seems to originate from a Sumatran exonym ‘*mantawei*’ used to identify Pagai islanders (Crisp 1799), itself deriving from the Pagai Mentawai word for ‘man’, *manteu*, which was then gradually picked up by colonial administrators and mainland Sumatrans alike to refer to the peoples of the four ‘Mentawai’ islands (Siberut, Sipora, and North and South Pagai) as a group, in lieu of an existing term with which to do so (Reeves 1999). This is not to imply that the use and spread of the term in more recent years has not been driven by an endogenous desire for a pan-archipelegic identifier, quite the opposite: the term has been used strategically by Mentawai intellectuals over the latter half of the 20th century in the fight for political and legal recognition from the Indonesian nation-state (Eindhoven 2007). This notion of ‘Mentawai’ coherence, distinctiveness, and unity as a group was a key factor in the 1999 decision to grant regency status to the Mentawai islands, leading to a greater level of self-determination in policymaking and, more intangibly, a bolstering of this sense of Mentawai cultural autonomy from the rest of Indonesia (Samaloisa 2020).

For better or worse, concomitant with the strategically deployed discourses of ‘Mentawai’ as a unified group for the purposes of political bargaining, discourses on the topic of the Mentawai islands have similarly frequently referred to a ‘Mentawai language’, presumed to be coextensive with this political, cultural, and geographic group. This notion of a unified ‘Mentawai’ language, however, is shown to be illusory even on the grounds of a simple test of mutual intelligibility. Previous surveys of the Mentawai languages which have been more realistic in this regard (Budiono, Novita and Syarfina 2021; Nothofer 1986) have tended to rely on lexicostatistics to determine which Mentawai varieties may appropriately be called ‘languages’ and which are more accurately ‘dialects’; this is an orthogonal question to all those which I am attempting to answer. I do not make use of lexicostatistics in my reconstruction, primarily because Sipora Mentawai has exerted a massively oversized influence on other Mentawai languages in the recent past, and so any lexicostatistical comparison would have to deal with the extremely thorny problem of identifying, without the aid of sound change, all such subfamily-internal loans. There are also legitimate reasons to doubt the use of lexicostatics for reconstruction in the Austronesian area on principle (Blust 2000; Smith 2024*b*).

language distributions.

 The following subsection provides approximate geographic descriptions of the distributions of the Mentawai languages referred to in later sections of the paper.

Sipora Mentawai.

Sipora Mentawai is by far the most widely spoken Mentawai language; all other Mentawai languages are endangered to varying degrees by language shift to Sipora Mentawai, as well as to Indonesian and Minang. Almost all people born this century across the island chain have some degree of passive competency in Sipora Mentawai, aside from those few who have switched to Indonesian and/or Minang entirely. The geographic extent of Sipora Mentawai is self-explanatory: it is the language spoken on Sipora island. The Mentawai of the Pagai islands, also referred to as Sikakap or Sakalagan/Sakalagat Mentawai[[2]](#footnote-2), is mutually intelligible and not separated by any sound changes; these should be considered one language with at most some minor lexical differences. ‘Sipora’ Mentawai could perhaps also be called ‘Sipora-Pagai’ Mentawai, but since the data in this paper all comes from Sipora and there is some limited lexical variation this is the term used here.[[3]](#footnote-3)

Sabirut Mentawai.

Sabirut Mentawai is perhaps the next most widespread Mentawai language after Sipora. It is spoken in the coastal regions of South and Southwest Siberut districts, including Maileppet, which is the largest town on Siberut, in its sister villages Muarasiberut and Muntei, as far inland as Rokdog and around the Southern tip to Taileleu. It is, by the standards of the divergence of the Mentawai languages, extremely similar to Sipora Mentawai: it is mutually intelligible, though a few lexical innovations separate the two. It is not separated from Sipora Mentawai by any sound changes. Tulius (2012*b*), a native speaker, notes that Morris (1900) more closely reflects Sabirut Mentawai despite apparently taking its data from Sipora. For a more contemporary source, Gil (2007) provides a corpus of ‘Muarasiberut’ Mentawai, which represents the same variety with a more localized name. I include it here mostly for its distinctness of speaker identity: anecdotally, though individuals from Sipora and Pagai often refer to each other as speaking ‘the same’ language, the same is seldom of Sipora and Sabirut.

Rereiket Mentawai.

Rereiket Mentawai is spoken in the inland regions of South Siberut district up to the border with Central Siberut district, in the regions of Buttui, Matotonan, and Madobag. Rereiket Mentawai is so far only treated in the recent dictionary published by Mendrofa, Salakkirat, and Henry (2019), which is fairly extensive, and in ongoing documentation work by the authors of that dictionary. Reimar Schefold has also conducted fieldwork in the area. It is in extensive contact with Sabirut Mentawai, and some language shift to Sabirut seems to be ongoing among younger speakers in downriver settlements such as Madobag, though it remains strong further inland. It has a fair degree of mutual intelligibility with Sabirut Mentawai, and some limited mutual intelligibility with Sipora Mentawai.

Sikabaluan Mentawai.

Sikabaluan Mentawai is spoken over a relatively wide and diverse area encompassing the whole of Central Siberut district, including Saibi, and the towns of Sikabaluan, Mongan Poula, and Sirilogui in coastal North Siberut regency. It could probably sensibly be called something like ‘Coastal Eastern Siberut’, but Sikabaluan seems to be the dominant endonym and is therefore used here, as in Cambielli (1998), in spite of the language’s geographic spread outside of Sikabaluan proper. It is not mutually intelligible with the aforementioned languages. It is best attested in Cambielli (1998), as well as in a wordlist used in the creation of Budiono, Novita and Syarfina (2021).[[4]](#footnote-4)

Simatalu Mentawai.

Simatalu Mentawai is spoken in and around the settlement of Simatalu, in West Siberut district. It is attested to my knowledge only in Cambielli (1998). Juniator Tulius and Reimar Schefold both seem to have done fieldwork in the area, but the primary sources from this are not available to me. Notably, Simatalu is widely understood in folktales as being the geographic origin of the Mentawai migrations (Schefold 1989; Tulius 2012*a*, 2012*b*).

Simalegi Mentawai.

Simalegi Mentawai is spoken in and around the remote hamlet of Simalegi, also in West Siberut district. It is only attested in Cambielli (1998) and the dialect surveys by Pampus (1989, 1994).

Terekan Mentawai.

Terekan Mentawai is spoken in the settlement of Terekan, on the Northern tip of Siberut island (North Siberut district). The extent of its distribution outside of Terekan proper is unknown. Terekan is a very isolated settlement; Cambielli (1998) provides the only substantial attestation. It is not separated from Sikabaluan Mentawai by any sound changes, and may be best understood as a part of Sikabaluan Mentawai itself; poor attestation at current prevents me from saying anything confident on this matter.

Paipajet Mentawai.

Paipajet Mentawai is spoken in the settlement of Paipajet, in coastal West Siberut district. It is attested in Cambielli (1998); Karl-Heinz Pampus has apparently conducted some fieldwork on this language (cited as p.c. in Blust 1997), but primary sources from this fieldwork are unavailable to me.

phonologies of mentawai languages.

 The following section describes the phonological systems of all of the above languages.

Vowel Inventories.

Since all of the Mentawai languages have the same five-vowel system, this is briefly described below.

table 1. shared vowel inventory of all mentawai languages

|  |  |  |  |
| --- | --- | --- | --- |
|  | Front | Central | Back |
| **High** | i |  | u |
| **Low-Mid** | ɛ |  | ɔ |
| **Low** |  | a |  |

Vowels do not reduce to schwa. For reasons that will become apparent in the reconstruction, I treat diphthongs diachronically as resulting from strings of vowel phonemes rather than being phonemic in their own right. Each diphthong is therefore interpreted as a vowel plus a semivowel created by the reduction of a corresponding vowel that loses the feature [+syl] according to phonotactic contraints. – namely, that the nucleus of a syllable may only contain one vowel, so any vowel phoneme in coda position loses its syllabic status. When desyllabified, /i/ becomes [j] and /u/ becomes [w]. When appearing before semivowels, both of which are high, the low-mid vowels move up to high-mid position.

table 2. diphthong sequences and their underlying vowel phonemes

|  |  |
| --- | --- |
| Dipthong | Constituent Phonemes |
| iw | i, u |
| ej | ɛ, i |
| ew | ɛ, u |
| aj | a, i |
| aw | a, u |
| oj | ɔ, i |
| ow | ɔ, u |

Syllable Structure.

In all Mentawai languages, all consonants can occur in the syllable onset. The syllable coda may be filled by any non-continuant consonant, and also not by any palatal consonant, a common restriction among Sumatran languages. If a vowel fills the coda then it is desyllabified to a semivowel. The glottal stop may come after any vowel, including coda vowels which have undergone desyllabification. That the semivowel truly does occupy the coda position in these cases is shown by the fact that no true consonants may come after the semivowel coda. The glottal stop must therefore be treated as more of an articulatory gesture: is the cessation of a vowel or semivowel segment achieved by closure of the glottis or mere cessation of voicing? – than a segment in its own right. This quasi-phonemic status is also demonstrated by the glottal stop’s apparent ‘invisibility’ to sound change environment rules (e.g. 3.1.5). It is still included in the phoneme inventories here due to its conditioned allophony with [k], but it should be viewed as a quasi-phoneme at best.

figure 1. the structure of a mentawai syllable

|  |
| --- |
| σ |
| ω | ρ |
|  | ν | κ |
| (C) | V | (C[-cont][-pal]) |

table 3. syllable types in Mentawai languages

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Structure** | **Glottal V release?** | **Secondary articulation?** | **Example** | **Language** | **Gloss** |
| V | No | No | [**ˈɔ.**tu] | Sipora | ‘hundred’ |
| CV | No | No | [**ˈbɔ**.ɟɛ] | Terekan | ‘brown’ |
| Yes | No | [ˈlej.**tiʔ**] | Simatalu | ‘snake’ |
| No | Yes | [kwa] | Simalegi | ‘say’ |
| Yes | Yes | [djɔʔ] | Simalegi | ‘stand’ |
| VC | No | No | [**ˈap**.pɛk] | Sikabaluan | ‘reap’ |
| CVC | No | No | [ˈu.**ban**] | Sabirut | ‘grey hair’ |
| [ˈsa.**baw**] | Sipora | ‘pass’ |
| Yes | No | [ˈŋajʔ] | Sabirut | ‘sand’ |
| No | Yes | [bwat] | Simalegi | ‘uncle’ |

Allophonic variation.

There are a few features of allophony which are shared throughout the Mentawai languages, and therefore reconstructed for PMEN:

1. The phoneme /ɡ/ has the allophones [ɰ] and [ɤ] intervocalically. Similarly, the phoneme /b/ has the allophones [β] and [β̞] in the same environment. These patterns of allophony can be generalized as a process of spreading of the feature [+cont] into voiced non-coronal plosives from adjacent vowels, which is blocked for voiceless plosives. Since the feature [+str] is underspecified throughout the inventory, both fricative and approximant realizations are freely available.
2. The phoneme /s/ has the allophone [c] after coronal consonants. For some speakers, particularly younger speakers, the cluster [tc] becomes [t͡ʃ], presumably under the influence of Indonesian, which has contributed numerous loan items which have [t͡ʃ] in syllable onsets, thereby ‘phonemicizing’ the cluster [tc] as /͡tc~t͡ʃ/. The native vocabulary, however, only allows [c] as an allophone of /s/, and does not treat the cluster [tc] as a phoneme, since it only appears at syllable boundaries where /t/ is the coda of one syllable and /s/ the onset of another.
3. If /n/ appears before a palatal consonant, it assimilates to the palatal place as [ɲ]. This includes [c] (allophone of /s/).
4. The quasi-phoneme /ʔ/ has the allophone [k] when it occurs in the onset of a syllable through morphological processes.
5. [r] is an allophone of /d/ word-initially. [r]- and [d]-initial realizations of words beginning with /d/ are in free variation.

/t/ tends to be phonetically dentalized by default, as is common in Austronesian languages. It is phonemically interpreted as the alternant of the alveolar consonant [d], and is never contrastive with a dental equivalent /t̪/.

Suprasegmentals.

In all Mentawai languages: stress falls on the penultimate syllable of the word (much like PAN, Smith 2023*b*, but without the exception for schwa since schwa is not present). Plosives in syllable codas have no audible release, as is common in Austronesian languages. Uniquely, voicing is non-constrastive for plosives in syllable-final position: all syllable-final plosives lack voicing, but evidence from diachrony and writing suggests that these are perceived in the phonology to be underspecified for voice rather than being voiceless phonemes per se.[[5]](#footnote-5) Much like PMP and PAN (Smith 2023*a*), the vast majority of open-class lexemes are disyllabic, with a few trisyllabic exceptions; quadrisyllabic[[6]](#footnote-6) and monosyllabic lexemes are also possible in some edge cases. Like in PAN (Blust 2013), closed classes may freely have monosyllabic items.

The following subsections detail language-specific sections of the phonetics and phonology.

Sipora and Sabirut Mentawai.

table 4. consonant inventory of sipora and sabirut mentawai

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | [+/-voice] | Bilabial | Alveolar | Palatal | Velar | Glottal |
| **Stop** | - | p | t |  | k | (ʔ) |
| + | b | d | ɟ | ɡ |  |
| **Nasal** | + | m | n | ɲ | ŋ |  |
| **Fricative** | - |  | s |  |  |  |
| **Trill** | + |  | r |  |  |  |
| **Lateral Approximant** | + |  | l |  |  |  |

Several aspects of allophonic variation are worth mentioning:

1. Some[[7]](#footnote-7) word-final voiceless plosives are in free variation with (voiced) nasals at the same place of articulation for some speakers - e.g., [ˈlu.lak~ˈlu.laŋ] *lulak~lulang* ‘wooden bowl’. When these voiceless stops appear as the onset of a word-internal syllable as the result of affixation, they are realized by some speakers as stop-nasal clusters, with only the latter segment bearing voicing. A similar phenomenon seems to also result from a sandhi effect in rapid speech, but never in careful speech. Meanwhile, if these plosives become the onset of a syllable through affixation, they become nasals.

table 5. allophony affecting erstwhile preploded nasals in sipora and sabirut mentawai

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Phoneme** | **Syllable position** | **Allophone** | **Example** | **Gloss** |
| /k/ | κ | [k] | ˈbu.ɟuk | ‘lie down’ |
|  | ω | [kŋ] | bu.ɟukˈŋa.kɛʔ | ‘make lie down’lie.down-CAUS |
|  | ω\_κ[+nas] | [ŋ] | a.ma.buˈɟu.ŋan | ‘already lying down’REAL-DUR.ITR-lie.down-PERF |

1. Syllable-final nasals are in free variation with stops at their respective places of articulation for some speakers, assimilating to the voicing[[8]](#footnote-8) of the following segment – e.g. [ˈman.tew~ˈmat.tew] *manteu~matteu* ‘man’; [ˈnɛn.da~ˈnɛd.da] *nenda~nedda* DEM.PROX.
2. Plosives may geminate syllable-finally, ‘copying’ the value of the plosive into the onset of the next syllable. Geminate syllable-final plosives may cluster with liquids – e.g. [ˈsuk.kra] *sukkra* ‘dig’. In these cases, the geminate plosive remains within the coda of the former syllable phonemically – i.e., [ˈsuk.kra] is /ˈsukː.ra/ - this is demonstrated by the fact that triconsonantal clusters are entirely illegal aside from this specific geminate stop-liquid combination.

Rereiket Mentawai.

Rereiket Mentawai has a very similar inventory to Sipora and Sabirut, with two major differences: the presence of voiced dental stop /d̪/[[9]](#footnote-9) and the absence of palatal nasal /ɲ/. The former of these adds a place of articulation which is not present in Sipora or Sabirut. All of the patterns of allophonic variation touched on for Sabirut and Sipora Mentawai also apply in Rereiket Mentawai.

table 6. consonant inventory of Rereiket mentawai

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | [+/-voice] | Bilabial | Dental | Alveolar | Palatal | Velar | Glottal |
| **Stop** | - | p |  | t |  | k | (ʔ) |
| + | b | d̪ | d | ɟ | ɡ |  |
| **Nasal** | + | m |  | n |  | ŋ |  |
| **Fricative** | - |  |  | s |  |  |  |
| **Trill** | + |  |  | r |  |  |  |
| **Lateral Approximant** | + |  |  | l |  |  |  |

All of the remaining Mentawai languages except Simalegi Mentawai share this same phoneme inventory as above, which for reasons of space is not repeated.

Paipajet Mentawai.

Paipajet Mentwai is identical to Rereiket Mentawai in terms of phoneme inventory and allophony.[[10]](#footnote-10)

Sikabaluan Mentawai.

Sikabaluan Mentawai has the same inventory as above. Some key differences must be noted, however,

1. All word-final nasals appear as stop-nasal clusters when they appear, through affixation, as the onset of a syllable whose coda is not a nasal.

table 7. allophony affecting erstwhile preploded nasals in sikabaluan mentawai

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Phoneme** | **Syllable position** | **Allophone** | **Example** | **Gloss** |
| /m/ | κ | [m] | ˈpɛ.rɛm | ‘sleep’[[11]](#footnote-11) |
| ω | [pm] | pɛ.rɛpˈma.kɛʔ | ‘put to sleep’sleep-CAUS |
| ω\_κ[+nas] | [m] | pɛˈrɛ.man | ‘bed’sleep-LOC.NMLZ |

1. There is no free variation between word- or syllable-final nasals and voiceless stops; Sikabaluan Mentawai maintains nasals in these positions. Stop-nasal clusters, however, arise under the same conditions as described in 2.2.5.
2. Prenasalized plosives may occupy the syllable coda, e.g. [ˈsuŋk.ra] *sungkra* ‘dig’ – the phonemic status of these plosives is the same as for the corresponding geminate plosives in other languages.

Terekan Mentawai.

Terekan Mentwai is identical in phonology to Sikabaluan Mentawai.

Simatalu Mentawai.

Simatalu Mentwai has the same inventory as Rereiket Mentawai. Additionally:

1. /k/ has the allophone [x] intervocalically.

Unlike the continuant allophones of /b/ and /ɡ/, the allophone [x] does not exist in free variation with the elsewhere form; /k/ is always [x] in this environment. Accordingly, there is no change to the phoneme inventory, though /k/ could probably be re-written as /k~x/, since the intervocalic environment is so common.

Simalegi Mentawai.

The inventory of Simalegi Mentawai is as below; it is the largest and the only to include both /ɲ/ and /d̪/.

table 8. consonant inventory of simalegi mentawai

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | [+/-voice] | Bilabial | Dental | Alveolar | Palatal | Velar | Glottal |
| **Stop** | - | p |  | t |  | k | (ʔ) |
| + | b | d̪ | d | ɟ | ɡ |  |
| **Nasal** | + | m |  | n | ɲ | ŋ |  |
| **Fricative** | - |  |  | s |  |  |  |
| + |  |  | (z) |  |  |  |
| **Trill** | + |  |  | r |  |  |  |
| **Lateral Approximant** | + |  |  | l |  |  |  |

The patterns of allophony are the same as in Sipora and Sabirut Mentawai. Simalegi Mentawai permits plosives in the syllable onset to add secondary features: palatalization reflecting diachronic front vowels and labialization reflecting diachronic back vowels.

Reconstructing Proto-Mentawai.

 The following section reconstructs Proto-Mentawai on the basis of a number of regular sound changes and numerous other irregular phenomena affecting the forms of lexemes.

sound change evidence.

 The substantial majority of all sound changes within the Mentawai languages occur between at the first order of branching affect only Simalegi Mentawai. While it is tempting to posit that on these grounds the original settling-place of the first Mentawaians was Simalegi, not Simatalu as in legend (cf. Schefold 1986, 1989; Tulius 2012*a*, 2012*b*), I am uncomfortable doing so for two reasons: 1) there is no strong evidence that Simalegi Mentawai represents a primary branch of PMEN; 2) Simalegi and Simatalu are geographically proximal (only a little over 20km apart, and connected by a beach), and it is reasonable to believe that the speaker populations might have moved around in the unknown timeframe since the original settlement took place; 3) the sheer number of changes affecting Simalegi Mentawai as compared to all other branches, and the fact that at least one (ɟ>z/V\_V) seems to have occurred this century, suggests that perhaps, for some unknown reason[[12]](#footnote-12), Simalegi Mentawai is just uniquely prone to sound change. For reasons of space only examples from a few relevant daughter languages and their proto-forms are cited throughout this section; a fuller Proto-Mentawai lexicon which provides the daughter forms motivating reconstructions is provided in Appendix A.

I also reconstruct several protolanguages which are themselves daughters of PMEN; these are named for the approximate location where they are likely to have been spoken. Simalegi, Terekan, and Sikabaluan Mentawai are all primary branches of PMEN, with all other languages being daughters of a shared ancestor Proto-Central Siberut; this splits into Proto-West Siberut and Proto-South Siberut, with a final split between PSS into Rereiket Mentawai and a common ancestor of Sipora and Sabirut Mentawai, Proto-Sipora-Sabirut.

This splitting pattern suggests initial settlement somewhere in North or Northwest Siberut at a fair time-depth, followed by quite recent and rapid expansion Southward, which is in keeping with expectations established by folklore.

table 9. sound changes affecting Mentawai languages

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| PMEN |  | PCS | \*ew>\*ow/cf. 3.1.4 | PSS | \*d̪>\*t | PSiSa | Sipora Mentawai |
| Sabirut Mentawai |
|  | Rereiket Mentawai |
| PWS |  | Paipajet Mentawai |
| \*k>x/V\_V | Simatalu Mentawai |
| Sikabaluan Mentawai |
| Terekan Mentawai |
| \*ɟ>d̪/V[+high]\_\*ɡ>ɟ/\_V[+front]\*ɟ>z/\_V\*s>\*t\*t>s/\_i\*k>s/V[+front]\*ɡ>Ø/V[+high]\_V\*ŋ>ɲ/\_e\*n>ɲ/\_i | Simalegi Mentawai |

Reduplicative versus Non-reduplicative Lemmas.

