# A preliminary archaeology of tone in Raja Ampat

Laura ARNOLD

The University of Edinburgh

At least three Austronesian languages spoken in the Raja Ampat archipelago have lexical tone: Maʿya, Matbat, and Ambel. The objective of this paper is to examine data from these three languages, in order to determine how tone originated and developed. Using comparative data from monosyllabic cognates, I will show that, in the case of Maʿya and Matbat, tone was inherited from a single common ancestor; but that tone developed separately in Ambel. Possible scenarios for tonogenesis in proto-Maʿya-Matbat and proto-Ambel will then be explored. I will conclude that, in the absence of evidence for spontaneous, independent tonogenesis, the most likely scenario was that proto-Maʿya-Matbat developed tone through contact with a now-extinct tonal Papuan substrate. Proto-Ambel also likely developed tone through contact; however, it is at present unclear whether this contact was also with a Papuan substrate, or with tonal proto-Maʿya-Matbat or one of its descendants.

## 1. Introduction[1]

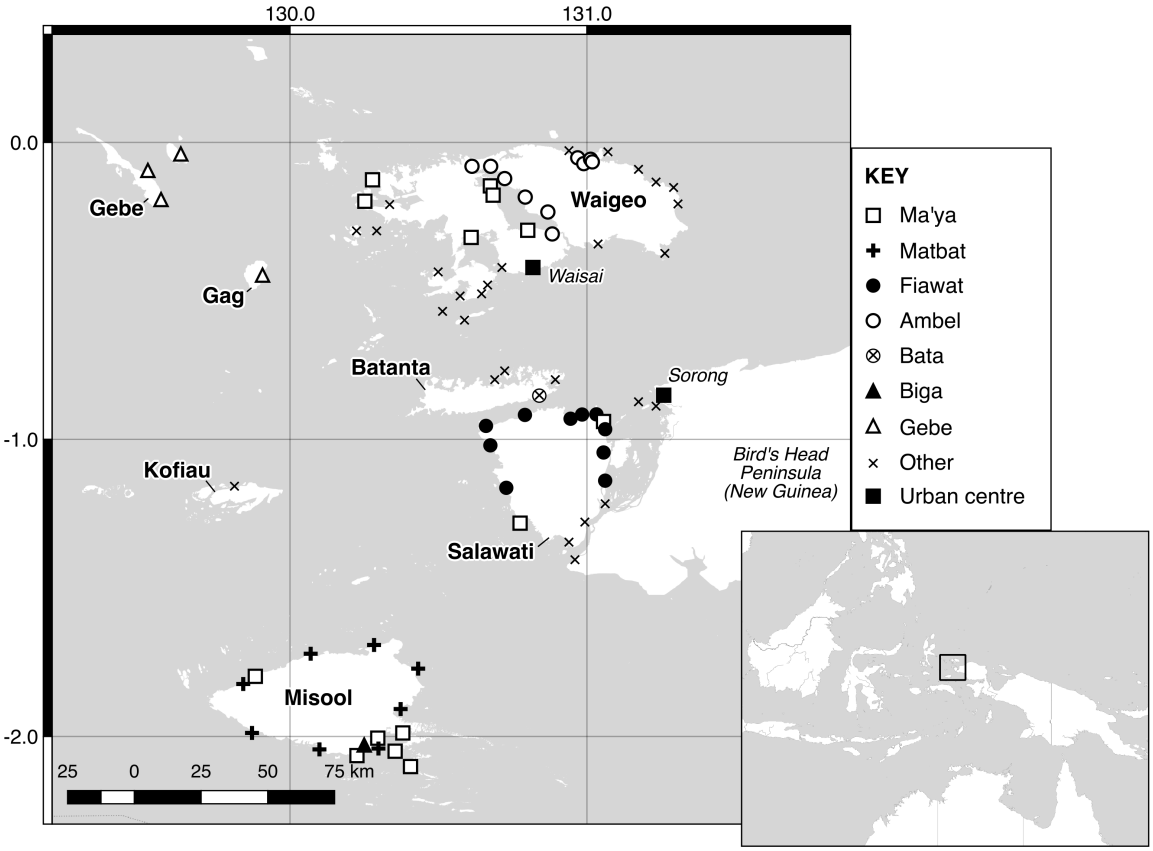
The Raja Ampat archipelago lies off the Bird's Head peninsula, at the western tip of New Guinea. As shown in Map 1, the archipelago consists of four large islands – Waigeo, Batanta, Salawati, and Misool – and hundreds of smaller islands. Administratively, the archipelago is part of West Papua province, in eastern Indonesia.