Lemmas in Mentawai languages may be either reduplicative or non-reduplicative. Reduplicative lemmas are created by the reduplication of a -CVC syllable and are always bisyllabic. The reduplication of monosyllables as a method of creating lemmas is well observed for PAN as well as in many daughter languages (Blust 1976; Chrétien 1965; Dempwolff 1934-8). This requires note because, for phonological and phonotactic purposes, reduplicative lemmas appear to treat the whole lemma as simply the reduplication of the latter syllable, which is ‘copied’ backwards into the former syllable. This can be observed through several patterns:

1. Sound changes which refer to a ‘word-final’ environment (4.2.1[[13]](#footnote-13), 4.2.5) also apply to the boundary of a monosyllable which is reduplicated to create a lemma.
2. Sound changes which specify some environment which references the value of a preceding segment (3.1.5, 3.1.16) will apply the change to both the latter syllable **and**its ‘copied’ preceding syllable, despite the fact that the copied syllable may place the relevant segment into what seems to be ‘word-initial’ position.
3. Clusters which may not occur within non-reduplicative syllables may occur within reduplicative ‘lemmas’, since any syllable which follows the rules in (2.2.2) may be ‘copied’ to form a reduplicative lemma without having to adhere to phonotactic constraints otherwise affecting lemmas.

This presents a rather intriguing challenge to our notion of ‘morphemehood’ for the open classes, which may be the inappropriate unit when dealing with Austronesian languages (Blust 1988): the latter syllable in such lemmas seems to be sensitive to the presence of a preceding syllable which is a copy of itself, but which is invisible to the phonotactics. We can get around this by positing that the phonological rule requiring open-class lemmas be (at least) bisyllabic, and phonological rules (including sound change, which is really just the diachronic alteration of phonological rules) are applied *after* the phonotactic rules affecting clusters, which here have no effect.

figure 2. phonotactic and phonological rules in reduplicative lemmas

|  |  |  |
| --- | --- | --- |
| (1) | Phonotactics: maximal syllable structure | σ*α* |
| (2) | Phonotactics: clustering rules (no effect) | σ*α* |
| (3) | Phonology: open-class lemmas before diachronic change must have 2+ syllables | σ*α*σ*α* |
| (4) | Phonology: application of phonological rules (including sound change)  | (C)1V1(C)2(C)1V1(C)2 |
| σ*α*σ*β* |
| (5) | Phonology: both syllables in reduplicative lemma must be identical | σ*β*σ*β* |

\*d̪>\*t (Proto-South Siberut>Proto-Sipora-Sabirut).

This change replaces the dental consonant \*d̪ in Sipora and Sabirut by means of devoicing, since \*t is still phonetically dental. The natural push towards symmetry of inventory was likely a motivator here.

table 10. evidence for \*d̪>\*t

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| PMEN | PSS | Rereiket | PSiSa | Sabirut | Sipora | Gloss |
| \*ɡid̪ik | \*ɡid̪iʔ | ɡid̪iʔ | \*ɡitiʔ |  | gitiʔ | ‘tickle’ |
|  | \*d̪ɛbbuk | d̪ɛbbuk | \*tɛbbuk | tɛbbuk | tɛbbuk | ‘float’ |

\*k>x/V\_V (Proto-West Siberut>Simatalu Mentawai).

This change likely originated from the spreading of [+cont] from adjacent vowels as attested for voiced /ɡ/ in other Mentawai languages, which became regularized unlike the effects of the phenomenon elsewhere. The change is not seen in any other branches, either of Paipajet from PWS or in PSS to the South. This sound change is more of a change in the phonological system than a sound change in the traditional sense, since it only creates an allophone rather than a new phoneme.

table 11. evidence for \*k>x/V\_V

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| PMEN | PWS | Simatalu | Paipajet | PSS | Gloss |
| \*bakɛ | \*bakɛ | baxɛ |  | \*bakɛ | ‘thread’ |
| \*bɔku | \*bɔku | bɔxu | bɔku | \*bɔku | ‘shoulder’ |
| \*lujkuʔ | \*lujkuʔ | lujxuʔ | lujkuʔ | \*lujkuʔ | ‘seed’ |
| \*saki | \*saki | saxi | saki | \*saki | ‘buy’ |
| \*tɛkaʔ | \*tɛkaʔ | tɛxaʔ | tɛkaʔ | \*tekaʔ | ‘caw (of chicken)’ |

\*ew>\*ow/unless ɛ…\_ (Proto-Mentawai>Proto-Central Siberut).

This sound change is observed in all daughter languages of PCS. The diphthong \*ew changes to an \*ow, unless the nucleus of the prior syllable is /ɛ/. Monosyllables seem not to obey this pattern.

table 12. evidence for \*ew>\*ow/ɛ…\_

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| PMEN | Simalegi | Sikabaluan | PCS | PSS | Sipora | Gloss |
| Sound change applied: \*ew>\*ow  |
| \*bilew | bilew | bilew | \*bilow | \*bilow | bilow | Kloss’s gibbon (*hylobates klossii*) |
| \*kalipew[[14]](#footnote-14) | kalipew | kalipew | \*kalipow | \*kalipow | kalipow | ‘forget’ |
| \*kasew | katew | kasew | \*kasow | \*kasow | kasow | ‘rafter’ |
| \*newnew | newnew | newnew | \*nownow | \*nownow | nownow | ‘propose’ |
| \*pewla | pewla | pewla | \*powla | \*powla | powla | ‘nipah palm’(*nypa frucitans*) |
| \*pusew | putew | pusew | \*pusow | \*pusow | pusow | ‘navel’ |
| \*sapew | tapew | sapew | \*sapow | \*sapow | sapow | ‘field hut’ |
| \*rewrew | rewrew | rewrew | \*rowrow | \*rowrow | rowrow | ‘hunt’ |
| \*unew | unew | unew | \*unow | \*unow | unow | ‘pith’ |
| \*ulew |  |  | \*bilow | \*ulow | ulow | ‘snake’ |
| Sound change blocked: \*ew/ɛ…\_ |
| \*ɛkew | ɛsew | \*ɛkew | ɛkew | \*ɛkew | ɛkew | 2SG |
| \*lɛɡew | lɛzew | \*lɛɡew | lɛɡew | \*lɛgew | lɛɡɡew | ‘dry season’ |
| \*lɛlew | lɛlew | \*lɛlew | lɛlew | \*lɛlew | lɛlew | ‘forested hill’[[15]](#footnote-15) |
| \*rɛŋɡew | rɛggew | \*rɛŋɡew | rɛŋɡew | \*rɛŋɡew | rɛggew | ‘narrow’ |
| Sound change blocked: Monosyllables |
| \*abew | abew | \*abew | abew | \*abew | abew | ‘large’[[16]](#footnote-16) |
| \*arew | arew | \*arew | arew | \*arew | arew | ‘far’ |
| \*sewʔ | tewʔ | \*sewʔ | sewʔ | \*sewʔ | sewʔ | ‘cook’ |

All of the remaining sound changes all occur at the level of PMEN>Simalegi Mentawai. Accordingly, the diachronic ordering of these sound changes is now discussed where relevant; the ordering of the above changes is taken to be demonstrated by the order of branching.

\*ɟ>d̪/V[-high]\_ (Proto-Mentawai>Simalegi Mentawai).

This sound change swaps one very uncommon phoneme in PMEN for another. The originality of \*ɟ is proven by cognates with \*j in PSUM (having applied sound change 4.2.12).

table 13. evidence for \*ɟ>d̪/V[-high]\_

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| PSUM | PMEN | Simalegi | PCS | Rereiket | Sipora | Gloss |
| Sound change applied: \*ɟ>d̪/V[-high]\_ |
| \*bajaw | \*baɟaʔ | bad̪aʔ | \*baɟaʔ | baɟaʔ | baɟaʔ | ‘elder sibling’ |
| \*lajaR | \*laɟɔ | lad̪ɔ | \*laɟɔ | laɟɔ | laɟɔ | ‘sail’ |
| \*laju | \*laɟuʔ | lad̪uʔ | \*laɟuʔ | laɟuʔ | laɟuʔ | ‘wither’ |
|  | \*ɟɔʔɟɔʔ | d̪ɔʔd̪ɔʔ | \*ɟɔʔɟɔʔ |  | ɟɔʔɟɔʔ | ‘dog’ |
|  | \*laɟɛ | lad̪ɛ | \*laɟɛ | laɟɛ | laɟɛ | ‘hungry’ |
|  | \*ɔɟuʔ | ɔd̪uʔ | \*ɔɟuʔ | ɔɟuʔ | ɔɟuʔ | ‘tide’ |
| Sound change blocked: \*ɟ/V[+high]\_ |
|  | \*ɟiʔɟiʔ | ziʔziʔ | \*ɟiʔɟiʔ | ɟiʔɟiʔ | ɟiʔɟiʔ | ‘baby girl’ |
|  | \*ɟajɟaj | zajzaj | \*ɟajɟaj | ɟajɟaj | ɟajɟaj | ‘similar’ |

\*ɡ>ɟ/V[+front] (Proto-Mentawai>Simalegi Mentawai).

This sound change would have particular motivation if some degree of palatalization persisted in PSUM and PMEN \*ɡ. The originality of \*ɡ is proven by cognates in PSUM with the same phoneme. Modern Simalegi Mentawai further shifts /ɟ/ to /z/ intervocalically, creating a chain shift. However, the intermediary stage of this chain shift is evidenced by a) forms given in Cambielli (1998), who provides /ɟ/ (orthographic <j>) for all forms below which in modern Simalegi Mentawai have /z/; and b) the preservation of /ɟ/ from this shift in word-initial position (cf. 3.1.7). Needless to say, this change must have happened first.

table 14. evidence for \*ɡ>ɟ(>z)/\_V[+front]

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| PSUM | PMEN | Simalegi | PCS | Simatalu | Sipora | Gloss |
| Sound change applied: \*ɡ>ɟ(>z)/V[+front] |
| \*siɡəm[[17]](#footnote-17) | \*siɡɛpm | sizɛp | \*siɡɛpm | siɡɛp | siɡɛp | ‘ant’ |
|  | \*baɡej | bazej | \*baɡej | baɡej | baɡej | ‘other’ |
|  | \*ɡɛla | zɛla | \*ɡɛla | ɡɛla | ɡɛla | ‘tired’ |
|  | \*ɡowɡiw | ɡowziw | \*ɡowɡiw | ɡowɡiw |  | ‘crab shell’ |
|  | \*lɛɡew | lɛzew | \*lɛɡew | lɛɡew | lɛɡew | ‘dry season’ |
| Sound change blocked: \*ɡ/V[-front] |
| \*[a]luɡa | \*luɡa | luɡa | \*luɡa |  | luɡa | ‘oar’ |
|  | \*ɡɔʔ | ɡɔʔ | \*ɡɔʔ | ɡɔʔ | ɡɔʔ | ‘swollen’ |

ɟ>z/V\_V (Proto-Mentawai>Simalegi: very recent).

This change is quite common cross-linguistically, and easy to justify: \*ɟ changes from stop to fricative; palatal stops in general have a tendency to fricate, meaning this change in practice only involves the loss of complete closure at the beginning of what is in many cases an affricate realization. Proof of environment is shown by the fact that word-initial \*ɟ, as created by 3.1.6 in one item (since it does not appear in PMEN), is unaffected. Simalegi Mentawai *zɛla* appears at first to be an exception, but given that this lemma is a stative intransitive verb, which must in all cases[[18]](#footnote-18) appear with stative *ma-*, creating a word-internal environment for \*ɟ.[[19]](#footnote-19)

table 15. evidence (by comparison) for ɟ>z/V\_V

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| PMEN | Simalegi | PCS | Simatalu | Gloss |
| \*ɡid̪aʔ | ɟilaʔ | \*ɡilaʔ | ɡilaʔ | ‘tickle’ |

\*s>\*t (Proto-Mentawai>Simalegi Mentawai).

This sound change, noted by Pampus (1994), is fairly intuive from an articulatory standpoint: the manner changes from fricative to plosive. Considering that /s/ is the only fricative in the inventory of PMEN, this is unremarkable and likely motivated, once again, by the draw towards symmetry. More noteworthy is the fact that \*t also subsequently changed to /s/ after /i/ (3.1.9), making it so that words which have \*s in this environment in PMEN also reflect it in Simalegi Mentawai despite the two intermediary sound changes. Since /s/ is a continuant and cannot appear in syllable codas, the former change has no specific conditioning environment.

table 16. evidence for \*s>\*t

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| PSUM | PMEN | Simalegi | PCS | Rereiket | Sipora | Gloss |
| Sound change applied: \*s>\*t |  |
| \*asah | \*asa | ata | \*asa | asa | asa | ‘whet’ |
| \*pusəɡ | \*pusew | putew | \*pusew | pusow | pusow | ‘navel’ |
| \*sawa | \*saba | taba | \*saba | saba | saba | ‘python’ |
| \*saɡu | \*saɡu | taɡu | \*saɡu | saɡu | saɡu | ‘sago’ |
| \*səraw | \*sow | tow | \*sow | sow | sow | ‘cry’ |
| \*suluh | \*sulu | tulu | \*sulu | sulu | sulu | ‘sun’ |
| There and back again: \*s(>\*t)>s/\_i |
| \*siwa | \*siba | siba | \*siba | siba | siba | ‘nine’ |
| \*siɡɛm | \*siɡɛpm | sizɛp | \*siɡɛpm | siɡɛp | siɡɛp | ‘ant’ |

\*t>s/\_i (Proto-Mentawai>Simalegi Mentawai).

This is a relatively intuitive sound change: it involves the same change as in 3.1.8 but in reverse. The high front vowel /i/ is a sensical conditioning environment as the tongue apex moving forward from the alveolar ridge/upper teeth to articulate /i/ would encourage frication.

table 17. evidence for \*t>s/\_i

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| PSUM | PMEN | Simalegi | Sikabaluan | Sipora | Gloss |
| \*b[əi]tis | \*biti | bisi | biti | biti | ‘calf (of leg)’ |
| \*tiktik | \*tiʔtiʔ | siʔsiʔ | tiʔtiʔ | tiʔtiʔ | ‘tattoo’ |
| \*tinahi | \*tinaj | sinaj | tinanaj | tinanaj | ‘intestine’ |
| \*rabun | \*tinɔbutn | sinɔbut | tinɔbun | tinɔbut | ‘fog’ |
|  | \*kisej | sitej | kisej | kisej | ‘surprised’ |
|  | \*tiɡɡɔt | siɡɡɔt | tiɡɡɔt | tiɡɡɔt | ‘hang’ |
|  | \*tinali | sinali | tinali | tinali | ‘frangipani’ |
|  | \*titi | sisi | titi |  | ‘sago leaf’ |

\*k>s/\_V[+front] (Proto-Mentawai>Simalegi Mentawai).

This sound change is much harder to justify through phonetic motivation, but it is amply evidenced. Once again, PSUM forms attest to the originality of \*k here. The specificity of environment makes some intermediary velar continuant stage seem unlikely for this change, since the restriction to the pre-front vowel environment is presumably influenced by the tongue position of alveolar consonants mirroring that of front vowels, as with change 3.1.9.

table 18. evidence for \*k>s/\_V[+front]

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| PSUM | PMEN | Simalegi | Terekan | Sipora | Gloss |
| \*kawaj | \*kabej | kabej | kabej | kabej | ‘hand’ |
| \*kəmpu | \*kɛmbuʔ | sɛbbuʔ | kɛmbuʔ | kɛbbuʔ | ‘elder sibling’ |
| \*kuniɡ | \*kiniw | siɲiw | kiniw | kiniw | ‘turmeric’ |
|  | \*ɛkew | ɛsew | ɛkew | ɛkew | 2SG |
|  | \*bakɛ | basɛ |  | bakɛ | ‘thread’ |
|  | \*kejkej | sejsej | kejkej | kejkej | ‘sacred’ |
|  | \*kɛmpa | sɛppa | pɛka | kɛppa | ‘armpit’ |
|  | \*kilat | silat |  | kilat | ‘cliff’ |
|  | \*lajkɛtn | lajsɛn | lajkɛn | lajkɛt | ‘seed’ |
|  | \*ukɛ | usɛ | ukɛ |  | ‘blue sky’ |
|  | \*kɔlɛ | kɔlɛ | kɔlɛ | kɔlɛ | ‘sugarcane’ |
|  | \*kukru | kukru | kukru | kukru | ‘chase’ |

\*ɡ>Ø/V[+front]\_V (Proto-Mentawai>Simalegi).

The reduction of \*ɡ to zero is a fairly unremarkable change; a similar change affects Gayo from PSUM (Billings and McDonnell 2024). The restriction to only occurring after front vowels can be explained by economy of articulation: the tongue dorsum requires more movement to reach from a depressed position (to articulate a high vowel) to a raised position to create a closure with the velum. This sound change likely occurred prior to the change in 3.1.6, since it affects \*ɡ before front and non-front vowels equally. Strings of vowels resulting from the loss of \*ɡ are subject to the rules of Simalegi Mentawai phonotactics.

table 19. evidence for \*ɡ>Ø/V[+front]\_V

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| PSUM | PMEN | Simalegi | PCS | Sipora | Gloss |
| Sound change applied: \*ɡ/V[+front]\_V |  |
| \*pəɡu | \*lappɛɡu[[20]](#footnote-20) | lappew | \*lappɛɡu | lappɛɡu | ‘bile’ |
| \*piɡa | \*piɡa | pja | \*piɡa | piɡa | ‘how many’ |
| Sound change blocked: \*ɡ/V[-front]\_V |
| \*baɡa | \*baɡa | baɡa | \*baɡa | baɡa | ‘stomach’ |
|  | \*bulaɡatn | bulaɡat | \*bulaɡatn | bulaɡat | ‘money’ |
|  | \*sɔɡaj | tɔɡaj | \*sɔɡaj | sɔɡaj | ‘call over’ |

\*ŋ>ɲ/\_ɛ (Proto-Mentawai>Simalegi).

The motivations for this change are similar to those influencing 3.1.11: less effort is required to move the tongue from a palatal than a velar position to articulate /ɛ/.

table 20. evidence for \*ŋ>ɲ/\_ɛ

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| PSUM | PMEN | Simalegi | PCS | Simatalu | Gloss |
| Sound change applied: \*ŋ>ɲ/\_ ɛ |
| \*tuŋaw | \*tuŋew | tuɲew | \*tuŋew | tiŋow | ‘mite’ |
|  | \*iŋɛp | iɲɛp | \*iŋɛp | iŋɛp | ‘shade’ |
|  | \*ŋɛtŋɛt | ɲɛtɲɛt | \*ŋɛtŋɛt | ŋɛtŋɛt | ‘pinch’ |
| Sound change blocked: \*ŋ>/V[not ɛ] |
| \*taliŋa | \*taliŋa | taliŋa | \*taliŋa | taliŋa | ‘ear’ |
|  | \*ŋitŋit | ŋitŋit | \*ŋitŋit | ŋitŋit | ‘gnat’ |

\*n>ɲ/\_i (Proto-Mentawai>Simalegi).

The articulatory motivations for this change are very similar to the previous two.

table 21. evidence for \*n>ɲ/\_i

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| PSUM | PMEN | Simalegi | Sikabaluan | PCS | Gloss |
| Sound change applied: \*n>ɲ/\_i |
| \*kuniɡ | \*kiniw | siɲiw | \*kiniw | kiniw | ‘turmeric’ |
|  | \*niʔniʔ | ɲiʔɲiʔ | \*niʔniʔ | niʔniʔ | ‘frugal’ |
| Sound change blocked: \*n/V[not i] |
| \*ənəm | \*ɛnɛm | ɛnɛm | \*ɛnɛm | ɛnɛm | ‘six’ |
| \*nusa | \*nusa | nuta | \*nusa | nusa | ‘island’ |
|  | \*tinali | sinali | \*tinali | tinali | ‘frangipani’ |

The various resolutions to preploded nasals, and their weakness as subgrouping arguments.

PMEN had pre-ploded nasals in word-final position, where PSUM had regular final nasal consonants, but only in syllables that did not also begin with a nasal (cf. Blust 1997). These are reflected in Sipora and Sabirut, Rereiket, Simalegi, Paipajet, and Simatalu Mentawai as word-final plosives which are for some speakers allophonic with nasals and in Terekan and Sikabaluan Mentawai as nasals word-finally; in all languages these become plosive-nasal clusters when derived into the coda of a syllable with a non-nasal onset and nasals when the onset of the host syllable is similarly a nasal.

table 22. realizations of proto-mentawai preploded nasals

|  |  |  |
| --- | --- | --- |
| PMEN | Trk., Sik. | Sim., Rrk., Smt., Ppj., Sip. Sab. |
| \*pm/κ | [m] | [p~m] |
| \*tn/κ | [n] | [t~n] |
| \*kŋ/κ | [ŋ] | [k~ŋ] |
| \*pm/ω…κ[-nas] | [pm] | [pm] |
| \*tn/ ω…κ[-nas] | [tn] | [tn] |
| \*kŋ/ ω…κ[-nas] | [kŋ] | [kŋ] |
| \*pm/ω…κ[+nas] | [m] | [m] |
| \*tn/ ω…κ[+nas] | [n] | [n] |
| \*kŋ/ ω…κ[+nas] | [ŋ] | [ŋ] |

At first glance, this may seem like evidence of a subgroup linking Simalegi with Rereiket, Sipora, and Sabirut. However, these changes may easily have occurred independently. This is, in all cases except Sikabaluan and Terekan which show de-ploding, a gradient denasalization, comparable to that currently underway in Seoul Korean (Yoo and Nolan 2020). Pre-ploded nasals are an inherently unstable manner type, which are prone to collapsing into either as plosives, as in Urak Lawoi’, Rogai, and Tsat (Blust 1997; Thurgood 1999), or as nasals, which seems likely to have occurred in at least some Land Dayak languages.

Metathesis.

This irregular but recurrent process is of limited subgrouping value, but seems worthwhile to point out for completeness of understanding change within the Mentawai languages.

table 23. metathesis

|  |  |  |  |
| --- | --- | --- | --- |
| PMEN | Metathesized form | Langs with metathesis | Langs without metathesis |
| \*birit | ribit | Smt. | Ppj., Sip. |
| \*kɛmpa | pɛka | Trk. | Sim., Sik., Smt., Rkt., Sab., Sip. |
| \*batra | tabra | Sim., Sik., Rkt | Sab., Sip. |
| \*tɛkaʔ | katɛʔ | Ppj. | Sim., Sik., Smt., Rkt., Sip. |
|  | birut (Sip., Sik., Ppj.)/burit (Sim.)[[21]](#footnote-21) |  |  |
|  | pojro (Sik.)/rojpo (Sab., Sip.) |  |  |

Irregular \*t-voicing.

There are two items in PMEN where the voiceless stop \*t becomes voiced adjacent to a high vowel. Since \*t is phonetically dental, this results not in its phonemic voice alternant \*d, but in the dental plosive \*d̪, a new phoneme. These represent the origin of all the instances of \*d̪ in PMEN (since the two other items are doublets from PMEN \*ɡid̪iʔ). Usefully, we seem to have caught this process in action affecting Simalegi Mentawai’s reflex of \*tuŋklu, giving credence to this highly irregular and unexpected process. In a way, this can be explained by a draw toward symmetry, since having two dental plosives is more symmetrical than having a dental and aveolar plosive which are in a pair only phonemically, but the lack of a d-devoicing process to derive alveolar [t] meant that the result was equally asymmetrical.

table 24. Irregular \*t-voicing

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| PSUM | PMEN | Simalegi | Sikabaluan | Gloss |
| \*ɡitik | \*ɡid̪iʔ | ɡid̪iʔ | ɡid̪iʔ | ‘tickle’ |
| \*tirtir | \*d̪awd̪aw | d̪awd̪aw | lawlaw | ‘shiver’ |
|  | \*tuŋklu | d̪uŋklu | tuŋklu | ‘pull’ |

Irregular \*t>ʔ/\_#.

There are a few lexical items where \*t becomes /ʔ/ unpredictably in some languages. This process is observed both at the level of PMEN and from PMEN to its daughter languages, much like the change in 3.1.16. Given the lack of audible release for /t/ syllable-finally, word-final environment seems ample motivation for it to erode significantly, but this does not occur as a generalizable sound change.

table 24. Irregular reduction of \*t/\_#

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| PSUM | PMEN | Simalegi | Sipora | Gloss |
| \*sədsəd | (>\*sɛtsɛt)\*sɛʔsɛʔ | tɛʔtɛʔ | sɛʔsɛʔ | ‘k.o. grass’ |
|  | \*buat | bwat | buaʔ | ‘uncle’ |
|  | \*katɛt | katɛʔ | katɛt | ‘love’ |
|  | \*lɛɡɡut | lɛɡɡut | lɛɡɡuʔ | ‘mosquito’ |

Doublets.

The issue of apparently unconditioned ‘doublets’ of lexical items is an issue which has long plagued Austronesianist linguistics, and which in some ways seems unique to the Austronesian family (Blust 1988, 2011). The uniqueness of doublets in the Austronesian family comes from the fact that doublets emerge from system-internal variation, typically unconditioned, and may comprise more than two variants with little to no semantic distance separating them (Blust 2011). At least two sets of system-internal doublets seem to have emerged in PMEN.

table 25. two sets of doublets within the mentawai family

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| PSUM | PMEN | Simalegi | Sikabaluan | Rereiket | Gloss |
| \*ɡitik | \*ɡid̪iʔ | ɡid̪iʔ |  | ɡid̪iʔ | ‘tickle’ |
| \*ɡid̪aʔ | ɟid̪aʔ | ɡid̪aʔ |  |
| \*ɡilaʔ | ɟilaʔ |  | ɡilaʔ |
| \*tirtir | \*d̪awd̪aw | d̪awd̪aw | d̪awd̪aw |  | ‘shiver’ |
| \*lawlaw |  | lawlaw | lawlaw |

These two sets are clearly observable as doublets for several reasons:

1. Individual languages preserve reflexes of multiple forms with no observable semantic distance;
2. The changes required to arrive at the existing forms cannot be explained through regular sound change;

Accordingly, we have no choice but to posit each of these sets in full at PMEN, in spite of the apparent redundancy (which is solved in some languages by the loss of some of these forms).

Irregular Items.

There are additionally several lexical items for which a change occurs that cannot currently be explained with reference to sound change, and must be interpreted for the time being as entirely irregular. Given the model of proto-language splitting in table 9, it is possible in some cases to indentify a node at which an irregular form was innovated. There are are numerous other items for which an individual language will innovate an irregular form (cf. Appendix A); this of no subgrouping utility but does demonstrate the presence of irregular change, which fuels processes like doubleting.

table 26. lexeme-level innovated irregularities.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| PMEN | Irregular innvation | Innovation | Innovation branch level | Retaining languages | Innovating languages | Gloss |
| \*nɛŋɡɛ | \*ɛɡɡɛ | \*n>Ø | PMEN>PCS | Sim., Trk. | Smt., Ppj., Rkt. | ‘wait’ |
| \*tinaj | \*tinanaj | Internal CV Reduplication | PMEN>PCS+Sik.? | Sim. *sinai*, Trk. | Sik., Sab., Sip. | ‘intestine’ |
| \*kɔmbajʔ | \*ɡɔbbajʔ | \*ɡ>\*k | PMEN>PCS | Sim. *kobbai’*, Trk., Sik. | Smt., Ppj., Sip. | ‘widow’ |
| \*muŋɡej | \*buɡɡej | \*m>\*b | PMEN>PCS | Sim., Sik. | Ppj., Rkt., Sab., Sip. | ‘beach’ |
| \*pusew | \*pisow | \*u>\*i | PCS>PWS | Sim. *putew*, Trk., Sik., Rkt./Sab./Sip. *pusow* | Ppj., Smt. | ‘navel’ |
| \*tɛkaʔ | \*tɛnaʔ | \*k>\*n | PCS>PSS | Trk., Sik., Smt. *teha’*, Ppj. *kate’* | Rkt., Sab., Sip. | ‘caw (of chicken)’ |
| \*tɔm | \*ɔm | \*t>Ø | PCS>PSS | Sim., Trk., Smt., Ppj. | Rkt., Sab., Sip. | ‘patient’ |

Innovated Phonemes.

Only one phoneme is innovated out of whole cloth (i.e., not arrived at through sound change) in any of the branches below PMEN: /ɲ/ at the level of Proto-Sipora-Sabirut. This phoneme appears only in a few innovated lexemes, with identical meanings in both daugher languages: \*paɲaɲat ‘star’; \*ɲɔaŋ ‘cough; of one gravely ill’; and \*ɲɔʔɲɔʔ ‘stab’. This is unfortunately of little subgrouping value, since the close relationship between Sipora and Sabirut is already clear from (lack of) sound change.

the phonology of proto-mentawai.

 The phoneme inventory and allophonic variation in PMEN was as follows:

table 27. consonant inventory of proto-mentawai

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | [+/-voice] | Bilabial | Dental | Alveolar | Palatal | Velar | Glottal |
| **Stop** | - | p |  | t |  | k | (ʔ) |
| + | b | (d̪) | d | ɟ | ɡ |  |
| **Nasal** | + | m |  | n |  | ŋ |  |
| **Fricatives** | - |  |  | s |  |  |  |
| **Trill** | + |  |  | r |  |  |  |
| **Lateral Approximant** | + |  |  | l |  |  |  |

1. Nasals incur preplosion when they occur syllable-finally.
2. The rules of allophony which are posited in all Mentawai languages (2.2.3) are accordingly also posited for PMEN.

In PMEN, only three items out of several hundred reconstructable show \*d̪; these are also non-coincidentally members of doublet pairs (3.1.18). This extremely low number of reconstructable items makes the phonemicity of \*d̪ questionable for PMEN; demonstrating contrastiveness through minimal pairs with \*d is out of the question. However, I treat PMEN \*d̪ as phonemic (at least in terms of being a segmental and distinctive unit in the phonology) for several reasons:

1. PMEN \*d̪ cannot be treated as an allophone of PMEN \*d or \*t;
2. PMEN \*d̪ is reflected clearly in all of its first-order branches;
3. First-order branches of PMEN innovate new forms with \*d̪, and in the case of Simalegi Mentawai exhibit a sound change in the direction of /d̪/, in a manner that makes most sense of \*d̪ was already ‘available’ in the phoneme inventories of these intervening proto-languages, even if was seldom used; and
4. Phonemic dental consonants are not otherwise seen in Sumatran languages, and this seems an unlikely innovation to have occurred independently in all branches of PMEN.

the phonotactics of proto-mentawai.

 The PMEN syllable worked much like that in the modern Mentawai languages: it has the structure (C)V(C[-cont][-pal]); vowels may appear in the coda but are desyllabified; the glottal stop is an articulatory gesture rather than a segment; and voiced plosives are non-contrastive for voicing in coda position.

table 28. possible onset and coda values for pmen syllables

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Phoneme | Onset | Gloss | Coda | Gloss |
| \*p | \*pana | ‘arrow’ | \*lɛlɛp | ‘submerged’ |
| \*t | \*tabɛ | ‘fat’ | \*ambit | ‘take’ |
| \*k | \*kɔlɛ | ‘sugarcane’ | \*pɔrak | ‘land’ |
| \*b | \*baɡa | ‘stomach’ |  |  |
| \*d̪ | \*d̪awd̪aw | ‘shiver’ |  |  |
| \*d | \*dua | ‘two’ |  |  |
| \*ɟ | \*ɟɔʔɟɔʔ | ‘dog’ |  |  |
| \*ɡ | \*ɡɛla | ‘tired’ |  |  |
| \*m | \*muŋɡej | ‘beach’ | \*lalɛpm | ‘house’ |
| \*n | \*nusa | ‘island’ | \*uratn | ‘rain’ |
| \*ŋ | \*ŋaŋa | ‘mouth’ | \*asakŋ | ‘nose’ |
| \*s | \*saba | ‘python’ |  |  |
| \*r | \*rewrew | ‘hunt’ |  |  |
| \*l | \*lɛlew | ‘forested hill’ |  |  |
| \*i |  |  | \*sakaj | ‘ascend’ |
| \*u |  |  | \*pusew | ‘navel’ |

Additionally, there are a few restrictions along featural lines, whereby the onset and coda phonemes interact.

1. If the onset of a syllable contains a nasal, its coda must **not** have the same place of articulation;
2. If the onset of a syllable contains a non-geminate voiced plosive, that syllable’s coda must **not** be a plosive;
3. If the onset of a syllable contains a non-geminate voiced coronal plosive, its coda cannot be contentful;
4. \*p may not appear as coda to a syllable with a contentful onset unless the onset is \*ŋ; conversely \*m may not appear as coda if the onset contains \*ŋ;[[22]](#footnote-22)
5. If the onset contains a nasal, the coda cannot be \*k.
6. If the onset contains a labial plosive, the coda cannot be \*m. This restriction also seems to be observed in PAN (Chrétien 1965:266).

With the exception of 4); 5), which may simply be a frequency effect (nasal onsets were already the least common manner in syllable onsets aside from semivowels in PAN, Chrétien 1965:254, and coda \*k is restricted to a few PMEN innovations); and 6), these are all restrictions which have arisen somwhere in between PAN and PMEN and which can be explained in terms of featural dissimilation. While it is difficult to characterize or explain these changes uniformly beyond this statement, this in itself demonstrates that onset-coda interaction in terms of features has been instrumental in shaping the evolution of Mentawai phonology. In this context, the otherwise puzzling environment for the plosivisation of syllable-final nasals in PMEN (‘only if the onset of the host syllable is not a nasal’, 2.2.5) seems less unusual.

The fourth restriction is curious, as it suggests that the nasal>pre-ploded nasal>plosive chain shift borne out in most Mentawai languages was preceded by a change in the reverse direction restricted to interaction between syllable-final \*p and syllable-initial \*ŋ. That \*p and \*ŋ do not uniquely share any features makes this difficult to reconcile with the other feature-dissimilating changes.

table 29. possible onset-coda combinations

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | \*m | \*n | \*ŋ | \*p | \*t | \*k |
| \*m | 1) | \*mantɛman | \*mɛŋmɛŋ | 1), 4) | \*ŋamut | 5) |
| \*n | \*nanam | 1) | \*ɛnuŋ | 4) | 1) | 5) |
| \*ŋ | 4) | \*ambaŋan | 1) | \*iŋɛp | \*ŋɛtŋɛt | 1), 5) |
| \*p | 6) | \*arɛpɛtn | \*kɔpukŋ | 4) | \*ɛpat | \*ampɛk |
| \*t | \*tɔpm | \*alutɛtn | \*iŋtɛkŋ | 4) | \*ɛtut | \*tuktuk |
| \*k | \*kɔpm | \*lajkɛtn | \*pɛlɛkakŋ | 4) | \*baŋkat | \*kukru |
| \*s | \*sɛsɛpm | \*sɔtn | \*asakŋ | 4) | \*pɔsɔt | \*sɔksɔk |
| \*r | \*rɔpm | \*ɡɔratn | \*ɡurikŋ | 4) | \*bɔrɔt | \*pɔrak |
| \*l | \*lalɛpm | \*tɔlatn | \*lukŋlukŋ | 4) | \*kilat | \*ɡilik |
| \*b | 6) | \*ubatn | \*buŋkuʔ | 2), 4) | 2) | 2) |
| \*ɡ | \*siɡɛpm | \*bulaɡatn | \*ɡɛɡɛkŋ | 2), 4) | 2) | 2) |
| \*d̪ | 3) | 3) | 3) | 2), 3), 4) | 2), 3) | 2), 3) |
| \*d | 3) | 3) | 3) | 2), 3), 4) | 2), 3) | 2), 3) |
| \*ɟ | 3) | 3) | 3) | 2), 3), 4) | 2), 3) | 2), 3) |

The only clusters can be reconstructed for generic lemmas in PMEN are stop-liquid clusters and nasal-stop clusters.

In stop-liquid clusters, the plosive must be voiceless whether or not it is prenasalized; this reflects the syllable-final non-distinctiveness of voicing. In theory any plosive may cluster with either liquid; not all are attested but this is likely a frequency effect.

The segmental status of nasal-stop sequences is somewhat ambiguous: in sound changes (3.1.15, fn. 28, 32), nasal-stop cluters are frequently treated as monosegmental units, i.e., prenasalized plosives. However, these clusters do show distinctiveness for voicing, meaning they cannot be unproblematically assigned to the coda of the preceding syllable synchronically. Nasal-stop clusters must share the same place of articulation, as in PAN (Chrétien 1965); they may reflect older prenasalizations or have recently undergone enigmatic prenasalization between the levels of PAN>PMP and PSUM>PMEN; the evidence presented in 4.2.13 suggests that this process continued sporadically but uninterrupted and with the same outcomes throughout each of these evolutions. Given Chrétien’s (1965:245) observation that such prenasalization in PAN, even if apparently segmental, is ‘secondary’, it follows that homorganic nasal-stop ‘clusters’ are really just prenasalized plosives regardless of how far back prenasalization occurred. Regardless, I write these as clusters rather than single segments since this more closely reflects the articulatory reality at the level of PMEN; when the monosegmentality of these ‘clusters’ for sound change purposes becomes relevant this is duly noted.

table 30. clusters in generic proto-mentawai lemmas

|  |  |  |  |
| --- | --- | --- | --- |
| Type | Cluster | Example | Gloss |
| Nasal-stop | mb | \*timbɔ | ‘smoke’ |
|  | mp | \*tumpaj | ‘add’ |
|  | nd | \*tandaʔ | ‘sole of foot’ |
|  | nt | \*mantɛman | ‘beauty spot’ |
|  | ɲɟ | \*laɲɟaw | ‘machete’ |
|  | ɲc | \*ɔɲutn | ‘coals’ |
|  | ŋɡ | \*lɛlɛŋɡu | ‘thunder’ |
|  | ŋk | \*suŋkaj | ‘lever’ |
| Stop-liquid | tr | \*batra | ‘sago worm’ |
|  | kr | \*kukru | ‘chase’ |
|  | mpr | \*ampra | ‘tie’ |
|  | ŋkl | \*bɔŋklɔ | ‘protrude’ |
|  | ŋkr | \*suŋkra | ‘dig’ |

table 31. correspondence sets for clusters

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Cluster type | PMEN | Sim. | Trk., Sik. | Ppj., Smt., Rkt., Sab., Sip. |
| Nasal-plosive | \*mb | bː~mb | mb | bː |
| \*mp | pː~mp | mp | pː |
| \*nd | dː~nd | nd | dː |
| \*nt | tː~nt | nt | tː |
| \*ɲɟ | ɟː~ɲɟ | ɲɟ | ɟː |
| \*ɲc | cː~ ɲc | ɲc | cː |
| \*ŋk | kː~ŋk | ŋk | kː |
| \*ŋɡ | ɡː~ŋɡ | ŋɡ | ɡː |
| Prenasalized plosive | \*mp | pː~mp | mp | pː |
| \*nt | tː~nt | nt | tː |
| \*ɲc | cː~ ɲc | ɲc | cː |
| \*ŋk | kː~ŋk | ŋk | kː |
| Prenasalized plosive-liquid | \*mpr | pːr~ mpr | mpr | pːr |
| \*ŋkl | kːl~ŋkl | ŋkl | kːl |
| \*ŋkr | kːr~ŋkr | ŋkr | kːr |

Geminate plosives also appear in PMEN, but must not be regarded as clusters. Instead, they occupy the onset of the latter syllable and are copied into the coda of the former syllable.

table 32. geminate plosives

|  |  |  |
| --- | --- | --- |
| Stop | Example | Gloss |
| dː | \*rɛddɛt | ‘obedient’ |
| tː | \*ŋɔttɔt | ‘gather flowers’ |
| kː | \*akkɛkejluʔ | ‘eyebrow’ |
| ɡː | \*tiɡɡɔt | ‘hang’ |

For phonotactic purposes, the boundary between the two syllables of a reduplicative lemma functions like a word boundary: all possible cluster types may appear across this boundary that would result from the combination from all possible syllable coda-onset combinations, though far from all are attested likely due simply to reduplicative lemmas making up a comparatively small part of the lexicon.[[23]](#footnote-23) Additionally, glottal-release vowel sequences, which otherwise may only appear lemma-finally, appear adjacent to the syllable boundary here.

table 33. otherwise illegal ‘clusters’ resulting from reduplicative lemmas

|  |  |  |
| --- | --- | --- |
| Type | Example | Gloss |
| Nasal-nasal | \*mɛnmɛn | ‘saliva’ |
|  | \*mɛŋmɛŋ | ‘silent’ |
|  | \*nɛmnɛm | ‘sunken’ |
| Plosive-nasal | \*ŋɛtŋɛt | ‘narrow’ |
| Plosive-plosive | \*patpat | ‘closed’ |
|  | \*tuktuk | ‘fingertip’ |
| Plosive-fricative | \*sɔksɔk | ‘change direction’ |

table 34. reduplicative lemmas with internal glottal vowel release

|  |  |
| --- | --- |
| Example | Gloss |
| \*bɛʔbɛʔ | ‘edge’ |
| \*tajʔtajʔ | ‘set alight’ |
| \*niʔniʔ | ‘frugal’ |
| \*ŋuʔŋuʔ | ‘nervous’ |
| \*laʔlaʔ | ‘sea snail’ |
| \*sɛʔsɛʔ | ‘k.o. grass’ |

There are at least two additional restrictions on how syllables may be combined to form generic (open-class, non-reduplicative, 2-3-syllabic) lemmas, namely:

1. The onsets of two successive syllables may not be filled by dissimilar bilabials – this is a common restriction across Austronesian languages (Blust 2022; Chrétien 1965);
2. If a lemma includes a stop-liquid cluster (all those that do are bisyllabic), the latter syllable must be open – i.e., if we interpret nasal stop clusters as being underlyingly monosegmental, the ‘heavy’ lemma structure CVCCVC is disallowed. This is identical to what Chrétien (1965:245) found for PAN, under a similar assumption about the monosegmentality of apparent nasal-stop sequences.

(proto-)Mentawai as a branch of sumatran.

 Having now described the phonology of PMEN, and in the process proven the cohesiveness of Mentawai as a subfamily, the next step is logically to figure out the relation of this subfamily to the overarching macrofamily – in this case, Sumatran.

the evidence for sumatran.

 Mentawai has been robustly shown (Billings and McDonnell 2024; Smith 2017) to be a member of the Sumatran language family, which also includes nearby Enggano and Nias; the languages of Simeulue (Haloban, Leukon, Sigulai, and Simeulue/Devayan); and on the Sumatran mainland Gayo, Nasal, and the Batak languages. Of this list, Enggano, Nasal, and Gayo represent first-order branches in their own right alongside PMEN, while others are grouped as members of subfamilies: Northern Barrier Islands (the languages of Simeulue with the exception of Sigulai); Central Barrier Islands (Sigulai and Nias); and Batak (Adelaar 1981; Billings and McDonnell 2024). Sumatran is one of the groupings that replace the deprecated ‘Western Malayo-Polynesian’ grouping (cf. Blust 1977, 2013), and is presumed (in the absence of higher-level subgrouping evidence) to be a primary branch of Malayo-Polynesian (Smith 2017).

The first detailed proposal of what would come to be called the Sumatran family appears in Nothofer (1986) as ‘Barrier Islands-Batak’: a group which covers several languages of the Barrier Islands: Sigulai, Simalur, Nias, ‘Mentawai’, and Enggano, plus the Batak languages on the Sumatran mainland. Nothofer’s primary evidence for this is the reflexes of PMP \*j in all of these languages, which he posited as having changed to \*x in all of the daughter languages. As Edwards (2015) points out, however, Mentawai does not observe this change: reflexes of PMP \*j appear intervocalically in (Proto-)Mentawai as /ɡ/. The true nature of this sound change, as identified by Smith (2017), is instead the merger of PMP \*j and PMP \*g as PSUM \*g. This merger has different reflexes synchonically in daughter languages, but in all cases reflects a merger to \*g in all positions. PSUM \*g was most likely /ɡ/ with some continuant allophone[[24]](#footnote-24) intervocalically, as reflected in continuant reflexes either as allophones in this position or as the elsewhere form in PMEN, PCBI, Nasal, and Old Enggano. It is on the basis of the \*j,\*g>\*g merger that Smith (2017) expanded the grouping to include Enggano (included by Nothofer (1986), but not by Edwards (2015)) and Gayo, in a paper which also coined the term ‘Sumatran’ to describe this expanded grouping.

table 35. reflexes of pMP \*j and \*g in primary branches of sumatran

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| PMP | PSUM | PMEN | PB | PNBI | PCBI | Gayo | Old Enggano[[25]](#footnote-25) | Nasal |
| \*j, \*g | \*ɡ | \*w/\_#\*ɡ/else | \*ɡ | \*k | \*x/V\_V\*ɡ/else | k/\_#Ø/else | h | ɡ/ə\_k/\_#χ/else |

In the cases of PB and PMEN, /ɡ/ is reflected straightforwardly and requires no further explanation. In PNBI, \*g loses [+voice] to become \*k; this is a straightfoward sound change which maintains all other features. PCBI turns \*g into \*x intervocalically; this strongly resembles the allophonic variation which occurs in PMEN, but with additional loss of [+voice] and regularization. Gayo changes PSUM \*g to /k/ word-finally and zero elsewhere; both of these are fairly unremarkable changes. Nasal, meanwhile, has the voiceless uvular fricative /χ/ as the elsewhere form for its \*g reflexes. In Old Enggano, the move to \*h could be explained as the regularization of a continuant allophone as in other branches, plus a movement of the place from velar (or uvular as in Nasal) to glottal. Note that since Old Enggano undergoes coda deletion (Smith 2020), the value of syllable-final \*g is not relevant, so if PSUM \*g did have a continuant allophone intervocalically this would be the more common realization in items which meaningfully preserve \*g into Old Enggano.