There are several languages spoken in Raja Ampat. At least seven of these belong to the South Halmahera-West New Guinea (SHWNG) subgroup of Austronesian, the lesser-known sister of Oceanic; within SHWNG, these languages are classified in the Raja Ampat-South Halmahera branch (RASH; Remijsen 2001a:34–7; Kamholz 2014). The RASH languages spoken in and around Raja Ampat are marked in Map 1, and are as follows: Maʿya, spoken in villages throughout the archipelago; Ambel, spoken on Waigeo; Bata, spoken on Batanta; Fiawat, spoken on Salawati; Matbat and Biga, both spoken on Misool; and Gebe, spoken on the small island of Gag and the Gebe islands between Waigeo and Halmahera.[2] Biak, a language belonging to the Cenderawasih Bay branch of SHWNG, is also spoken throughout Raja Ampat; as, increasingly, are dialects of Malay. These non-RASH Austronesian languages are relatively recent incomers to the archipelago (Remijsen 2001a:30–1; Arnold 2018a:17–8); as such, they will not receive further comment in this paper. Remijsen (2001a:30–1) also lists two non-Austronesian, Papuan languages that are spoken to a limited extent in the archipelago, both of which have arrived recently. The first is Duriankari, whose speakers migrated

[1] The Ambel data presented in this paper are the result of five periods of fieldwork between January 2014 and June 2017. I would like to extend my gratitude to the Ambel community, for their hospitality, hard

[2] As, which is also a RASH language, is spoken in three villages on the coast of the Bird's Head. As is not marked in Map 1.

ARNOLD, Laura, 2018. 'A preliminary archaeology of tone in Raja Ampat'. In Antoinette SCHAPPER, ed. *Contact and substrate in the languages of Wallacea* PART 2. NUSA 64: 7–37. [Permanent URL: http://hdl.handle.net/10108/92289]
[doi: 10.5281/zenodo.1450778]




**Map 1. The Raja Ampat archipelago (based on Remijsen 2001a:16 and Arnold 2018a:5)**

from the south coast of the Bird's Head to a single village on south Salawati (Polansky 1957, cited in Remijsen 2001a:20; Voorhoeve 1975); de Vries (1998:644) reports that Duriankari may now be extinct. The second is Moi, speakers of which have migrated from the Bird's Head to east and south Salawati.

The internal classification of the RASH subgroup is still in its early stages: while the RASH languages spoken in the south of Halmahera form a primary branch of RASH (Kamholz 2014), it is unclear whether those spoken in and near Raja Ampat form a separate primary branch. Kamholz (2014), based on phonological and morphological innovations in the languages, concludes that they constitute several primary branches of RASH: Ambel-Biga, Maʿya-Matbat, Fiawat, and As. However, Kamholz (2015), a reconstruction of proto-SHWNG morphology, casts doubt on the validity of the Maʿya-Matbat branch, and Kamholz (2017:10 f.n. 4) has since retracted the Ambel-Biga branch.

To date, the best described of the RASH languages of Raja Ampat are Maʿya (van der Leeden 1993, n.d.; Remijsen 2001a, 2001b, 2002), Matbat (Remijsen 2001a, 2007, 2010), and Ambel (Arnold 2018a, 2018b). One curious feature of these three languages is that they all have systems of lexical tone. Tone is only sporadically attested in Austronesian languages; the question therefore arises as to how tone developed in the Austronesian languages of Raja Ampat. This paper aims to provide a preliminary answer to this question, and in the process contribute to the question of the internal subgrouping of the RASH branch. Only data from these three languages will be compared in the first instance because phonological analyses of the word-prosodic systems of the other RASH languages of Raja Ampat have not yet been carried out.

This paper is structured as follows. Following a brief description of the segmental phonologies and word-prosodic systems of Maʿya, Matbat, and Ambel in section 2, monosyllabic cognates in the three languages will be compared in section 3, in order to determine the extent to which tone has been inherited, and hence the point at which the languages became tonal. I will present evidence to support the hypothesis that tone in Maʿya and Matbat has been inherited from a common ancestor, but that the tone system of Ambel developed independently. In section 4, two possible mechanisms for tonogenesis in proto-Maʿya-Matbat and proto-Ambel will be considered: independent innovation through the spontaneous phonemicisation of previously allophonic pitch differences; and contact-induced tonogenesis. I will show that tone likely developed in the Austronesian languages of Raja Ampat as the result of contact with a now-extinct Papuan substrate. In section 5, I will discuss the timescale and nature of Austronesian-Papuan contact in Raja Ampat. Finally, the conclusions of this paper are summarised in section 6, where I also outline directions for future research.

## 2. Outline of the phonological systems

In this section, the segmental and suprasegmental phonologies of Maʿya, Matbat, and Ambel are outlined. This information will provide the necessary background for the comparison of the prosodic systems in section 3.

### 2.1 Maʿya

Of the three Raja Ampat languages discussed in this paper, the phonology of Maʿya has received the most attention. There are five dialects of Maʿya: Kawe, Wauyai, and Laganyan Maʿya, all spoken on Waigeo; Salawati Maʿya, spoken on Salawati; and Misool Maʿya, spoken on Misool. Maʿya has 14 consonant phonemes (/p b t d k g f s m n l r j w/) and five vowel phonemes (/i e a o u/; van der Leeden 1993).

While the segmental phonology of Maʿya is unremarkable, the prosodic system is unusual, in that the Laganyan, Misool, and Salawati dialects of the language have both lexical tone and lexical stress systems (Remijsen 2001a, 2001b, 2002).[3] The following is a summary of the word-prosodic system of Salawati Maʿya.