If PSUM \*g is /ɡ/, then it makes sense for the value of merged \*j to be [ɡj], as suggested by Blust (2013) – the merger between \*j [ɡj] and \*g /ɡ/ would simply require the latter losing [+pal]. Smith (2024*a*) has recently suggested, on the basis of Formosan evidence, that PAN \*j was actually [ʎ], contradicting Blust’s (2013) proposal of [ɡj]. While the strength of this evidence that PAN had a sonorant value (Sagart 2024; Smith 2024*a*) is quite strong, this has little bearing on what \*j was by the time it reached PSUM[[26]](#footnote-26), since though a shift from [ɡj] to nasal values (as seen in East Formosan languages) is unmotivated (Blust 1999; Smith 2024*a*), the \*j>\*g merger is equally unmotivated if PSUM \*j is a sonorant. I therefore suggest that the theory most according with the PSUM merger is that PAN \*j became [ɡj] sometime on its path to PSUM, though I remain agnostic on the value of \*j in PMP. Since \*j,\*d>\*d is found in various languages of Western Indonesia, as a preservation, rather than innovation, of some earlier stage (Blust 2010), but which has not been coherently linked to a single merger event (Smith 2017), the timeline at this level is muddy and well beyond the scope of this paper.

Whatever the value of PMP \*j, the grouping created by its merger with \*g neatly excludes surrounding languages in Sumatra, which merge \*j and \*d to \*d, (Smith 2017, 2024*a*). Despite the neatness of this discovery, however, Smith (2017:459) remained uncomfortable with the situation that “the entire subgrouping proposal rest[ed] upon a single sound change”. Thankfully, other sound changes and merges have since been identified by Billings and McDonnell (2024), who provide three additional mergers (\*z,\*d>\*d; \*R,\*r>\*r;\*ñ,\*n>n) and two additional sound changes (\*h>Ø; \*q>\*h). On the basis of these sound changes, they expanded the family to include Nasal, another language of the Sumatran mainland.

table 36. reflexes of pMP \*d and \*z in primary branches of sumatran

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| PMP | PSUM | PMEN | PB | PNBI | PCBI | Gayo | Old Enggano | Nasal |
| \*d, \*z | \*d | \*l/{ω1=\*d, ω2=\*l} \*r/V\_V\*d/else | \*d | \*r/V\_V\*r/\_#d/else | r/V\_Vl/\_Vld/else | t/\_#r/else | d | d/ə\_t/\_#χ/else |

The merger \*z,\*d>\*d, as shown above, is very strong evidence of subgrouping complementary to \*j,\*g>\*g.

However, all the other changes and merger are much weaker as evidence. The triviality of the change \*h>Ø hardly need be demonstrated; this change affects almost all branches of PMP outside the Philippines (Smith 2024*b*). The merger \*ñ,\*n>n is so far evidenced in only three convincingly regular PSUM reconstructions; the phoneme \*ñ is in any case quite rare in PMP. The change of \*q>\*h is also not unique within the Sumatran (geographic) area: it is seen in Proto-Malayo-Chamic (Adelaar 2005). The merger \*R,\*r>\*r is problematic as the distinctiveness of the two rhotics in PMP is far from agreed upon (Adelaar 2005).

Billings and McDonnell (2024:136) claim that the new strength of Sumatran as a grouping comes not from any one change or merger, “but rather from the combination of all six”. While it certainly preferable that a subgrouping be rested on two distinctive sound changes than one, the Sumatran grouping as it currently stands is not the most robust grouping in my opinion given the commonness of all changes except \*j,\*g>\*j and \*d,\*z>\*d. Given the high rate of lexical replacement in many languages and the massive time-depth of the split between these languages, there may unfortunately be little more evidence to discover, leaving us with only a few sound changes evidenced in a few items.

sound changes affecting proto-mentawai from proto-sumatran.

 Having reconstructed PMEN in section 3, this subsection enumerates the sound changes that occurred between PSUM and PMEN and, where possible, comments on their relative ordering diachronically. (Proto-)Mentawai is identified as a primary branch of Sumatran in Billings and McDonnell (2024), a conclusion I share despite substantial differences in our reconstructions.[[27]](#footnote-27)

\*C[+nas]>C[+plo]C[+nas]/{ω=C[-nas]}.

This sound change gives nasals pre-plosion, but only if they appear word-finally and if the onset of their syllable does not also contain a nasal. This sound change and conditioning environment is not as unexpected as it may seem given the attestation of near-identical changes in Lom, Urak Lawoi’, Ragoi, and Manyukai Dayak (Blust 1997), all Austronesian languages, and the onset-coda interactions described in 3.3.

table 37. evidence for \*C[+nas]>C[+plo]C[+nas]/{ω=C[-nas]}

|  |  |  |  |
| --- | --- | --- | --- |
| PMP | PSUM | PMEN | Gloss |
| \*ənəm | \*ənəm | \*ɛnɛm | ‘six’ |
| \*sijəm | \*siɡəm | \*siɡɛpm | ‘ant’ |
| \*quzan | \*hudan | \*uratn | ‘rain’ |
| \*qambawaŋ | \*hambawaŋ | \*ambaŋan[[28]](#footnote-28) | ‘wild mango’ |
| \*hasaŋ | \*asaŋ | \*asakŋ | ‘gills/nose’ |
|  | \*unɛŋ | \*unɛŋ | ‘asleep’ |

\*C[+plo][+voice], \*C[+plo][-voice]>\*C[+plo]/\_$.

In PMEN, voicing became non-constrastive for plosives syllable-finally. This can be understood as a ‘merger’ between voiced and voiceless plosives in this environment, which are unmarked for voice. This merger could have contributed to the plosivization of preploded nasals in some Mentawai languages, since the merger leaves only one set of plosives – a contrast therefore arises between ‘hidden nasal’ plosives and regular plosives, rather than a voicing distinction, in the languages which apply plosivization, though this is contrast is only visible through patterns of allophony. This merger is unfortunately quite weakly evidenced by lexical items, syllable-final plosives not being all that frequent in PSUM, but the point is adequately demonstrated by the lack of voicing distinction for plosives syllable-finally in PMEN.

table 38. evidence for \*C[+plo][+voice], \*C[+plo][-voice]>\*C[+plo]/\_$

|  |  |  |  |
| --- | --- | --- | --- |
| PMP | PSUM | PMEN | Gloss |
| \*huab | \*huab | \*ɔap | ‘yawn’[[29]](#footnote-29) |
| \*sədsəd | \*sədsəd | \*sɛʔsɛʔ | ‘k.o. grass’ |

\*ə>\*ɛ.

This sound change affects the phoneme \*ə in PSUM in all positions. Schwa is not re-innovated in PMEN, and disappears from the phoneme inventory. This is a fairly intuitive sound change that can be explained with reference to the pull towards symmetry of the vowel space (Boersma 1997). The emergence of \*ɔ, both through change 4.2.10 and numerous innovations, also reflects this draw for vowel symmetry in both its existence (two front vowels with only a single back counterpart would be highly unusual), and its height.

table 38. evidence for \*ə>\*ɛ

|  |  |  |  |
| --- | --- | --- | --- |
| PMP | PSUM | PMEN | Gloss |
| \*bələk | \*bələk | \*bɛlɛʔ | ‘fall’ |
| \*əpat | \*əpat | \*ɛpat | ‘four’ |
| \*ənəm | \*ənəm | \*ɛnɛm | ‘six’ |
| \*tabɛ | \*tabəh | \*tabɛ | ‘fat’ |

\*h>Ø.

This sound change occurs in all positions in PMEN; it is extremely common cross-linguistically and requires no further comment.

table 39. evidence for \*h>Ø

|  |  |  |  |
| --- | --- | --- | --- |
| PMP | PSUM | PMEN | Gloss |
| \*hasaŋ | \*hasaŋ | \*asakŋ | ‘nose’ |
| \*huab | \*huab | \*ɔap | ‘yawn’ |
| \*suluq | \*suluh | \*sulu | ‘sun’ |
| \*panaq | \*panah | \*pana | ‘arrow’ |

\*s>Ø/\_#.

This sound change is restricted to word-final position, but is well-evidenced in PMEN. This is the simplest available change which complies with PMEN’s prohibition of continuants in coda position.

table 40. evidence for \*s>Ø/\_#

|  |  |  |  |
| --- | --- | --- | --- |
| PMP | PSUM | PMEN | Gloss |
| Sound change applied: \_# |
| \*atas | \*atas | \*ata | ‘tall’[[30]](#footnote-30) |
| \*nipis | \*nipis | \*nimpi | ‘thin’ |
|  | \*baɡas | \*baɡa | ‘stomach’ |
| Sound change blocked: elsewhere |
| \*usiR | \*usir | \*usi | ‘chase’ |

\*k>ʔ/\_#.

This sound change occurs in word-final position, and is unremarkable considering the propensity of unreleased /k/ to become /ʔ/ in Austronesian languages. Word-final \*k is then re-innovated in PMEN.

table 41. evidence for \*k>ʔ/\_$

|  |  |  |  |
| --- | --- | --- | --- |
| PMP | PSUM | PMEN | Gloss |
| \*bak | \*bak | \*ba**ʔ** | NEG.IMP |
| \*ləpak | \*ləpak | \*lɛpaʔ | PERF |
| \*qutək | \*utək | \*utɛ**ʔ** | ‘head’ |
| \*tandak | \*tandak | \*tanda**ʔ** | ‘sole of foot’ |
| \*tiktik | \*tiktik | \*ti**ʔ**ti**ʔ** | ‘tattoo’ |

\*w>\*b/\_V.

Although comparatively uncommon, this change is easy to explain in articulatory terms: the velar coarticulation disappears and the manner becomes plosive.

table 42. evidence for \*w>\*b/{#|V}\_V

|  |  |  |  |
| --- | --- | --- | --- |
| PMP | PSUM | PMEN | Gloss |
| Sound change applied: \*w>\*b/{#|V}\_V |
| \*sawa | \*sawa | \*saba | ‘python’ |
| \*siwa | \*siwa | \*siba | ‘nine’ |
| \*walu | \*walu | \*balu | ‘eight’ |
| \*wakat | \*wakat | \*baŋkat | ‘root’ |
| Sound change blocked: \*w/\_# |
| \*takaw | \*takaw | \*taŋkew | ‘steal’ |
| \*tuŋaw | \*tuŋaw | \*tuŋew | ‘mite’ |

\*d>\*r/V\_V.

This sound change occurs intervocalically. It is worth keeping in mind that even phonemic \*d was in free variation as [d~r] in PMEN, so this can be thought of as an “unfinished” unconditioned sound change at PMEN.

table 45. evidence for \*d>\*r/V\_V

|  |  |  |  |
| --- | --- | --- | --- |
| PMP | PSUM | PMEN | Gloss |
| Sound change applied: \*d>\*r/V\_V |
| \*adat | \*adat | \*arat | ‘tradition’ |
| \*tuzuq | \*tuduh | \*turu | ‘index finger’ |
| \*zauq | \*[a]dauh | \*arew | ‘far’ |
|  | \*sada | \*sara | ‘one’ |
| Sound change blocked: \*d/#\_ |
| \*duha | \*dua | \*dua | ‘two’ |

\*a>\*e/\_V[+high].

This sound change subsumes changes affecting two vowel-semivowel strings (‘dipthongs’), but is a useful generalization: \*a becomes \*e before semivowels [w] and [j] (high vowels will always be desyllabified after /a/, since both possibilities yield legal diphthongs): so PSUM \*aj gives PMEN \*ej and PSUM \*aw gives PMEN \*ew. As usual, economy of articulation proves a useful explanation.

table 47. evidence for \*a>\*e/\_V[-syl]

|  |  |  |  |
| --- | --- | --- | --- |
| PMP | PSUM | PMEN | Gloss |
| \*abaw | \*abaw | \*abew | ‘large’ |
| \*kasaw | \*kasaw | \*kasew | ‘rafters’ |
| \*matay | \*mataj | \*matej | ‘dead’ |
| \*qatay | \*hataj | \*atej | ‘liver’ |
| \*sapaw | \*sapaw | \*sapew | ‘field hut’ |
| \*tuŋaw | \*tuŋaw | \*tuŋew | ‘mite’ |
| \*zauh | \*[a]dauh | \*arew | ‘far’ |

\*r>various.

The changes that affect \*r are by far the most complex, the least intuitive, and defy explanation as a ‘simple’ one-step sound change. The patterns governing PSUM \*r can be described informally as follows: 1) If schwa adjacent to \*r can be removed with the resulting string being phonotactically legal, this is done with no change to the value of \*r; 2) if the lemma contains a back vowel either adjacent to \*r or in the prior syllable, then \*r goes to zero with no change to surrounding vowels; else 3) \*r and the nearest non-back vowel, preferencing subsequent over prior vowel if both are available, both go to \*ɔ, except 4) after \*t, where the same process instead produces \*u – likely resultant from height spreading given the unique dental place of articualtion. Strings of vowels produced after these processes have affected \*r are then subject to the phonotactic constraints described in section (2.2); if a string of three successive vowels is produced involving schwa then schwa is simply deleted, as in rule 1). The below table displays each of these options for sound change in practice.

table 48. evidence for \*r>various

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Type | PMP | PSUM | PMEN | Gloss |
| (1) | \*qabatəR | \*habatər | \*batra[[31]](#footnote-31) | ‘sago worm’ |
| (2) | \*baRu | \*baru | \*baw | ‘new’ |
| (2) | \*buRiq | \*burih | \*bui | ‘wash’ |
| (2) | \*duRi | \*duri | \*dui | ‘thorn’ |
| (2) | \*qatəluR | \*hantəlur | \*antɛlu | ‘egg’ |
| (2) | \*qusiR | \*husir | \*usi | ‘chase’ |
| (2) | \*Rumaq | \*rumah | \*uma | ‘house’ |
| (2) | \*Rusuk | \*rusuk | \*usuʔ | ‘rib’ |
| (3) | \*bibiR | \*bibir | \*bibɔ | ‘lip’ |
| (3) | \*diRi | \*diri | \*diɔʔ | ‘stand’ |
| (3) | \*layaR | \*lajar | \*laɟɔ | ‘sail’ |
| (3) | \*Rabun | \*rabun | \*tinɔbutn | ‘fog’ |
| (3) | \*Ratus | \*ratus | \*ɔtu | ‘hundred’ |
| (3) | \*Rapus | \*rapus | \*ɔmpu | ‘bind’ |
| (3) | \*rəbuŋ | \*rəbuŋ | \*ɔmbuŋ | ‘thin bamboo’ |
| (3) |  | \*s[ə]raw | \*sow | ‘cry’ |
| (3) |  | \*timbər | \*timbɔ | ‘smoke’ |
| (4) | \*buntəR | \*buntər | \*lajmuntu[[32]](#footnote-32) | ‘round’ |
| (4) | \*tirtir | \*tirtir | \*d̪awd̪aw[[33]](#footnote-33) | ‘shiver’ |

\*ɡ>\*w/\_#.

This sound change creates a secondary (labial) articulation, in the process switching the manner to approximant. Given both that PMEN prohibits voiced plosives in word-final position and that PSUM \*ɡ likely had a continuant allophone, this change is well motivated – if the original environment for this allophone (probably [ɤ]) was V\_ and inclusive of the word-final environment, the change to [w] would require only labialization.

table 49. evidence for \*ɡ>w/\_#

|  |  |  |  |
| --- | --- | --- | --- |
| PMP | PSUM | PMEN | Gloss |
| \*dələj | \*dələɡ | \*lelew | ‘forested hill’ |
| \*pusəj | \*pusəɡ | \*pusew | ‘navel’ |
| \*quləj | \*huləɡ | \*ulew | ‘snake’ |
| \*qunəj | \*hunəɡ | \*unew | ‘pith’ |

\*j>\*ɟ/\_V.

This change is posited intervocalically on the basis of the three items below. This change is fairly intuive considering the phonotactics of PMEN, which prohibits semivowels in syllable onsets – conversion to plosive while maintaining the place is an easy method to accord syllable-initial instances of \*j with this rule. This seemingly being the origin of \*ɟ in PMEN (recalling that PSUM has neither \*ɟ nor \*c) also gives a reason for \*ɟ to have emerged as the sole palatal plosive without a voiceless alternant. Since PSUM \*j does not appear word-initially, and word-final \*j only appears as a part of ‘diphthongal’ vowel-semivowel sequences (and in this regard remains unchanged in PMEN, cf. 4.2.10, the prevocalic environment must be assigned.

table 50. evidence for \*j>\*ɟ/\_V

|  |  |  |  |
| --- | --- | --- | --- |
| PMP | PSUM | PMEN | Gloss |
| \*bayaw | \*bajaw | \*baɟaʔ | ‘brother-in-law’ |
| \*layaR | \*lajar | \*laɟɔ | ‘sail’ |
| \*layu | \*laju | \*laɟuʔ | ‘wither’ |

\*d>l/{ω1=d,ω2=l}.

Since both consonants are alveolar, this change can be conceived of as featural assimilation across syllable onsets. The same change occurs in PCBI.

table 51. evidence for \*d>l/{ω1=d,ω2=l}

|  |  |  |  |
| --- | --- | --- | --- |
| PMP | PSUM | PMEN | Gloss |
| \*dələj | \*dələɡ | \*lɛlew | ‘forested hill’ |
| \*daləm | \*daləm | \*lal**ɛ**pm | ‘inside’ |
| \*dələp | \*dələp | \*lɛlɛp | ‘submerged’ |
| \*dila | \*dila | \*lila | ‘tongue’ |

Enigmatic prenasalization.

PMEN prenasalized plosives sometimes appear where PSUM has plosives with no prenasalization. It is not possible to explain the appearance of prenasality as a feature with reference to the environment of these plosives. Instead, this is an example of a phenomenon noticed as far back as Dempwoff (1922) in his work on Ngaju Dayak, and dubbed ‘enigmatic prenasalization’[[34]](#footnote-34): the emergence of prenasalization on word-internal plosives in Austronesian languages with no clear phonetic motivation. Enigmatic prenasalization occurs both between PMP and PSUM (1), and between PSUM and PMEN (2).

table 52. enigmatic prenasalization

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Type | PMP | PSUM | PMEN | Gloss |
| (1) | \*hajək | \*aŋɡəh |  | ‘kiss’ |
| (1) | \*qatəluR | \*hantəluR | \*antɛlu | ‘egg’ |
| (1) | \*takaw | \*taŋkaw | \*taŋkew | ‘steal’ |
| (2) | \*nipis | \*nipis | \*nimpi | ‘thin’ |
| (2) | \*Rapus | \*Rapus | \*ɔmpu | ‘bind’ |
| (2) | \*wakat | \*wakat | \*baŋkat | ‘root’ |
| [[35]](#footnote-35) | \*rəbuŋ | \*rəbuŋ | \*ɔmbuŋ | ‘thin bamboo’ |

Excrecent Glottal Stop.

There are three instances where an item in PMEN with a word-final glottal stop descends from a PSUM form with no apparent motivation for the glottal stop to appear. In all instances the glottal stop appears after a back vowel, which I take to be non-coincidental: the glottal stop as an articulatory gesture requires much less effort from a back tongue position than a mid or front one. It is not possible to generalize the appearance of the glottal stop in terms of environment, so I belive this must be interpreted as yet another irregular process.

table 53. excrecent glottal stop

|  |  |  |  |
| --- | --- | --- | --- |
| PMP | PSUM | PMEN | Gloss |
| \*diRi | \*diri | \*diɔʔ | ‘stand’ |
| \*kəmpu | \*kəmpu | \*kɛmbuʔ[[36]](#footnote-36) | ‘elder sibling’ |
| \*layu | \*laju | \*laɟuʔ | ‘wither’ |

Exceptions for -CVC ‘roots’.

Blust (1988) provides 231 -CVC ‘roots’ which are attested throughout Austronesian languages in the formation of semantically linked (typically bisyllabic) lemmas, either reduplicatively or non-reduplicatively. Although it is unclear how many of these can be reconstructed all the way back to PAN, a fair number of these roots are clearly visible in PMEN. What is of particular interest for our purposes, however, is how a number of these roots are used productively in the formation of new lemmas in PMEN, in ways which seem to ignore sound changes that have affected existing lemmas. Rather, -CVC roots seen elsewhere in AN are reproduced in their closest possible approximation in PMEN phonology, in some cases ignoring phonotactic restrictions otherwise in place.

table 54. cvc root ‘exceptions’

|  |  |  |  |
| --- | --- | --- | --- |
| CVC root | Expected form if sound change applied | PMEN examples | Gloss |
| \*-ɡuk ‘throaty sound’ | \*-ɡuʔ | \*ɡuk | ‘swallow’ |
| \*-laq ‘split’ | \*-la | \*silaʔ | ‘split’ |
| \*-rit ‘scratch a line’ | \*-ɔt | \*birit | ‘rip’ |
| \*kirit | ‘grate’ |
| \*-tək ‘mud’ | \*-tɛʔ | \*lottɛk | ‘mud’ |

There are intriguing resemblances to these forms elsewhere in AN: \*biriC is reconstructed to PAN with the same meaning on the basis of Paiwan *birits*, and has the doublet \*bərit reflected in various languages of Borneo, but is seen nowhere else in Sumatran or anywhere West of Borneo to my knowledge; it should therefore be interpreted as a PMEN innovation based on a preserved -CVC root and possibly also some other ‘gestalt’ (cf. Blust 1988:59-61) template. We can make similar observations for PMEN \*kirit (Malay *kerit*, but no attestation within Sumatran; probably independent innovation given ‘gestalt’ template) and \*lɔttɛk (PB \*litək, within Sumatran but cannot be justified as reflecting the same form).

This suggests something rather intriguing: that the -CVC roots proposed by Blust (1988) not only have psychological reality for speakers, they are stored somewhere apart from where the remainder of the lexicon is stored, and ignore sound changes which have proliferated fully throughout the ‘regular’ open-class lemmas.

mentawai as a primary branch of sumatran.

 Billings and McDonnell (2024), the only detailed phylogeny of Sumatran at current, identifies Mentawai as a primary branch of the family. However, there are a few commonalities between sound changes affecting PMEN and those affecting other branches of PSUM which deserve attention as potential evidence for subgrouping of Mentawai and other first-order branches, namely Enggano and Central Barrier Islands.

Proto-Southern Barrier Islands? (Mentawai-Enggano Subgrouping)

In terms of sound change, Enggano and PMEN have a few intriguing similarities. The change \*w>\*b occurs also in Enggano, but only alongside a merger that turns all nasals to voiced stops; Edwards (2015) nonetheless cites this as the best evidence for the subgrouping of Enggano with any other Sumatran languages. However, similar mergers occur nearby in Rejang (Blust 1984) and Dairi Batak (Adelaar 1981), making it seem plausible that the same change affected Mentawai and Enggano separately; Edwards (2015) similarly considers this the most likely scenario. Billings and McDonnell (2024:143) identify the ‘aberrant’ change of \*s>\*t as having occurred in both Simalegi Mentawai and Enggano, but, as they correctly diagnose, Enggano shares none of the other changes that PMEN has undergone, meaning these must be independent changes.

One area in which PMEN and Enggano do bear a striking resemblance is in the highly idiosyncratic development of a set of pronominal indexing prefixes reflecting PSUM free forms.

table 55. pronominal indexing prefixes in psbi?

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| PMP | PSUM | PSBI? | PMEN[[37]](#footnote-37) | Old Enggano | Gloss |
| \*aku | \*aku | \*ku- | \*ku- | ʔu- | 1SG |
| \*ita | \*ita | \*ta- | \*ta- | ka-  | 1PL.INCL |
| \*kami | \*kami | \*ku- kai | \*ku- kai | ʔu- ʔai | 1PL.EXCL |
|  |  | \*nu- | \*nu- | u- | 2SG |
| \*kamu | \*kamu | \*nu- kamu | \*nu- kam | u- aʔa | 2PL |
| \*ia | \*ia | \*i- | \*i- | i- | 3SG |
| \*sida | \*sida | \*da- | \*da- | da- | 3PL |

Other branches of PSUM (PCBI, PNBI, viz. Akoli and Arka 2025; Mullan 2025) innovate paradigms of pronominal indexing prefixes by grammaticalizing free forms, but these appear to be the result of discrete grammaticalization events independent of one another. This is most clearly visible for our purposes in three prefixes which are idiosyncratic to PSBI: 1PL.EXCL; 2SG; and 2PL. The 1PL.EXCL form in PSBI features irregular loss of \*m, and also has a highly unusual structure of the free form acting as a resumptive pronoun *after* the verb, agreeing with the 1 prefix *\**ku-which is unmarked for number. The 2SG form \*nu-is presumably modelled after the GEN/NOM2 pronominal index of the same form, inherited from PMP, since the synchronous PSUM free pronoun for 2SG \*əkawdid not grammaticalize. Loss of \*n- in Old Enggano is irregular, but not extraordinary considering other irregular changes in Old Enggano. The 2PL form is created with the PL free form aʔa in Old Enggano rather than a reflex of \*kamas in PMEN, but the structure of a person-only prefix being paired with a modifying postposed free form providing number is retained, and is highly idiosyncratic.

If these two branches do indeed subgroup, the level of similarity incurred by this relation is very small, even trivial: it comprises no more than a single sound change and some irregular pronominal innovations. These similarities dwarf in comparison to the changes that have since affected Enggano. Comparing the lexicon and sound systems of PMEN (herein) and Enggano (Edwards 2015, 2020), and presuming that they split from a common ancestral ‘PSBI’ which resembled PSUM in all manners except \*w>\*b and the pronominal system, one is immediately struck by how much Enggano has diverged from this ancestor, and how little PMEN has done the same. This fits, however, with what we might expect of a small, highly isolated speech community such as that of Enggano island, where there is little to no dialect variation and lexical innovations can very quickly become known throughout the entire speaking population and thus become entrenched. If PSBI did exist, it was likely spoken for an extremely short period of time, probably in Siberut, directly following the settlement of Siberut from mainland Sumatra, and after which the ancestors of Enggano speakers, extremely small in number, left to develop their own language in an extremely small speech community with no contact with speakers of related languages, allowing their own language to become radically reconfigured.

Mentawai-PCBI Subgrouping?

The change \*d>l/{ω1=d, ω2=l} also occurs in PCBI. This may seem at first a highly unique and perhaps even *sui generis* sound change, but given the propensity for cross-syllable effects in sound change that we have already witnessed, this seems less the case. This kind of supersyllabic lateralizing change also has precedent within Austronesian in Thao, which lateralizes \*d>ɬ in the same environment (Blust 2003:77). The unexpectedness also decreases when we consider the probable [d~r] allophony of \*d in PSUM. Also in PCBI, \*r takes surrounding vowels to \*o in all environments except for a\_#, which goes to \*a (Billings and McDonnell (2024:131). The similarity here is remarkable, but overall the patterns do not line up (PCBI has no restriction requiring the presence of a back vowel for \*r, capturing a nearby vowel, to produce \*o). These changes may have a shared origin but are implemented in rather different manners. There are a few lexeme-level innovations that seem to be shared between both (enigmatic nasalization of PSUM \*rəbuŋ>\*ombuŋ; irregular PSUM \*bajaw>\*bajak), but these could easily be coincidences. Nevertheless, the potential for a link here should be given further attention, especially in light of folkloric myths about the original Mentawaian settlers having originated from Nias (Schefold 1989).

Other (Weaker) Evidence.

The change \*k>\*ʔ/\_# is shared, and indeed superceded, by the more general change C[+stop][-voice]> ʔ/\_# in Gayo (Billings and McDonnell 2024). However, the lack of audible release for syllable-final consonants in either branch (Eades and Hajek 2006) makes this change unremarkable. The change \*d>r in Gayo, and \*d>\*r/V\_{V|#} in PNBI, are very similar to the change \*d>r/V\_V in PMEN, especially considering that PMEN had [r] as an allophonic variant of \*d even outside of intervocalic position. The change \*d>\*r/V\_V in PCBI is near-identical. However, none of these preclude the possibility that PSUM had [r] as an allophone for \*d, as in PMEN, which would explain all of these resulting sound changes if regularized. The change \*h>Ø is shared with PNBI and Old Enggano, but this is extremely common cross-linguistically and of little subgrouping value.

One intriguing possibility is that the loss of voicing contrast for syllable-final plosives and/or prohibition on syllable-final continuants in PMEN occurred at some earlier stage as a precursor to coda loss, which reached its completion only in either Old Enggano or PCBI – this hypothesis works equally well with either branch. In the same breath, coda loss in these branches may well have destroyed otherwise valuable evidence for subgrouping, since we have no way of knowing if any of the unique onset-coda phonotactic interactions that characterize PMEN (3.3) were shared with these branches.

conclusion.

 This reconstruction of Proto-Mentawai should demonstrate adequately the imprecision of referring to the Mentawai languages as a single linguistic unit. The reconstruction fits with a picture of Mentawai as sitting within the Sumatran family (Billings and McDonnell 2024), with a potential link to Enggano. Whether this link can be more robustly demonstrated; whether any additional linkages within the Mentawai family can be proven; and why Simalegi Mentawai shows such a truly disproportionate number of changes compared to all other branches of PMEN all remain fascinating open questions to be tackled by future research.

In addition to dealing with local issues of phylogenics, the reconstruction here has broader implications for the study of sound change in the Sumatran and Austronesian areas: a number of irregular and unconditioned processes are shown to be at play recurrently and at multiple levels of branching, supporting Blust’s (2005) notion that sound change in the Austronesian area need not always be phonologically conditioned in the manner that the Neogrammarian hypothesis has taught us to expect.

appendix a. proto-mentawai reconstruction of selected items

All forms come either directly from the speakers listed in the acknowledgements or Cambielli (1998) (all languages except Rereiket and Sabirut Mentawai) and Mendrofa, Salakkirat, and Henry (2019) (Rereiket Mentawai). Terekan, Rereiket, and Paipajet Mentawai items are from written sources alone.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| PMEN | Sim. | Trk. | Sik. | PCS | PWS | Ppj. | Smt. | PSS | Rkt. | PSiSa | Sab. | Sip. | Gloss |
| \*abew | abew |  | abew | \*abew | \*abew |  | abew | \*abew | abew | \*abew | abew | abew | ‘large’ |
| \*akkɛkejluʔ | akkɛkejluʔ |  | akkɛkejluʔ | \*akkɛkejluʔ |  |  |  | \*akkɛkejluʔ | akkɛkejluʔ | \*akkɛkejluʔ | akkɛkejluʔ | akkɛkejluʔ | ‘eyebrow’ |
| \*alaʔ | alaʔ | alaʔ | alaʔ | \*alaʔ | \*alaʔ | alaʔ | alaʔ | \*alaʔ | alaʔ | \*alaʔ | alaʔ | alaʔ | ‘take’ |
| \*alutɛtn | alutɛt | alutɛn | alutɛn | \*alutɛtn | \*alutɛtn | alutɛn | alutɛn | \*alutɛt | alutɛt | \*alutɛt | alutɛt | alutɛt | ‘fire’ |
| \*ambit |  |  |  | \*ambit |  |  |  | \*ambit |  | \*ambit | abbit | abbit | ‘bring’ |
| \*ampra | appra |  |  | \*ampra | \*ampra |  |  | \*appra | appra | \*appra | appra | appra | ‘tie’ |
| \*antɛlu | attɛlu | attɛlu | antɛlu | \*antɛlu | \*antɛlu |  | attɛlu | \*attɛlu | attɛlu | \*attɛlu | attɛlu | attɛlu | ‘egg’ |
| \*arat | arat | arat | arat | \*arat | \*arat | arat | arat | \*arat | arat | \*arat | arat | arat | ‘tradition’ |
| \*arɛpɛtn | arɛpɛn | arɛpɛt | arɛpɛn | \*arɛpɛtn |  |  |  | \*arɛpɛt | arɛpɛt | \*arɛpɛt |  | arɛpɛt | ‘eyelash’ |
| \*arew | arew | arew | arew | \*arew | \*arew | arew | arew | \*arew | arew | \*arew | arew | arew | ‘far’ |
| \*asa | ata | asa | asa | \*asa | \*asa | asa | asa | \*asa | asa | \*asa | asa | asa | ‘whet’ |
| \*asakŋ |  | asaŋ | asaŋ | \*asakŋ | \*asakŋ | asaŋ |  | \*asak | asak | \*asak | asak | asak | ‘nose’ |
| \*ata |  | ata | ata | \*ata | \*ata | ata | ata | \*ata | ata | \*ata | ata | ata | ‘tall’ |
| \*atej | atej | atej | atej | \*atej | \*atej | atej | atej | \*atej | atej | \*atej | atej | atej | ‘liver’ |
| \*baʔ | baʔ | baʔ | baʔ | \*baʔ | \*baʔ | baʔ | baʔ | \*baʔ | baʔ | \*baʔ | baʔ | baʔ | NEG.IMP |
| \*baɡa | baɡa | baɡa | baɡa | \*baɡa | \*baɡa | baɡa | baɡa | \*baɡa | baɡa | \*baɡa | baɡa | baɡa | ‘stomach’ |
| \*baɟaʔ | bad̪aʔ | baɟaʔ | baɟaʔ | \*baɟaʔ | \*baɟaʔ | baɟaʔ | baɟaʔ | \*baɟaʔ | baɟaʔ | \*baɟaʔ | baɟaʔ | baɟaʔ | ‘brother-in-law’ |
| \*bakɛ | bakɛ |  |  | \*bakɛ | \*bakɛ |  | baxɛ | \*bakɛ | bakɛ | \*bakɛ | bakɛ | bakɛ | ‘thread’ |
| \*baɡej | baɟej | baɡej | baɡej | \*baɡej | \*baɡej | baɡej | baɡej | \*baɡej | baɡej | \*baɡej | baɡej | baɡej | ‘other’ |
| \*balu | balu | balu | balu | \*balu | \*balu | balu | balu | \*balu | balu | \*balu | balu | balu | ‘eight’ |
| \*baŋkat |  |  | baŋkat | \*baŋkat |  |  |  | \*bakkat | bakkat | \*bakkat | bakkat | bakkat | ‘root’ |
| \*batra | kamra |  | tabra | \*tabra | \*tabra |  |  | \*tabra | tamara | \*batra | batra | batra | ‘sago worm’[[38]](#footnote-38) |
| \*baw | baw | baw | baw | \*baw | \*baw | baw | baw | \*baw | baw | \*baw | baw | baw | ‘new’ |
| \*bɛʔbɛʔ | bɛʔbɛʔ | bɛʔbɛʔ | bɛʔbɛʔ | \*bɛʔbɛʔ | \*bɛʔbɛʔ | bɛʔbɛʔ | bɛʔbɛʔ | \*bɛʔbɛʔ | bɛʔbɛʔ | \*bɛʔbɛʔ | bɛʔbɛʔ | bɛʔbɛʔ | ‘edge’ |
| \*bibɔ | bibɔ | bibɔ | bibɔ | \*bibɔ | \*bibɔ |  | bibɔ | \*bibɔ | bibɔ | \*bibɔ | bibɔ | bibɔ | ‘lip’ |
| \*bilew | bilew | bilew |  | \*bilow | \*bilow | bilow |  | \*bilow |  | \*bilow | bilow | bilow | ‘Kloss’s gibbon (*hylobates klossii*)’ |
| \*birit | ribit | birit | birit | \*birit | \*birit | birit | ribit[[39]](#footnote-39) | \*birit | birit | \*birit | birit | birit | ‘rip’ |
| \*biti | bisi | biti | biti | \*biti | \*biti | biti | biti | \*biti | biti | \*biti | biti | biti | ‘calf (of leg)’ |
| \*bɔku | bɔku | bɔku | bɔku | \*bɔku | \*bɔku | bɔku | bɔxu | \*bɔku |  | \*bɔku | bɔku | bɔku | ‘shoulder’ |
| \*bɔŋklɔ | bɔkklɔ |  |  | \*bɔŋklɔ | \*bɔŋklɔ | bɔkklɔ | bɔkklɔ | \*bɔkklɔ | bɔkklɔ |  |  |  | ‘protrude’ |
| \*bɔrɔt | bɔrɔt |  | bɔrɔt | \*bɔrɔt | \*bɔrɔt | bɔrɔt | bojrɔt[[40]](#footnote-40) |  |  |  |  |  | ‘straight’ |
| \*buat | bwat | buat | buatʔ | \*buaʔ | \*buaʔ |  | buaʔ | \*buaʔ | buaʔ | \*buaʔ | buaʔ | buaʔ | ‘uncle’ |
| \*bui | bui |  | bui | \*bui |  |  |  | \*bui | bui | \*bui | bui | bui | ‘wash’ |
| \*bulaɡatn | bulaɡat | bulaɡan | bulaɡan | \*bulaɡatn | \*bulaɡatn | bulaɡan | bulaɡan | \*bulaɡan | bulaɡat | \*bulaɡat | bulaɡat | bulaɡat | ‘money’ |
| \*diɔʔ | djɔʔ | riɔʔ | riɔʔ | \*diɔʔ | \*diɔʔ | riɔʔ | riɔʔ | \*diɔʔ | riɔʔ | \*diɔʔ | riɔʔ | riɔʔ | ‘stand’ |
| \*dua | dua | dua | dua | \*dua | \*dua | dua | dua | \*dua | dua | \*dua | dua | dua | ‘two’ |
| \*dui | dui | dui | dui | \*dui | \*dui | dui | dui | \*dui | dui | \*dui | dui | dui | ‘thorn’ |
| \*duruʔ | duruʔ | duruʔ | duruʔ | \*duruʔ | \*duruʔ | duruʔ | duruʔ | \*duruʔ | duruʔ | \*duruʔ | duruʔ | duruʔ | ‘gather’ |
| \*d̪awd̪aw | d̪awd̪aw | d̪awd̪aw | d̪awd̪aw | \*d̪awd̪aw |  |  |  |  |  |  |  |  | ‘shiver’ |
| \*ɛkew | ɛsew | ɛkew | ɛkew | \*ɛkew | \*ɛkew | ɛkɔ[[41]](#footnote-41) | ɛxew | \*ɛkew | ɛkew | \*ɛkew | ɛkew | ɛkew | 2SG |
| \*ɛnɛm | ɛnɛm | ɛnɛm | ɛnɛm | \*ɛnɛm | \*ɛnɛm | ɛnɛm | ɛnɛm | \*ɛnɛm | ɛnɛm | \*ɛnɛm | ɛnɛm | ɛnɛm | ‘six’ |
| \*ɛnuŋ | ɛnuŋ | ɛnuŋ | ɛnuŋ | \*ɛnuŋ | \*ɛnuŋ | ɛnuŋ | ɛnuŋ | \*ɛnuŋ | ɛnuŋ | \*ɛnuŋ | ɛnuŋ | ɛnuŋ | ‘walk’ |
| \*ɛpat | ɛpat | ɛpat | ɛpat | \*ɛpat | \*ɛpat | ɛpat | ɛpat | \*ɛpat | ɛpat | \*ɛpat | ɛpat | ɛpat | ‘four’ |
| \*ɛtut |  |  | ɛtut | \*ɛtut | \*ɛtut |  |  | \*ɛtut | ɛtut | \*ɛtut | ɛtut | ɛtut | ‘fart’ |
| \*ɡɛla | zɛla | ɡɛla |  | \*ɡɛla | \*ɡɛla | ɡɛla |  |  |  |  |  |  | ‘tired’ |
| \*ɡɛɡɛkŋ | ɟɛzɛŋ | ɡɛɡɛŋ | ɡɛɡɛŋ |  |  |  |  |  |  |  |  |  | ‘watch’ |
| \*ɡid̪aʔ | ɟid̪aʔ |  | ɡid̪aʔ |  |  |  |  |  |  |  |  |  | ‘tickle’ |
| \*ɡilaʔ | ɡilaʔ |  |  | \*ɡilaʔ | \*ɡilaʔ |  |  | \*ɡilaʔ | ɡilaʔ | \*ɡilaʔ |  | ɡilaʔ | ‘tickle’ |
| \*ɡilik | ɟilik | ɡilik | ɡilik | \*ɡilik |  |  |  | \*ɡilik | ɡilik | \*ɡilik |  | ɡilik | ‘sunlight’ |
| \*ɡid̪iʔ | ɡid̪iʔ |  |  | \*ɡid̪iʔ | \*ɡid̪iʔ |  | ɡid̪iʔ | \*ɡid̪iʔ | ɡid̪iʔ | \*ɡitiʔ | ɡitiʔ | ɡitiʔ | ‘tickle’ |
| \*ɡɔʔ | ɡɔʔ |  | ɡɔʔ | \*ɡɔʔ | \*ɡɔʔ |  | ɡɔʔ | \*ɡɔʔ | ɡɔʔ | \*ɡɔʔ | ɡɔʔ | ɡɔʔ | ‘swollen’ |
| \*ɡɔratn | ɡɔrat |  | ɡɔrat | \*ɡɔratn | \*ɡɔratn |  | ɡɔran | \*ɔrat | ɔrat | \*ɔrat | ɔrat | ɔrat | ‘stairs’ |
| \*ɡowɡiw | ɡowziw |  |  | \*ɡowɡiw | \*ɡowɡiw | ɡowɡiw | ɡowɡiw |  |  |  |  |  | ‘crab shell’ |
| \*ɡuk | ɡuk | ɡuk | ɡuk | \*ɡuk | \*ɡuk | ɡuk | ɡuk | \*ɡuk |  | \*ɡuk | ɡuk | ɡuk | ‘swallow’ |
| \*ɡurikŋ | ɡurik |  |  | \*ɡurikŋ | \*ɡurikŋ | ɡuriŋ | ɡuriŋ | \*ɡurik | ɡurik | \*ɡurik |  | ɡurik | ‘striped’ |
| \*ina | ina | ina | ina | \*ina | \*ina | ina |  | \*ina | ina | \*ina | ina | ina | ‘mother’ |
| \*iŋɛp | iɲɛp | iŋɛp | iŋɛp | \*iŋɛp | \*inɛp | inɛp | inɛp | \*inɛp | inɛp | \*inɛp | inɛp | inɛp | ‘shade’ |
| \*iŋtɛkŋ | iktɛk |  |  | \*iŋtɛkŋ | \*iŋtɛkŋ | ittɛŋ | ittɛt | \*iŋtɛk |  | \*iŋtɛk |  | iktɛk | ‘bowstring’ |
| \*ɟajɟaj | zajzaj | ɟajɟaj | ɟajɟaj | \*ɟajɟaj | \*ɟajɟaj |  |  | \*ɟajɟaj |  | \*ɟajɟaj |  | ɟajɟaj | ‘similar’ |
| \*ɟiʔɟiʔ | ziʔziʔ | ɟiʔɟiʔ |  | \*ɟiʔɟiʔ | \*ɟiʔɟiʔ |  |  | \*ɟiʔɟiʔ | ɟiʔɟiʔ | \*ɟiʔɟiʔ |  | ɟiʔɟiʔ | ‘baby girl’ |
| \*ɟɔʔɟɔʔ | d̪ɔʔd̪ɔʔ | ɟɔʔɟɔʔ | ɟɔʔɟɔʔ | \*ɟɔʔɟɔʔ | \*ɟɔʔɟɔʔ |  | ɟɔʔɟɔʔ | \*ɟɔʔɟɔʔ |  | \*ɟɔʔɟɔʔ | ɟɔʔɟɔʔ | ɟɔʔɟɔʔ | ‘dog’ |
| \*kabej | kabej | kabej | kabej | \*kabej | \*kabej | kabej | kabej | \*kabej | kabej | \*kabej | kabej | kabej | ‘hand’ |
| \*kalipew | kalipew | kalipew | kalipew | \*kalipow | \*kalipow |  | kalipew[[42]](#footnote-42) | \*kalipew | kalipow | \*kalipow | kalipow | kalipow | ‘forget’ |
| \*kasew | katew | kasew | kasew | \*kasow | \*kasow | kasow |  | \*kasow | kasow | \*kasow |  | kasow | ‘rafters’ |
| \*katɛt | katɛʔ |  | katɛt | \*katɛt | \*katɛt |  | katɛt | \*katɛt |  | \*katɛt |  | katɛt | ‘love’ |
| \*kejkej | sejsej | kejkej | kejkej | \*kejkej | \*kejkej | kejkej |  | \*kejkej | kejkej | \*kejkej | kejkej |  | ‘sacred’ |
| \*kɛmbuʔ | sɛbbuʔ | kɛmbuʔ | kɛmbuʔ | \*kɛmbuʔ | \*kɛbbuʔ | kɛbbuʔ | kɛbbuʔ | \*kɛbbuʔ | kɛbbuʔ | \*kɛbbuʔ | kɛbbuʔ | kɛbbuʔ | ‘older sibling’ |
| \*kɛmpa | sɛppa | pɛka | kɛmpa | \*kɛmpa | \*kɛppa | kɛppa |  | \*kɛppa | kɛppa | \*kɛppa |  | kɛppa | ‘armpit’ |
| \*kilat | silat |  |  | \*kilat | \*kilat | kilat | kilat | \*kilat | kilat |  |  |  | ‘cliff’ |
| \*kirit |  |  |  | \*kirit | \*kirit |  |  | \*kirit |  |  |  | kirit | ‘grate’ |
| \*kiniw | siɲiw | kiniw | kiniw | \*kiniw | \*kiniw | kiniw | kiniw | \*kiniw | kiniw | \*kiniw | kiniw | kiniw | ‘turmeric’ |
| \*kisej | sitej |  | kisej | \*kisej | \*kisej | kisej | kisej | \*kisej | kisej | \*kisej |  | kisej | ‘surprised’ |
| \*kɔlɛ | kɔlɛ | kɔlɛ | kɔlɛ | \*kɔlɛ | \*kɔlɛ | kɔlɛ | kɔlɛ | \*kɔlɛ | kɔlɛ | \*kɔlɛ | kɔlɛ | kɔlɛ | ‘sugarcane’ |
| \*kɔmbajʔ | kɔmbajʔ | kɔmbajʔ | kɔmbajʔ | \*kɔmbajʔ | \*ɡɔbbajʔ | ɡɔbbajʔ | ɡɔbbajʔ | \*kɔmbajʔ |  | \*kɔmbajʔ | kɔmbajʔ | kɔmbajʔ | ‘widow’ |
| \*kɔpm | kɔp | kɔm | kɔm | \*kɔm | \*kɔm | kɔm | kɔm | \*kɔp |  | \*kɔp | kɔp | kɔp | ‘eat |
| \*kɔpukŋ | kɔpuk |  |  | \*kɔpukŋ | \*kɔpukŋ | kɔpuŋ | kɔpuk | \*kɔpuk | kɔpuk | \*kɔpuk |  | kɔpuk | ‘aromatic ginger (*kaempferia galanga*)’ |
| \*kua | kwa | kua | kua | \*kua | \*kua | kua | kua | \*kua | kua | \*kua | kua | kua | ‘say’ |
| \*kukru | kukru |  | kukru | \*kukru | \*kukru | kukru |  | \*kukru | kukru | \*kukru | kukru | kukru | ‘chase’ |
| \*laʔlaʔ | laʔlaʔ | laʔlaʔ | laʔlaʔ | \*laʔlaʔ |  |  |  | \*laʔlaʔ |  | \*laʔlaʔ |  | laʔlaʔ | ‘sea slug’ |
| \*lajkɛtn | lajsɛn | lajkɛn | lajkɛn | \*lajkɛn | \*lajkɛtn | lajkɛn | lajkɛn | \*lajkɛt |  |  |  | lajkɛt | ‘seed’ |
| \*laɟɛ | lad̪ɛ | laɟɛ | laɟɛ | \*laɟɛ | \*laɟɛ | laɟɛ | laɟɛ | \*laɟɛ | laɟɛ | \*laɟɛ | laɟɛ | laɟɛ | ‘hungry’ |
| \*laɟɔ | lad̪ɔ | laɟɔ | laɟɔ | \*laɟɔ | \*laɟɔ | laɟɔ | laɟɔ | \*laɟɔ | laɟɔ | \*laɟɔ | laɟɔ | laɟɔ | ‘sail’ |
| \*laɟuʔ | lad̪uʔ | laɟuʔ | laɟuʔ | \*laɟuʔ | \*laɟuʔ | laɟuʔ | laɟuʔ | \*laɟuʔ | laɟuʔ | \*laɟuʔ | laɟuʔ | laɟuʔ | ‘wither’ |
| \*lalɛpm | lalɛp |  | lalɛm | \*lalɛpm | \*lalɛm |  |  | \*lalɛp | lalɛp | \*lalɛp | lalɛp | lalɛp | ‘house’ |
| \*laɲɟaw | laɟɟaw |  |  | \*laɲɟaw | \*laɲɟaw | laɲɟaw |  |  |  |  |  |  | ‘machete’ |
| \*lappɛɡu | lappew |  |  | \*lappɛɡu | \*lappɛɡu | lappɛɡu | lappɛɡu | \*lappɛɡu |  | \*lappɛɡu | lappɛɡu | lappɛɡu | ‘bile’ |
| \*lawlaw |  |  | lawlaw | \*lawlaw | \*lawlaw | lawlaw | lawlaw |  |  |  |  |  | ‘shiver’ |
| \*lɛɡew | lɛzew |  | lɛɡew | \*lɛɡew | \*lɛɡew |  | lɛɡew | \*lɛɡew | lɛɡew | \*lɛɡew | lɛɡew | lɛɡew | ‘dry season’ |
| \*lɛɡɡut | lɛɡɡut | lɛɡɡut | lɛɡɡuʔ | \*lɛɡɡuʔ | \*lɛɡɡuʔ | lɛɡɡuʔ | lɛɡɡuʔ | \*lɛɡɡuʔ | lɛɡɡuʔ | \*lɛɡɡuʔ | lɛɡɡuʔ | lɛɡɡuʔ | ‘mosquito’ |
| \*lɛlɛŋɡu | lɛlɛŋɡu | lɛlɛŋɡu | lɛlɛŋɡu | \*lɛlɛŋɡu | \*lɛlɛɡɡu | lɛlɛɡɡu | lɛlɛɡɡu | \*lɛlɛɡɡu | lɛlɛɡɡu | \*lɛlɛɡɡu | lɛlɛɡɡu | lɛlɛɡɡu | ‘thunder’ |
| \*lɛlɛp | lɛlɛp |  | lɛlɛp | \*lɛlɛp | \*lɛlɛp | lɛlɛp | lɛlɛp | \*lɛlɛp | lɛlɛp | \*lɛlɛp | lɛlɛp | lɛlɛp | ‘submerged’ |
| \*lɛlew | lɛlew | lɛlew | lɛlew | \*lɛlew | \*lɛlew | lɛlew | lɛlew | \*lɛlew | lɛlew | \*lɛlew | lɛlew | lɛlew | ‘forested hill’ |
| \*lɛpaʔ |  |  | lɛpaʔ |  | \*lɛpaʔ |  |  |  \*lɛpaʔ | lɛpaʔ | \*lɛpaʔ | lɛpaʔ | lɛpaʔ | PERF |
| \*lila | lila | lila | lila | \*lila | \*lila |  | lila | \*lila | lila | \*lila | lila | lila | ‘tongue’ |
| \*lɔttɛk |  |  |  | \*lɔttɛk | \*lɔttɛk |  |  | \*lɔttɛk |  | \*lɔttɛk | lɔttɛk | lɔttɛk | ‘mud’ |
| \*luɡa | luɡa |  |  | \*luɡa | \*luɡa | luɡa | luɡa | \*luɡa | luɡa | \*luɡa | luɡa | luɡa | ‘oar’ |
| \*lujkuʔ | lujkuʔ | lujkuʔ | lujkuʔ | \*lujkuʔ | \*lujkuʔ | lujkuʔ | lujxuʔ | \*lujkuʔ |  | \*lujkuʔ |  | lujkuʔ | ‘seed’ |
| \*lukŋlukŋ | lukluk |  | lukluk | \*lukŋlukŋ | \*lukŋlukŋ | lukluk | lukluk |  |  |  |  |  | ‘chicken meat’ |
| \*mantɛman | mantɛman | antɛman | antɛman | \*antɛman | \*antɛman | attɛman | attɛman | \*attɛman | attɛman | \*attɛtɛman | attɛman | attɛman | ‘beauty spot’ |
| \*mɛnmɛn |  |  | mɛnmɛn | \*mɛnmɛn | \*mɛnmɛn |  |  | \*mɛnmɛn | mɛnmɛn |  |  | mɛnmɛn | ‘saliva’ |
| \*mɛŋmɛŋ |  |  | mɛŋmɛŋ | \*mɛŋmɛŋ | \*mɛŋmɛŋ |  |  | \*mɛŋmɛŋ | mɛŋmɛŋ |  |  | mɛŋmɛŋ | ‘silent’ |
| \*muŋɡej | muŋɡej |  | muŋɡej | \*muŋɡej | \*buɡɡej | buɡɡej |  | \*buɡɡej | buɡɡej | \*buɡɡej | buɡɡej | buɡɡej | ‘beach’ |
| \*ŋamut | ŋamut | ŋamut | ŋamut | \*ŋamut | \*ŋamut | ŋamut | ŋamut | \*ŋamut |  | \*ŋamut |  | ŋamut | ‘dream’ |
| \*nanam | nanam | nanam | nanam | \*nanam | \*nanam | nanam | nanam | \*nanam | nanam | \*nanam | nanam | nanam | ‘tasty’ |
| \*nɛmnɛm |  |  | nɛmnɛm | \*nɛmnɛm |  |  |  | \*nɛmnɛm | nɛmnɛm | \*nɛmnɛm |  | nɛmnɛm | ‘sunken’ |
| \*nɛŋɡɛ | nɛŋɡɛ |  |  | \*nɛŋɡɛ | \*ɛɡɡɛ | ɛɡɡɛ | ɛɡɡɛ | \*ɛɡɡɛ | ɛɡɡɛ |  |  |  | ‘wait’ |
| \*nimpi |  |  | nimpi | \*nimpi | \*nippi |  |  | \*nippi |  | \*nippi | nippi | nippi | ‘thin’ |
| \*nusa | nuta |  | nusa | \*nusa | \*nusa |  |  | \*nusa | nusa | \*nusa | nusa | nusa | ‘island’ |
| \*ŋaŋa | ŋaŋa | ŋaŋa | ŋaŋa | \*ŋaŋa | \*ŋaŋa | ŋaŋa | ŋaŋa | \*ŋaŋa | ŋaŋa | \*ŋaŋa | ŋaŋa | ŋaŋa | ‘mouth’ |
| \*newnew | newnew |  | newnew | \*nownow | \*nownow |  | nownow | \*nownow |  | \*nownow | nownow | nownow | ‘propose’ |
| \*ŋɛtŋɛt | ɲɛtɲɛt |  |  | \*ŋɛtŋɛt | \*ŋɛtŋɛt |  | ŋɛtŋɛt | \*ŋɛtŋɛt |  |  |  |  | ‘narrow’ |
| \*niʔniʔ | ɲiʔɲiʔ |  | niʔniʔ | \*niʔniʔ | \*niʔniʔ | niʔniʔ | niʔniʔ |  |  |  |  |  | ‘frugal’ |
| \*ŋitŋit | ŋitŋit |  | ŋitŋit | \*ŋitŋit | \*ŋitŋit | ŋitŋit | ŋitŋit | \*ŋitŋit | ŋitŋit | \*ŋitŋit | ŋitŋit | ŋitŋit | ‘ghat’ |
| \*ŋɔttɔt | ŋɔttɔt |  | ŋɔttɔt | \*ŋɔttɔt | \*ŋɔttɔt | ŋɔttɔt | ŋɔttɔt | \*ŋɔttɔt |  | \*ŋɔttɔt |  | ŋɔttɔt | ‘gather flowers’ |
| \*ɔap |  |  | ɔap | \*ɔap | \*ɔap |  |  | \*ɔap |  | \*ɔap | ɔap | ɔap | ‘yawn’ |
| \*ɔɟuʔ | ɔd̪uʔ | ɔɟuʔ | ɔɟuʔ | \*ɔɟuʔ | \*ɔɟuʔ | ɔɟuʔ | ɔɟuʔ | \*ɔɟuʔ | ɔɟuʔ | \*ɔɟuʔ | ɔɟuʔ | ɔɟuʔ | ‘tide’ |
| \*ɔmbuŋ | ɔmbuŋ |  | ɔmbuŋ | \*ɔmbuŋ | \*ɔmbuŋ | ɔbbuk | ɔbbuŋ | \*ɔbbuk | ɔbbuk | \*ɔbbuk | ɔbbuk | ɔbbuk | ‘thin bamboo’ |
| \*ɔmpu |  | ɔppu | ɔmpu | \*ɔmpu | \*ɔppu | ɔppu | ɔppu | \*ɔmpu | ɔppu | \*ɔppu |  | ɔppu | ‘bind’ |
| \*ɔɲcutn | ɔtcut |  | ɔɲcun | \*ɔɲcutn | \*ɔtcut | ɔtcut | ɔtcut | \*ɔtcut | ɔtcut | \*ɔtcut | ɔtcut | ɔtcut | ‘coals’ |
| \*ɔtu | ɔtu | ɔtu | ɔtu | \*ɔtu | \*ɔtu | ɔtu | ɔtu | \*ɔtu | ɔtu | \*ɔtu | ɔtu | ɔtu | ‘hundred’ |
| \*pana | pana |  | pana | \*pana | \*pana |  |  | \*pana |  | \*pana | pana | pana | ‘arrow’ |
| \*patpat | patpat |  | patpat[[43]](#footnote-43) | \*patpat | \*patpat |  |  | \*patpat | patpat | \*patpat | patpat | patpat | ‘closed’ |
| \*pɛrɛpm | pɛrɛm | pɛrɛm | pɛrɛm | \*pɛrɛpm | \*pɛrɛpm | pɛrɛm | pɛrɛm | \*pɛrɛp | pɛrɛp | \*pɛrɛp | pɛrɛp | pɛrɛp | ‘sleep’ |
| \*pewla | pewla | pewla | pewla | \*powla | \*powla | powla | powla | \*powla | powla | \*powla | powla | powla | ‘nipah palm (*nypa frucitans*) |
| \*pɔrak | pɔrak | pɔrak | pɔrak | \*pɔrak | \*pɔrak | pɔrak | pɔrak | \*pɔrak | pɔrak | \*pɔlak[[44]](#footnote-44) | pɔlak | pɔlak | ‘land’ |
| \*pɔsɔt | pɔtɔt | pɔsɔt | pɔsɔt |  |  |  |  |  |  |  |  |  | ‘sunken’ |
| \*pusew | putew | pusew | pusew | \*pusow | \*pisow | pisow | pisow | \*pusow | pusow | \*pusow | pusow | pusow | ‘navel’ |
| \*rɛddɛt | rɛddɛt | rɛddɛt | rɛddɛt | \*rɛddɛt | \*rɛddɛt | rɛddɛt | rɛddɛt | \*rɛddɛt | rɛddɛt | \*rɛddɛt | rɛddɛt | rɛddɛt | ‘obedient’ |
| \*rɛŋɡew | rɛɡɡew | rɛŋɡew | rɛŋɡew | \*rɛŋɡew | \*rɛɡɡew | rɛŋɡew | rɛɡɡew | \*rɛɡɡew |  | \*rɛɡɡew | rɛɡɡew | rɛɡɡew | ‘narrow’ |
| \*rewrew | rewrew | rewrew | rewrew | \*rowrow | \*rowrow | rowrow | rowrow | \*rowrow | rowrow | \*rowrow | rowrow | rowrow | ‘hunt’ |
| \*rɔpm | rɔp | rɔm | rɔm | \*rɔm |  |  |  | \*rɔp | rɔp | \*rɔp | rɔp | rɔp | ‘help’ |
| \*saba | taba | saba | saba | \*saba | \*saba | saba | saba | \*saba | saba | \*saba | saba | saba | ‘python’ |
| \*sakaj |  |  |  | \*sakaj | \*sakaj |  |  | \*sakaj |  | \*sakaj |  | sakaj | ‘ascend’ |
| \*saki | tasi | saki | saki | \*saki | \*saki | saki | saki | \*saki | saki | \*saki | saki | saki | ‘buy’ |
| \*saɡu | taɡu |  | saɡu | \*saɡu | \*saɡu | saɡu |  | \*saɡu | saɡu | \*saɡu | saɡu | saɡu | ‘sago’ |
| \*sapew | tapew | sapew |  | \*sapow | \*sapow |  |  | \*sapow | sapow | \*sapow |  | sapow | ‘farmhut’ |
| \*sara |  |  |  | \*sara | \*sara |  |  | \*sara | sara | \*sara | sara | sara | ‘one’ |
| \*sɛʔsɛʔ | tɛʔtɛʔ |  | sɛʔsɛʔ | \*sɛʔsɛʔ | \*sɛʔsɛʔ |  |  | \*sɛʔsɛʔ | sɛʔsɛʔ | \*sɛʔsɛʔ | sɛʔsɛʔ | sɛʔsɛʔ | ‘k.o. grass’ |
| \*sewʔ | tewʔ | sewʔ | sewʔ | \*sewʔ | \*sewʔ | sewʔ | sewʔ | \*sewʔ | sewʔ | \*sewʔ | sewʔ | sewʔ | ‘cook’ |
| \*siba | siba | siba | siba | \*siba | \*siba |  |  | \*siba | siba | \*siba | siba | siba | ‘nine’ |
| \*siɡɛpm | sizɛp |  | siɡɛm | \*siɡɛpm | \*siɡɛp |  | siɡɛm | \*siɡɛp | siɡɛp | \*siɡɛp | siɡɛp | siɡɛp | ‘ant’ |
| \*silaʔ |  | silaʔ | silaʔ | \*silaʔ | \*silaʔ | silaʔ | silaʔ | \*silaʔ | silaʔ | \*silaʔ | silaʔ | silaʔ | ‘split’ |
| \*sɔɡaj | tɔɡaj | sɔɡaj | sɔɡaj | \*sɔɡaj | \*sɔɡaj | sɔɡaj | sɔɡaj | \*sɔɡaj | sɔɡaj | \*sɔɡaj | sɔɡaj | sɔɡaj | ‘call over’ |
| \*sɔksɔk | tɔktɔk |  |  | \*sɔksɔk | \*sɔksɔk | sɔksɔk | sɔksɔk | \*sɔksɔk |  | \*sɔksɔk | sɔksɔk | sɔksɔk | ‘change direction’ |
| \*sɔtn | tɔt | sɔt | sɔn | \*sɔtn | \*sɔtn |  | sɔn | \*sɔt | sɔt | \*sɔt | sɔt | sɔt | ‘tooth’ |
| \*sow | tow | sow | sow | \*sow | \*sow | sow | sow | \*sow | sow | \*sow | sow | sow | ‘cry’ |
| \*sulu | tulu | sulu | sulu | \*sulu | \*sulu | sulu | sulu | \*sulu | sulu | \*sulu | sulu | sulu | ‘sun’ |
| \*suŋkaj | tukkaj | sukkaj | suŋkaj | \*suŋkaj | \*sukkaj | sukkaj | sukkaj | \*sukkaj | sukkaj | \*sukkaj |  | sukkaj | ‘lever’ |
| \*suŋkra | tuŋkra | sukkra | suŋkra | \*suŋkra | \*sukkra |  | sukkra | \*sukkra |  | \*sukkra |  | sukkra | ‘dig’ |
| \*taʔ | taʔ | taʔ | taʔ | \*taʔ | \*taʔ | taʔ | taʔ | \*taʔ | taʔ | \*taʔ | taʔ | taʔ | NEG |
| \*tabɛ | tabɛ |  | tabɛ | \*tabɛ | \*tabɛ |  |  | \*tabɛ |  | \*tabɛ |  | tabɛ | ‘fat’ |
| \*tajʔtajʔ | tajʔtajʔ | tajʔtajʔ | tajʔtajʔ | \*tajʔtajʔ | \*tajʔtajʔ | tajʔtajʔ | tajʔtajʔ | \*tajʔtajʔ | tajʔtajʔ | \*tajʔtajʔ |  | tajʔtajʔ | ‘set alight’ |
| \*tandaʔ | taddaʔ | tandaʔ | tandaʔ | \*tandaʔ | \*taddaʔ | taddaʔ | taddaʔ | \*taddaʔ | taddaʔ | \*taddaʔ |  | taddaʔ | ‘sole of foot’ |
| \*taŋkew |  |  | taŋkow | \*taŋkew | \*takkow |  |  | \*takkow |  | \*takkow | takkow | takkow | ‘steal’ |
| \*tɛkaʔ | tɛkaʔ | tɛkaʔ | tɛkaʔ | \*tɛkaʔ | \*tɛkaʔ | katɛʔ | tɛxaʔ | \*tɛnaʔ | tɛnaʔ | \*tɛnaʔ |  | tɛnaʔ | ‘caw (of chicken)’ |
| \*tiʔtiʔ | siʔsiʔ | tiʔtiʔ | tiʔtiʔ | \*tiʔtiʔ | \*tiʔtiʔ | tiʔtiʔ | tiʔtiʔ | \*tiʔtiʔ | tiʔtiʔ | \*tiʔtiʔ | tiʔtiʔ | tiʔtiʔ | ‘tattoo’ |
| \*tiɡɡɔt | siɡɡɔt | tiɡɡɔt | tiɡɡɔt | \*tiɡɡɔt | \*tiɡɡɔt | tiɡɡɔt | tiɡɡɔt | \*tiɡɡɔt |  | \*tiɡɡɔt |  | tiɡɡɔt | ‘hang’ |
| \*timbɔ |  |  | timbɔ | \*timbɔ | \*timbɔ |  |  | \*timbɔ |  | \*timbɔ |  | timbɔ | ‘smoke’ |
| \*tinaj | sinaj | tinaj | tinanaj | \*tinanaj | \*tinanaj |  |  | \*tinanaj |  | \*tinanaj | tinanaj | tinanaj | ‘intestine’ |
| \*tinali | sinali |  | tinali | \*tinali |  |  |  |  |  |  |  |  | ‘frangipani’ |
| \*tinɔbutn | sinɔbun |  | tinɔbut | \*tinɔbutn | \*tinɔbutn | tinɔbun | ɔbun[[45]](#footnote-45) | \*tinɔbut | tinɔbut | \*kinɔbut[[46]](#footnote-46) |  | kinɔbut | ‘fog’ |
| \*titi | sisi | titi | titi | \*titi |  |  |  |  |  |  |  |  | ‘sago leaf’ |
| \*tɔlatn | tɔlat | tɔlan | tɔlan | \*tɔlatn | \*tɔlatn | tɔlan | tɔlan | \*tɔlat | tɔlat | \*tɔlat | tɔlat | tɔlat | ‘bone’ |
| \*tɔm | tɔp | tɔm | tɔm | \*tɔm | \*tɔm | tɔm | tɔm | \*ɔm | ɔm | ɔm | ɔm | ɔm | ‘patient’ |
| \*tuktuk | tuktuk | tuktuk | tuktuk | \*tuktuk | \*tuktuk | tuktuk | tuktuk |  |  |  |  |  | ‘fingertip’ |
| \*tumpaj | tuppaj | tuppaj | tumpaj | \*tumpaj | \*tuppaj |  | tuppaj | \*tuppaj | tuppaj | \*tuppaj |  | tuppaj | ‘add’ |
| \*tuŋklu | d̪uŋklu | tuŋklu[[47]](#footnote-47) | tuŋklu | \*tuŋklu | \*tuŋklu |  | tukklu | \*tukklu | tukklu | \*tukklu | tukklu | tukklu | ‘push’ |
| \*tuŋew | tuɲew | tuŋow[[48]](#footnote-48) | tuŋew | \*tuŋow | \*tuŋow | tuŋow | tiŋow[[49]](#footnote-49) | \*tuŋow |  | \*tuŋow |  | tuŋow | ‘mite’ |
| \*turu |  |  |  | \*turu | \*turu |  |  | \*turu |  | \*turu |  | turu | ‘index finger’ |
| \*ukɛ | usɛ | ukɛ |  |  |  |  |  |  |  |  |  |  | ‘blue sky’ |
| \*ulew |  |  |  | \*ulow | \*ulow |  |  | \*ulow | ulow | \*ulow |  | ulow | ‘snake’ |
| \*uma | uma | uma | uma | \*uma | \*uma | uma | uma | \*uma | uma | \*uma | uma | uma | ‘longhouse’ |
| \*unɛŋ | unɛŋ |  |  | \*unɛŋ | \*unɛŋ | unɛŋ |  | \*unɛŋ | unɛŋ | \*unɛŋ |  | unɛŋ | ‘asleep’ |
| \*unew | unew |  |  | \*unow | \*unow |  |  | \*unow | unow | \*unow |  | unow | ‘pith’ |
| \*uratn | urat | uran | uran | \*uratn | \*uratn | uran | uran | \*urat | urat | \*urat |  | urat | ‘rain’ |
| \*usi |  |  |  | \*usi | \*usi |  |  | \*usi |  | \*usi |  | usi | ‘chase’ |
| \*usuʔ | usuʔ |  | usuʔ | \*usuʔ | \*usuʔ |  |  | \*usuʔ | usuʔ | \*usuʔ | usuʔ | usuʔ | ‘rib’ |
| \*utɛʔ | utɛʔ | utɛʔ | utɛʔ | \*utɛʔ | \*utɛʔ | utɛʔ | utɛʔ | \*utɛʔ | utɛʔ | \*utɛʔ | utɛʔ | utɛʔ | ‘head’ |

appendix b. proto-sumatran reconstruction of selected items

All PSUM items referred to in the main text can be found here or in Billings and McDonnell (2024). Forms outside of PMEN are drawn from Warneck (1977) (Toba Batak); Neumann (1951) (Karo Batak); Eggink (1936) (Angkola Batak); Manik (1977) (Dairi Batak); Syarfina et al. (2016) (Simalungun Batak); Tim Kamus Balai Bahasa Banda Aceh (2013) (Haloban); Candrasari and Khalsiah (2018) (Leukon); Sundermann (1905) (Nias); Kähler (1955) (Sigulai); Hazeu (1907) (Gayo); Kähler (1987) and Rajeg et al. (2025) (Old Enggano); and Anderbeck and Aprilani (2013) (Nasal). PMP forms are taken from the Austronesian Comparative Dictionary (Blust, Trussel and Smith 2023). Billings and McDonnell (2024) provide details on sound changes affecting lower branches outside of Mentawai.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| PMP[[50]](#footnote-50) | PSUM | PMEN | PB | PNBI | PCBI | Gayo | Old Enggano | Nasal | Gloss |
| \*abaw | \*abaw | \*abew |  |  |  |  |  |  | ‘large’[[51]](#footnote-51) |
| \*alaq | \*alak[[52]](#footnote-52) | \*alaʔ |  |  | \*ala[[53]](#footnote-53) |  |  |  | ‘take’ |
| \*aluja | \*[a]luɡa[[54]](#footnote-54) | \*luɡa | \*luɡa[[55]](#footnote-55) |  | \*aluxa[[56]](#footnote-56) | luɡə |  |  | ‘oar’ |
| \*ambit | \*ambit | \*ambit | \*ambit[[57]](#footnote-57) |  |  |  |  |  | ‘bring’ |
| \*bak | \*bak | \*baʔ |  |  |  |  |  |  | NEG.IMP |
| \*bayaw | \*bajaw | \*baɟaʔ[[58]](#footnote-58) | \*bajow[[59]](#footnote-59) |  | \*baja[[60]](#footnote-60) |  |  |  | ‘brother-in-law’ |
| \*dələp | \*dələp | \*lɛlɛp |  |  |  |  |  |  | ‘submerged’ |
| \*diRi | \*diri | \*diɔʔ | \*diri[[61]](#footnote-61) |  |  |  |  |  | ‘stand’ |
| \*ɡidik | \*ɡidik |  | \*ɡidik[[62]](#footnote-62) |  |  |  |  |  | ‘tickle’ |
| \*ɡidik | \*ɡitik | \*ɡid̪iʔ | \*ɡitik[[63]](#footnote-63) |  |  |  |  |  | ‘tickle’ |
| \*hasaq | \*asah | \*asa | \*asah[[64]](#footnote-64) | \*asa[[65]](#footnote-65) |  |  |  |  | ‘whet’ |
| \*ia | \*ia | \*ia | \*ia[[66]](#footnote-66) |  | \*ia[[67]](#footnote-67) |  | kia | iyo | 3SG |
| \*(i)kahu | \*əkaw | \*ɛkew | \*əŋkow[[68]](#footnote-68) |  |  | ko |  | kaw | 2SG |
| \*ina | \*ina | \*ina | \*ina[[69]](#footnote-69) | \*ina[ŋ][[70]](#footnote-70) | \*ina[[71]](#footnote-71) | inə | naẽ[[72]](#footnote-72) |  | ‘mother’ |
| \*kamu | \*kamu | \*kam | \*kamu[[73]](#footnote-73) |  |  | kam |  |  | 2PL |
| \*kasaw | \*kasaw | \*kasew | \*kasow[[74]](#footnote-74) |  |  |  |  |  | ‘rafters’ |
| \*kaway | \*kawaj | \*kabej |  |  |  |  | ʔapo[[75]](#footnote-75) |  | ‘hand’[[76]](#footnote-76) |
| \*kəmpu | \*kəmpu | \*kɛmbuʔ | \*kəmpu[[77]](#footnote-77) |  |  |  |  |  | ‘elder sibling’[[78]](#footnote-78) |
| \*kua | \*kua | \*kua |  |  |  |  |  |  | ‘say’[[79]](#footnote-79) |
| \*ləpak | \*ləpak | \*lɛpaʔ | \*ləpak[[80]](#footnote-80) | \*ləpat[[81]](#footnote-81) |  | lepak |  |  | PERF[[82]](#footnote-82) |
| \*ŋaŋa | \*ŋaŋa | \*ŋaŋa | \*ŋaŋa[[83]](#footnote-83) |  | \*ŋaŋa[[84]](#footnote-84) |  |  |  | ‘mouth’[[85]](#footnote-85) |
| \*nipis | \*nipis | \*nimpi | \*nipis[[86]](#footnote-86) | \*nihik[[87]](#footnote-87) | \*nifi[[88]](#footnote-88) | nipis |  |  | ‘thin’ |
| \*nusa | \*nusa | \*nusa |  |  | \*nusa[[89]](#footnote-89) |  |  |  | ‘island’ |
| \*panaq | \*panah[[90]](#footnote-90) | \*pana |  | \*pana[[91]](#footnote-91) | \*fana[[92]](#footnote-92) |  |  |  | ‘arrow’ |
| \*qabatəR | \*habatər | \*batra |  | \*batəl[[93]](#footnote-93) |  |  |  |  | ‘sago worm’ |
| \*qapəjux | \*pəɡu | \*lappɛɡu | \*pəɡu[[94]](#footnote-94) |  | \*fəxu[[95]](#footnote-95) |  |  |  | ‘bile’ |
| \*Rabun | \*rabun | \*tinobutn | [[96]](#footnote-96) |  |  |  |  |  | ‘fog’ |
| \*Rapus | \*rapus | \*ɔmpu | \*rapus[[97]](#footnote-97) |  | \*rafu[[98]](#footnote-98) | rapus |  |  | ‘bind’ |
| \*sakay | \*sakaj | \*sakaj |  |  |  |  |  |  | ‘ascend’ |
| \*saɡu | \*saɡu | \*saɡu | \*saɡu[[99]](#footnote-99) |  |  |  |  |  | ‘sago’ |
| \*sapaw | \*sapaw | \*sapew | \*sapow[[100]](#footnote-100) |  |  |  |  |  | ‘field hut’ |
| \*sədsəd[[101]](#footnote-101) | \*sədsəd | \*sɛʔsɛʔ |  |  |  |  | kẽkẽ |  | ‘k.o. grass’ |
| \*sida | \*sida | \*sia[[102]](#footnote-102) | \*sida | \*sira[[103]](#footnote-103) | \*ira[[104]](#footnote-104) |  | ki |  | 3PL |
| \*takaw | \*taŋkaw | \*taŋkew | \*taŋkow |  | \*taɡə |  |  |  | ‘steal’ |
| \*taq | \*tak | \*taʔ |  |  |  |  |  |  | NEG |
| \*tiktik | \*tiktik | \*tiʔtiʔ | \*tiktik[[105]](#footnote-105) |  |  |  |  |  | ‘tattoo’[[106]](#footnote-106) |
| \*t<in>aqi | \*tinahi | \*tinaj |  |  |  |  |  |  | ‘intestines’ |
| \*qusiR | \*husir | \*usi |  |  |  |  |  |  | ‘chase’ |
| \*zauq | \*[a]dauh | \*arew | \*dauh | \*daw | \*dəw |  |  |  | ‘far’ |
|  | \*adat | \*arat | \*adat[[107]](#footnote-107) |  | \*ara[[108]](#footnote-108) |  |  |  | ‘tradition’ |
|  | \*duduk | \*duruʔ |  |  | \*ruru[[109]](#footnote-109) |  |  |  | ‘gather’ |
|  | \*ŋitŋit[[110]](#footnote-110) | \*ŋitŋit | \*ŋitŋit[[111]](#footnote-111) |  |  |  |  |  | ‘gnat’[[112]](#footnote-112) |
|  | \*tandak | \*tandaʔ | \*tandak[[113]](#footnote-113) |  |  |  |  |  | ‘sole of feet’[[114]](#footnote-114) |
|  | \*unəŋ | \*unɛŋ | \*unəŋ[[115]](#footnote-115) |  |  |  |  |  | ‘asleep’[[116]](#footnote-116) |

References

Adelaar, K. Alexander. 1981. Reconstruction of Proto-Batak phonology. In *Historical linguistics in Indonesia*, ed. Robert Blust, 1-20. Jakarta: Atma Jaya University.

———. 2005. Malayo-Sumbawan. *Oceanic Linguistics* 44.2: 357-88.

Akoli, Marcelinus Yeri F. and I Wayan Arka. 2025. Pronominal and alignment systems in Sikule. Paper presented at *Kongres Internasional Masyarakat Linguistik Indonesia,* Depok, Indonesia.

Anderbeck, Karl R. and Herdian Aprilani. 2013. *The improbable language: Survey report on the Nasal language of Bengkulu, Sumatra* (Electronic Survey Report 2013-012). Dallas: SIL International.

Billings, Blaine and Bradley McDonnell. 2024. Sumatran. *Oceanic Linguistics* 63.1:112-174.

Blust, Robert A. 1976. Dempwolff’s reduplicated monosyllables. *Oceanic Linguistics* 15.1/2:107-30.

———. 1977. The Proto-Austronesian pronouns and Austronesian subgrouping: A preliminary report. *Working Papers in Linguistics* 9.2:1-15. Honolulu: University of Hawai’i.

———. 1984. On the history of the Rejang vowels and diphthongs. *Bijdragen tot de Taal-, Land- en Volkenkunde* 140:422-50.

———. 1988. *Austronesian Root Theory: An essay on the limits of morphology*. Amsterdam/Philadelphia: John Benjamins.

———. 1997. Nasals and nasalization in Borneo. *Oceanic Linguistics* 36.1:149-179.

———. 1999. Subgrouping, circularity and extinction: Some issues in Austronesian comparative linguistics. In *Selected Papers from the Eighth International Conference on Austronesian Linguistics* (Symposium Series of the Institute of Linguistics, Academia Sinica 1). ed. Elizabeth Zeitoun and Paul J. K. Li, 31-94. Taipei: Acaemia Sinica.

———. 2000. Why lexicostatistics doesn’t work: The ‘universal constant’ hypothesis and the Austronesian languages. In *Time Depth in Historical Linguistics*, ed. Colin Renfrew, April McMahon, and Larry Trask, 311-331. Cambridge: McDonald Institute for Archaeological Research.

———. 2003. *Thao Dictionary.* Tapiei: Academia Sinica.

———. 2005. Must sound change be linguistically motivated? *Diachronica* 22.2:219-69.

———. 2010. The Greater North Borneo hypothesis. *Oceanic Linguistics* 49:44-118.

———. 2011. The problem of doubleting in Austronesian languages. *Oceanic Linguistics* 50:399-457.

———. 2013. *The Austronesian Languages*. Canberra: Pacific Linguistics.

———. 2022. Repair and drift in Austronesian languages: Avoidance of dissimilar labials as the onsets of successive syllables. *Oceanic Linguistics* 62.2: 304-23.

Blust, Robert A., Stephen Trussel, and Alexander D. Smith. 2023. CLDF dataset derived from Blust’s “Austronesian Comparative Dictionary” (v1.2) [Data set]. Zendono. https://doi.org/10.5281/zendono.7741197.

Boersma, Paul. 1997. Sound change in functional phonology. MS.

Budiono, Satwiko, Rita Novita, and Tengku Syarfina. 2023. Mentawai language variations in the Mentawai Islands regency, West Sumatra province. *Jurnal Arbitrer* 10.1:8-18.

Cambielli, Daniele. 1998. *Kamus Indonesia-Mentawai.* MS.

Candrasari, Ratri and Khalsiah. 2018. *Kamus Indonesia-Leukon.* Lhokseumawe: Sefa Bumi Persada.

Chrétien, C. Douglas. 1965. The statistical structure of the Proto-Austronesian morph. *Lingua* 14:243-270.

Crisp, John. 1799. An account of the inhabitants of the Poggy islands lying off Suamtra. *Asiatick Researches* 6:77-91.

Dempwolff, Otto. 1922. Entstehung von Nasalen und Nasalverbindungen im Ngadju (Dajak). *Zeitschrift für Eingeborenen-Sprachen* 13:161-205.

———. 1934-8. *Vergleichende Lautlehre des Austronesischen Wortschatzes*. Berlin: Vohsen.

Eades, Domenyk and John Hajek. 2006. Gayo. *Journal of the International Phonetic Association* 36.1:107-15.

Edwards, Owen. 2015. The position of Enggano within Austronesian. *Oceanic Linguistics* 54.1:54-109.

Eggink, H. J. 1936. *Angkola- en Mandailing-Bataksch-Nederlandsch Woordenboek*. Bandung: A. C. Nix & Co.

Eindhoven, Myra. 2007. New colonizers? Identity, represenation and government in the post-New Order Mentawai archipelago. In *Renegotiating Boundaries: Local politics in post-Suharto Indonesia*. ed. Henk Schulte Nordholt and Gerry van Klinken, 67-90. Leiden: KITLV.

Gil, David. 2007. Item “Mentawai” in collection “MPI EVA copora”. The Language Archive.

Hazeu, Godard Arend Johanes. 1907. *Gajōsch-Nederlandsch woordeboek*. Jakarta: Landsdrukkerij.

Kähler, Hans. 1955. *Die Sichule-Sprache auf der Insel Simalur an der Westküste von Sumatra*. Berlin: Dietrich Reimer.

———. 1987. *Enggano-Deutsches Wörterbuch*. Veröffenlichungen des Seminars für Indonesische und Südseesprachen der Universität Hamburg 14. Berlin/Hamburg: Dietrich Reimer Verlag.

Lenggang, Zainuddin H. R., Be Kim Ho Nio, Mohammad Ansyar Zainil, and Syofyan Adam. 1978. *Bahasa Mentawai*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa.

Manan, Umar, Anas Yasin, Anan Syafei, Muhammad Yamin, Bomantoro, and Jurlisman Rajab. 1984. *Kata Tugas Bahasa Mentawai*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa.

Manik, Tindi Radja. 1977. *Kamus bahasa Dairi Pakpak-Indonesia*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa.

Mendrofa, Fransiskus Yanuarius, Gabriel Terri Salakkirat, and Rob Henry. 2019. *Buko Rereiket Mattaoi-Indonesia: Kamus Rereiket Mentawai-Indonesia*. Rokdog: Yayasan Pendidikan Budaya Mentawai.

Morris, Max 1900. *Die Mentawai-Sprache*. Berlin: Conrad Skopnik.

Mullan, Keira. 2025. Person marking in Barrier Islands languages. Paper to be presented at *Australian Linguistics Society conference 2025*.

Mustafa G., Rizanur Gani, Sarwono Kartodipuro, Busri, Mohammad Atar Semi, and S. Zaura Gusmali. 1993. *Sastra Lisan Mentawai.* Jakarta: Pusat Pembinaan dan Pengembangan Bahasa.

Neumann, Joshua H. 1951. *Karo-bataks Nederlands Woordenboek.* Medan: VAREKAMP & Co.

Nothofer, Bernd. 1986. The Barrier Island languages in the Austronesian Language family. In *FOCAL II: Papers from the Fourth International Conference on Austronesian Linguistics*, ed. Geraghty, Paul, Lois Carrington, and S. A. Wurm. Canberra: Pacific Linguistics.

Pampus, Karl-Heinz. 1989. Zur dialektgeographischen Gliederung des Mentawai-Archipels. In *Mentawai: Indentität im Wandel auf Indonesischen Außeniseln*. Ed. Wilfred Wagner, 61-102. Bremen: Übersee-Museum.

———. 1994. Sabirut-Taburit ‘Siberut’: Die Lautgesetze der Simalegi-Sprache im Mentawai-Archipel (Westsumatra). *Zeitschrift der Deutschen Morgenländischen Gesellschaft* 114(1): 114-29.

Rajeg, Gede Primahadi Wijaya, Cokorda Rai Adi Pramartha, Ida Bagus Gede Sarasvanada, Putu Wahyu Widiatmika, Ida Bagus Made Ari Segara, Yul Fulgensia Rusman Pita, et al. (2015) The Enggano-German Dictionary online derived from Kähler’s (1987) “Enggano-Deutsches Wörterbuch. Online resource: University of Oxford.

Reeves, Glenn. 1999. History and ‘Mentawai’: Colonialism, scholarship and identity in the Rereiket, West Indonesia. *Australian Journal of Anthropology* 10.1:34-55.

SABDA. 1996. *Alkitab Bahasa Mentawai: Bahasa sehari-hari 1996. Perjanjian Baru*. Jakarta: Software Alkitab, Biblika, dan Alat-alat.

Sagart, Laurent. 2024. The phonetic nature of PAN \*j. *Oceanic Linguistics* 63.1:175-81.

Samaloisa, Rijel. 2020. Pemerintahan laggai paham “Arat Sabulungan” di Kabupaten Kepulauan Mentawai provinsi Sumatera Barat. *Governabilitas* 1.1:84-109.

Schefold, Reimar. 1986. The unequal brothers-in-law: Indonesia as a field of anthropological study and the case of Mentawai. *Bijdragen tot de Taal-, Land- en Volkenkunde* 142.1:69-86.

———. 1989. The origins of the woman on the raft: On the prehistory of the Mentawaians. In *Mentawai: Identität im Wandel auf Indonesisischen Außeniseln*. Ed. Wilfred Wagner, 1-26. Bremen: Übersee-Museum.

Smith, Alexander D. 2017. The Western Malayo-Polynesian problem. *Oceanic Linguistics* 56.2:435-490.

———. 2020. Nasalization in Enggano historical phonology. *Oceanic Linguistics* 59.2: 347-65.

———. 2023*a.* Constraint reranking in diachronic OT: Binary-feet and word-minimum phenomena in Austronesian. *Journal of East Asian Linguistics* 32:341-72.

———. 2023*b.* Reconstructing non-contrastive stress in Austronesian and the role of the mora in stress shift, gemination, and vowel shift. *Diachronica* 40.1:111-52.

———. 2024*a*. Proto-Austronesian \*j once again. *Oceanic Linguistics* 63.2:319-337.

———. 2024*b.* Proto-Malayo-Polynesian: Its place within the Austronesian language family, reconstruction, and daughters. In *The Oxford Guide to the Malayo-Polynesian langauges of Southeast Asia*, eds. Alexander Adelaar and Antoinette Schapper, 9-24. Oxford: Oxford University Press.

Sundermann, Heinrich D. 1905. *Niassich-Deutcshes Wörterbuch.* Moers: J. W. Spaarmann.

Syarfina, Tengku, Amrin Saragih, Amran Purba, Zufri Hidayat, Anharuddin Hutasuhut, Chairani Nasution, Sri Asrianti and Juliana. 2016. *Kamus bahasa Simalungun-Indonesia*. Medan: Badan Pengembangan dan Pembinaan Bahasa.

Tim Kamus Balai Bahasa Banda Aceh. 2013. *Kamus bahasa Haloban*. Banda Aceh: Balai Bahasa Banda Aceh.

Thurgood, Graham. 1999. From Ancient Cham to modern dialects: Two thousand years of language contact and change, with an appendix of Chamic reconstructions and loanwords. Oceanic Linguistics Special Publications, no. 28. Honolulu: University of Hawai’i Press.

Tulius, Juniator. 2012*a*. Family stories: Oral tradition, memories of the past, and contemporary conflicts over land in Mentawai – Indonesia. PhD thesis, Leiden University.

———. 2012*b*. Stranded people: Mythical narratives about the first inhabitants of Mentawai islands. *Wacana, Journal of the humanities of Indonesia.* 14.2:215-240.

Warneck, Johannes. 1977. *Toba Batak-Deutsches Wörterbuch*. The Hague: Koninklijk Institut voor Taal-, Land- en Volkenkunde.

Yoder, Brendon E. 2014. Phonological and phonetic aspects of Enggano vowels (SIL eBook 62). Dallas: SIL International.

Yoo, Kayeon and Francis Nolan. 2020. Sampling the progression of domain-initial denasalization in Seoul Korean. Laboratory Phonology: Journal of the Association for Laboratory Phonology 11.1:1-32.

1. . I am greatly indebted to the speakers of the various Mentawai dialects who have provided wordforms for this paper: Sumario “Mario” Tatubeket (Sipora); Srikandi “Cici” Putri and Edo Saleleubaja (Sabirut); Mona Sagurung (Sikabaluan); Eni Magdalena (Simatalu); and Brygita Taelegat (Simalegi), for sitting through several hours of targeted questioning about the meanings and pronunciations of hundreds of often obscure words with what I can only imagine must have seemed like no rhyme or reason. I must also thank Hieronimus “Bang Nimus” Eko Zebua, Markolinus Sagulu, and everyone at *FORMMA (Forum Mahasiswa Mentawai)* for bringing together, at such short notice, a group spanning such diversity of geographically distant varieties in one place. Enormous thanks are also due to Pastor Daniele Cambielli and the Catholic diocese in Padang for allowing access to Cambielli’s unpublished mutlidialectal dictionary of Mentawai, an unparalleled source of Mentawai dialectology without which this reconstruction simply would not have been possible; and to Zoe Orlanda Elwyn, Keira Mullan, and Yosep Sagari for assisting with acquiring copies of the Karo Batak, Sigulai, and Rereiket Mentawai materials respectively. I am grateful also to Blaine Billings and Floofy McFloofkins for comments. This research was supported by the Australian Resarch Council grant *Languages of Barrier Islands, Sumatra: Description, Typology, and History* (DP230102019), and an Australian Government Research and Training Program. Any errors in the text remain my own. [↑](#footnote-ref-1)
2. Earlier sources use the names *Chagalelegat*, *Tschagallelegat* (Reeves 1999). [↑](#footnote-ref-2)
3. The interested reader is directed to the several existing grammatical materials published by the *Pusat Bahasa* on ‘Sikakap’ Mentawai (Lenggang et al. 1984; Manan et al., 1984; inter alia); the translation of the bible (SABDA 1996), which reflects Sikakap/Pagai Mentawai; and Mustafa G. et al.’s (1993) text collection, which explicitly treats Sipora and Pagai as one. All of these confirm the impression that ‘Sipora-Pagai’ Mentawai can be sensibly treated as a single language with at most dialectal differences, for example in the intonation. In general, extant materials in ‘Mentawai’ which do not specify the location further typically reflect varieties of this language. [↑](#footnote-ref-3)
4. I am greatful to Rita Novita for the provision of this wordlist, which confirms my statement that the Mentawai of Mongan Poula can be considered representative of Sikabaluan Mentawai. [↑](#footnote-ref-4)
5. Though there is no official orthography for any of the Mentawai languages the vast majority of speakers nowadays are literate in Indonesian and many write their own languages in idiosyncratic *ad hoc* orthographies influenced by the Indonesian orthography; syllable-final plosives are in free variation between graphemes indicating voiceless and voiced equivalents in writing, suggesting these are underspecified for voice in the minds of speakers rather than being phonemically voiceless. Furthermore, ‘continuants and voiced plosives’ is not a well-defined natural class to be excluded from appearing in the syllable coda (2.2.2), suggesting that rather than there being a phonotactic prohibition on voiced plosives in syllable codas, as with continuants, voiced plosives are unable to appear there through reasons of noncontrastiveness rather than prohibition. The merger in 4.2.2, whereby PSUM voiced plosives became voiceless through the necessary contrast reduction, proves this point. [↑](#footnote-ref-5)
6. Of the few quadrisyllabic lemmas in Mentawai languages that are not transparently loans, at least one seems to the result of a PMP compound becoming lexicalized: PMP \**qali-pətpət* > PMEN \**alupɛtpɛt*,‘firefly’, and another is a transparent lexicalization of a compound from an earlier stage of Mentawai: *ata* ‘long’ + *paipai* ‘posterior’ > PMEN \**atapaipai* ‘Mentawai langur; *presbytis potenziani*’*.* Others are less transparent in their origins but likely arose through similar processes. [↑](#footnote-ref-6)
7. I.e., those which reflect PMEN pre-ploded nasals, cf. later sections. [↑](#footnote-ref-7)
8. This assimilation affects VOT but has no phonemic reality, since syllable-final plosives are underspecified for voice. [↑](#footnote-ref-8)
9. Mendrofa, Salakkirat and Henry (2019:8) describe this as fricative /ð/, but my short video corpus confirms this to be /d̪/, as in other Mentawai languages with have a voiced dental consonant. [↑](#footnote-ref-9)
10. Karl-Heinz Pampus (cited as p.c. in Blust 1997) claims that word-final plosives in Paipajet Mentawai, even those reflecting PMEN plosives, are replaced with nasals in careful speech; Cambielli (1998) does not attest this and it is not treated here. [↑](#footnote-ref-10)
11. The root here has previously (Billings and McDonnell 2024:160; Blust 1997:159) as having an initial *m-*, but it has *p-*; in REALis main clauses this is converted to *m-* by the affix process *m>p*, which appears on voice morphology in Mentawai languages in such clauses; since this is one of a few unaccusative verbs in Mentawai languages which may occur without voice morphology (others: *bele’* ‘fall’; *matei* ‘to be dead’) this affix process occurs directly on the stem. It is the only member of this unaccusative voice-free verb subclass I have been able to identify which begins with *p-*. This also makes the change PSUM \*pədəm > PMEN \*pɛrɛpm, previously with unexplained \*p>\*m, fully regular. [↑](#footnote-ref-11)
12. One possibility, which seems to accord with both the linguistic evidence and the general reputation of the Simalegians as the most isolated and indeed isolation*ist* group among the Mentawai peoples, is that all of the Mentawai languages outside of Simalegi existed as a *sprachbund* for most of their history, with widespread mutlilingualism and contact between different Mentawai languages discouraging sound change. Simalegi Mentawai could have emerged as a breakaway from this *sprachbund* at some point. This fits with the general impressions that 1) Simalegi Mentawai has changed a great deal phonologically since the age of PMEN; 2) all other Mentawai languages have changed remarkably little in the same time, considering the time-depth at hand; 3) there is a strong sense throughout Mentawai folklore and historical myth that settlement began at Simatalu, but Simatalu Mentawai does not diverge from PMEN to anywhere near the level of Simalegi Mentawai. [↑](#footnote-ref-12)
13. If syllable-final PSUM nasals in reduplicative lemmas became preploded in PMEN, we would still not expect to see nasal-stop clusters emerging allophonically in modern Mentawai languages, since reduplicated monosyllables necessarily have a contentful onset, leaving no slot for the plosive segment to be ‘pushed out’ into. Variation between nasal and voiceless plosives syllable-finally in reduplicative lemmas adequately demonstrates that nasal preplosion did also occur here in PMEN, meaning both ‘word-final’ changes are also ‘root-final’. [↑](#footnote-ref-13)
14. Almost certainly a fossillization of \*ka- ‘adversative passive’ (Blust 2013) + \*lipaw (?) ‘forget’ – doublet of \*lupa? [↑](#footnote-ref-14)
15. Mentawai *lelew* is frequently translated (as in Cambielli 1998; Lenggang et al. 1978) as both ‘forest’ and ‘hill’; the actual meaning is closer to ‘forested hill’: the kind of semi-impenetrable hilly bushland that is ubiquitous on the Mentawais. The term is frequenly used to contrast with navigable paths as a sort of ‘elsewhere’ biome – compare the use of the word ‘bush’ in regional Australian English. [↑](#footnote-ref-15)
16. \*abew and \*arew are not historically monosyllables. However, they are both stative verbs expressing ‘immutable’ states, which in Mentawai languages obligatorily appear with stative *ma-*. Additionally, a rule of sequential identical vowel deletion causes these to surface as \*mabew and \*marew, respectively. I therefore posit that they were treated by sound change as monosyllablic despite their history. [↑](#footnote-ref-16)
17. Since the Simalegi form here seems to have had \*ɟ in this form which originally had PMP \*j, the curious reader may wonder if this is evidence that Simalegi Mentawai (and therefore PMEN) had escaped the Sumatran \*j,\*g>\*g merger, and \*ɟ is original to all these forms in PMEN. However, forms like Simalegi Mentawai *luɡa* disprove this; /ɟ/ in forms like (\*siɟɛp>)*sizɛp* really did go ‘there and back again’ to and from \*ɡ in front of \*i, in an interesting parallel to changes 3.1.8-9. [↑](#footnote-ref-17)
18. The imperative mood has zero voice morphology in Mentawai languages, but stative verbs cannot appear in the imperative mood; this makes them the only subclass of verbs to *never* appear without voice morphology. [↑](#footnote-ref-18)
19. Despite this pseudo-intervocalic environment for \*ɟɛla having triggered the sound change, in modern Simalegi Mentawai the root is still given as *zela* even in citation form without *ma-*. Thus, /z/ in Simalegi Mentawai is probably best understood as a phoneme in its own right on distributional grounds, although it is unlikely to be contrasted from /ɟ/ by any minimal pairs. [↑](#footnote-ref-19)
20. Addition of *\*laC-* is shared with \*buntər>\*lajmuntu – some fossilized morpheme? [↑](#footnote-ref-20)
21. Since C1V1C2>C2V1C1 and V1V2>V2V1 metathesis are an information-preserving transformations and there are no apparent cognates in other branches or PSUM, it is not possible to tell which of these forms is a transformation of the other. [↑](#footnote-ref-21)
22. Pampus (cited in Blust 1997:159) provides the stative verb <*mai-ŋep>*, which he takes to be reflective of earlier (PMEN?) \*ŋepm. However, this is an mal-segmentation; the lemma in PMEN is \*iŋɛp, stative \*ma-iŋɛp ‘to be dawn’. [↑](#footnote-ref-22)
23. Based on Chrétien’s (1965) classic statistical study of the PAN morph, I estimate reduplicative lemmas to be in the ballpark of 10% of the lexicon. [↑](#footnote-ref-23)
24. The principle of fewest changes suggests this was [ɤ], since this is reflected directly in PMEN and requires only one further change to reach its apparent reflexes in PCBI and Nasal: devoicing and uvularization, respectively. In this light, Nothofer’s (1986) mistaken assertion that \*j became /x/ in the proto-language of many of these same languages is more understandable than it may first seem. [↑](#footnote-ref-24)
25. Contemporary Enggano has undergone a number of rapid additional sound changes since the stage of ‘Old Enggano’ as attested by Kähler (1978), however these have little bearing on the subgrouping of Enggano since the older form naturally more closely reflects PSUM. Cf. Yoder (2014) and Edwards (2015) for descriptions of how the two differ. [↑](#footnote-ref-25)
26. It should be noted that in the wake of the deprecation of the WMP group as a first-order branch of MP, our understanding of the relation of its erstwhile subgroups (including Sumatran) to MP is not strong; Sumatran may well not be a first order branch of MP and may share this value of \*g with other branches. [↑](#footnote-ref-26)
27. Billings and McDonnell (2024:170) posit the following changes from PSUM to their ‘Pre-Mentawai’: 1) \*h>Ø; 2) \*d>l/…l; 3) s>Ø/\_#; 4) \*d>\*r/V\_V; 5) \*ɡ>\*w/V\_V; 6) \*w>\*b; 7) \*j>\*ɟ; 8) \*ə>\*e; 9) r>Ø ‘with changes to adjacent vowels’; 10) \*aw>\*ew; and 11) \*b>\*p/\_#. I accept the first five with no modifications; modify the environment or target value of 6)-9); subsume 10)-11) under broader changes targeting whole natural classes; and propose one new regular change \*k>\*ʔ/\_# in addition to nasal preplosion and several irregular but recurrent processes. [↑](#footnote-ref-27)
28. The change \*w>\*b (4.2.6) occurred first, creating the sequence <mbVb>; which would have run up against the prohibition against dissimilar nasals in the onsets of successive syllables – treating the nasal-stop cluster as a segmental unit \*mb, ‘dissimilar’ to a regular plosive \*b. The change \*b>\*ŋ may have been conditioned by the onset, pushed away from its erstwhile value but with no clear target value, imitating the coda of its host syllable, as in changes like 4.2.1. Afterwards, word-final \*ŋ changed to \*n to avoid violating the local (Mentawai-specific, possibly recent) phonotactic constraint against like nasals in the same host syllable ({ω[+nas][α place], κ[α place]}, 3.2). The ordering of changes suggests that the change in 4.2.6 occurred prior to the change in 4.2.1, since the final syllable of \*hambawaŋ did not have a nasal onset until the change \*b>\*ŋ had occurred and would thus have incurred preplosion if 4.2.1 came first. [↑](#footnote-ref-28)
29. After \*h disappeared (4.2.4), \*u was the onset of the resulting monosyllable, and subsequently desyllabified to \*w, since high vowels undergo desyllabification outside of nucleus position. However, \*w was unable to become \*b here, due to prohibition 2 in section 3.2 ({ω[+plo][α place][+voice], κ[+plo][α place]}). It thefore lowered in height to \*ɔ and became the nucleus of its own syllable, since non-high vowels do not desyllabify. [↑](#footnote-ref-29)
30. Semshift: ‘above’>’tall’. [↑](#footnote-ref-30)
31. After \*h disappeared, \*a was metathesized to word-final position. [↑](#footnote-ref-31)
32. \*b>\*m avoids prohibition 2) (3.2), since \*nt is interpreted synchronically as a (prenasalized) plosive. [↑](#footnote-ref-32)
33. \*a was inserted here before \*u to maintain the CVC structure of the reduplicative lemma root. [↑](#footnote-ref-33)
34. Blust’s (1997) translation. [↑](#footnote-ref-34)
35. The prenasalization here seems to be shared with PCBI (cf. Billings and McDonnell 2024:162) – this contributes towards evidence of subgrouping (4.3.2). [↑](#footnote-ref-35)
36. Irregular \*mp>mb. [↑](#footnote-ref-36)
37. The pronominal indexing morphemes have the same forms in all Mentawai languages so the reconstruction is trivial and not presented here. [↑](#footnote-ref-37)
38. Parallel and repeated metathesis in multiple branches; several unexplained changes. [↑](#footnote-ref-38)
39. Parallel metathesis in Simalegi and Simatalu Mentawai; one could easily be a borrowing of the other given geographic proximity. [↑](#footnote-ref-39)
40. Irregular \*ɔ>oj. [↑](#footnote-ref-40)
41. Irregular \*ew>ɔ. [↑](#footnote-ref-41)
42. Irregular \*ew. [↑](#footnote-ref-42)
43. Semshift ‘closed’>’pig cage’; shared by Rereiket Mentawai. [↑](#footnote-ref-43)
44. Irregular r>l. [↑](#footnote-ref-44)
45. Irregular syllable loss. [↑](#footnote-ref-45)
46. Irregular t>k. [↑](#footnote-ref-46)
47. Cambielli (1998:64) provides *<tungglu>*. [↑](#footnote-ref-47)
48. Irregular \*ew>ow. [↑](#footnote-ref-48)
49. Irregular \*u>i. [↑](#footnote-ref-49)
50. I use <y> for /j/ in PMP so as not to create confusion with \*j, but <j> for /j/ elsewhere. [↑](#footnote-ref-50)
51. Semshift: ‘high, lofty’>’large’. [↑](#footnote-ref-51)
52. Irregular q>k. [↑](#footnote-ref-52)
53. Sigulai *ala*. [↑](#footnote-ref-53)
54. Semshift ‘to paddle’>’oar’. [↑](#footnote-ref-54)
55. TB, KB, SB *luɡa*. [↑](#footnote-ref-55)
56. Nias *aluxa*, Sigulai *eluxa*. [↑](#footnote-ref-56)
57. TB, AB, DB *ambit*. [↑](#footnote-ref-57)
58. Seemingly reflects prior *\*bajak*, as does PCBI. [↑](#footnote-ref-58)
59. TB *bao*, DB *bajo*. [↑](#footnote-ref-59)
60. Nias *baja*. [↑](#footnote-ref-60)
61. TB, KB, AB, DB, SB *diri*. [↑](#footnote-ref-61)
62. TB, KB, AB, DB *ɡidik*; semshift ‘tickle’>’armpit’ in TB. [↑](#footnote-ref-62)
63. TB *ɡitik*; also attested in GNB so may be as old as PMP. Semshift in TB ‘tickle’>’irritate’. [↑](#footnote-ref-63)
64. TB, *asa* (irregular h>Ø), KB, SB *asah*. [↑](#footnote-ref-64)
65. Haloban *asa*. [↑](#footnote-ref-65)
66. TB, KB, AB, DB, SB *ia*. [↑](#footnote-ref-66)
67. Nias *ia.* [↑](#footnote-ref-67)
68. TB *ho*, KB *engko*, DB, SB *ko*. [↑](#footnote-ref-68)
69. TB, KB *ina*, AB, DB *inaŋ*. [↑](#footnote-ref-69)
70. Haloban *inaŋ.* [↑](#footnote-ref-70)
71. Nias, Sigulai *ina*. [↑](#footnote-ref-71)
72. Metathesis; irregular i>e (influenced by nasal environment?). [↑](#footnote-ref-72)
73. TB *hamu*, KB *kam*, SB *ham*. [↑](#footnote-ref-73)
74. KB *kaso*. [↑](#footnote-ref-74)
75. Irregular (?) \*aj>o – I am aware of only one other Old Enggano item reflecting PSUM \*aj: *bake* (Billings and McDonnell 2024:160) – this could easily be phonologically conditioned difference. [↑](#footnote-ref-75)
76. Semshift ‘to wave the hand or arms’>’hand’. [↑](#footnote-ref-76)
77. TB, SB, AB *hompu*, semshift granchild>grandparent; KB, DB *kempu*, [↑](#footnote-ref-77)
78. Semshift ‘grandchild’>’elder sibling’. [↑](#footnote-ref-78)
79. Semshift filler word>’say’; also attested in Tae’ and Makassarese. [↑](#footnote-ref-79)
80. KB, SB *lepak*; semshift ‘break’>’to be mistaken’, also seen in Gayo. [↑](#footnote-ref-80)
81. Haloban *lepat*. Semshift ‘break’>PERF, shared with PMEN. Irregular k>t. [↑](#footnote-ref-81)
82. Semshift ‘break’>PERF. [↑](#footnote-ref-82)
83. KB *ŋaŋa,* AB, DB, SB *ŋaŋaŋ*. [↑](#footnote-ref-83)
84. Nias *ŋaŋa*, semshift ‘with open mouth’>’chew’. [↑](#footnote-ref-84)
85. Semshift ‘with mouth open’>’mouth; language’. [↑](#footnote-ref-85)
86. TB, AB, DB *nipis,* KB *nipes*. [↑](#footnote-ref-86)
87. Leukon *nɛhɛk*. Irregular k>s. Numerous i~e interchanges within PNBI are listed in Billings and McDonnell (2024); these are currently poorly understood. [↑](#footnote-ref-87)
88. Nias *nifi.* [↑](#footnote-ref-88)
89. Nias *nuza.* [↑](#footnote-ref-89)
90. Semshift ‘to shoot with bow and arrow’>‘(bow and) arrow’. [↑](#footnote-ref-90)
91. Haloban *pana.* [↑](#footnote-ref-91)
92. Nias *fana.* [↑](#footnote-ref-92)
93. Reconstruction from Billings and McDonnell (2024:161). PB *badir* and Old Enggano *~paki* reflect PMP doublet \*qabatir. [↑](#footnote-ref-93)
94. TB, AB, SB *poɡu,* DB *peɡu*. [↑](#footnote-ref-94)
95. Nias *fɤxɤ*. [↑](#footnote-ref-95)
96. TB *ombun*, SB *hombun* reflect PMP doublet *\**əmbun, possibly arising through an old enigmatic nasalization. [↑](#footnote-ref-96)
97. TB *rapus*; semshift ‘bind’>’hold on’, SB *rapus*; semshift ‘bind’>’capture’. [↑](#footnote-ref-97)
98. Nias *rafɤ.* [↑](#footnote-ref-98)
99. TB, KB, AB, DB, SB *saɡu*. [↑](#footnote-ref-99)
100. KB, DB *kaso*. [↑](#footnote-ref-100)
101. The ACD reconstructs this for the now-deprecated PWMP on the basis of Phillipine forms; I therefore interpret it as a PMP form which survived in both the common ancestor of these languages and PSUM. [↑](#footnote-ref-101)
102. Irregular r>Ø. [↑](#footnote-ref-102)
103. Haloban, Leukon, Simeulue *sira*. [↑](#footnote-ref-103)
104. Nias *ira*, Sigulai *ila.* [↑](#footnote-ref-104)
105. TB, KB *tiktik*. [↑](#footnote-ref-105)
106. Semshift ‘to strike lightly on a hard surface’>’tattoo’; seen also in Bontok. [↑](#footnote-ref-106)
107. TB, DB, SB *adat*. [↑](#footnote-ref-107)
108. Nias *ara*, semshift ‘tradition’>’ancient; a long time ago’. [↑](#footnote-ref-108)
109. Nias *ruru.* If Mentawai does subgroup with PCBI, this could be a shared innovation. [↑](#footnote-ref-109)
110. Connection to PMP *ñikñik* seems likely. [↑](#footnote-ref-110)
111. TB, KB, AB *ŋitŋit*; semshift mosquito>moth. [↑](#footnote-ref-111)
112. Semshift ‘mosquito’>’gnat’. [↑](#footnote-ref-112)
113. TB, AB *tandak*, semshift ‘to stamp feet’>’to kick backwards’. [↑](#footnote-ref-113)
114. Semshift ‘to stamp feet’>’sole of feet’. [↑](#footnote-ref-114)
115. TB *unoŋ*, semshift ‘motionless’>’still; of water’. [↑](#footnote-ref-115)
116. Semshift ‘motionless’>’asleep’. [↑](#footnote-ref-116)