Lexical tone is restricted to word-final syllables. There is a three-way contrast between High (/$^{3}$/), Rise (/$^{12}$/), and toneless syllables.[4] Phonetically, High syllables are realised with high pitch in all utterance positions; Rise syllables are realised utterance-finally with rising pitch, and utterance-medially with low pitch; and toneless syllables are realised with falling pitch. Lexical stress is restricted to the penultimate and final syllables of a word; the primary acoustic correlate is length. If a word has penultimate stress, the final syllable cannot bear the Rise toneme. There is a strong tendency for words with penultimate stress to have the same vowels in the penultimate and final syllables (e.g., /ˈtala$^{3}$/ ‘banana’, /ˈlomo$^{3}$s/ ‘blood’). Finally, words with final syllables

[3] The Kawe and Wauyai dialects do not have lexical tone, only lexical stress (Remijsen 2001a:87; 2001b:479). Remijsen (2001a:494–5) concludes that the Kawe and Wauyai dialects also once had tone systems, which they have subsequently lost.

[4] Remijsen (2001a, 2001b, 2002) describes toneless, stressed syllables as bearing a ‘Fall’ toneme. However, as these syllables do not have an underlying specification for tone, in this paper I refer to them as ‘toneless stressed’ syllables. While superscript numerals may be used to transcribe phonetic pitch, in Remijsen (2001a, 2001b, 2002, 2007) and throughout this paper they are representative of phonological tone.

that are both stressed and toneless are realised with an epenthetic final *-o* in sentence-final position.

Remijsen (2001b) describes the variation in the tone systems of the three tonal Maʿya dialects. There is one structural difference between the Misool dialect and the Salawati dialect: the Rise toneme in Salawati Maʿya is cognate with a Low toneme in Misool Maʿya. This Low toneme is realised with low pitch both utterance-medially and utterance-finally. Another difference between the Salawati and Misool dialects is distributional: whereas polysyllabic words with penultimate stress tend to have High tone on the final syllable in Salawati Maʿya, the final syllable bears Low tone in the Misool dialect.

There are two differences between the Laganyan dialect and the Salawati dialect described in Remijsen (2001b). First, there is an utterance-final fall boundary tone in Laganyan Maʿya, which is realised on voiced codas. This boundary tone means that High syllables with a voiced coda are realised with high falling pitch utterance-finally, and that utterance-final Rise syllables with a voiced coda are realised with rise-fall pitch. Second, if a Rise syllable in Laganyan Maʿya is followed by another prosodically marked syllable (i.e., a syllable marked for either stress or tone), the realisation is nearly identical with utterance-medial toneless syllables.

Remijsen concludes that the differences in the Misool and Laganyan dialects are recent developments, and that the tone system of Salawati Maʿya is the most conservative of the three dialects, reflecting most closely the tone system of proto-Maʿya (2001b:492). As there has not to date been any systematic reconstruction of proto-Maʿya, data from the Salawati dialect will be used as representative of Maʿya for the remainder of this paper. Henceforth, unless otherwise noted, I will use the term ‘Maʿya’ to refer to the Salawati Maʿya dialect.

### 2.2 Matbat

There are two dialects of Matbat: Magey Matbat, and Tomolol Matbat (Remijsen 2007:9–10). Only Magey Matbat has so far been described (Remijsen 2007, 2010). Data from Magey Matbat will therefore be used as representative of Matbat in this study.

Magey Matbat has 16 consonant phonemes (/p b t d k g m n ŋ s f h l w j/, plus a shallowly-integrated loan phoneme /r/) and seven vowels (/i e ɛ a ɔ o u/; Remijsen 2010). The syllable structure of Matbat is (C)V(C), and there is a strong preference for monosyllabic words. Matbat does not have lexical stress; it does, however, have an unusually rich tone system. Remijsen (2007) analyses Matbat with six tonemes: Low (/$^{1}$/), High (/$^{3}$/), Extra-high Fall (/$^{41}$/), Low Fall (/$^{21}$/), Low Rise (/$^{12}$/), and Rise-Fall (/$^{121}$/). The domain of tonal specification in Matbat is the syllable; while toneless syllables are permitted, tonal specification is obligatory in content words. If the final syllable of a word is specified with the Low Fall, an epenthetic final *-o* occurs in utterance-final position; as will be returned to in section 3.3 below, this *-o* is related to the epenthetic *-o* found in Maʿya.

### 2.3 Ambel

There are two dialects of Ambel: Metsam Ambel, spoken in two villages on Waigeo; and Metnyo Ambel, spoken in nine further villages on the island. Like Maʿya and Matbat, the segmental phonology of Ambel is simple. There are 14 native consonant

phonemes in Metnyo Ambel (/p b t d k g s h m n l r j w/);[5] Metsam Ambel has /f/ for Metnyo /h/. Both dialects have a simple five-vowel system (/i e a o u/; Arnold 2018a).

Both Metsam and Metnyo Ambel have tone systems. Tone in Metnyo Ambel is binary and privative: High syllables contrast with toneless syllables. The system is culminative, in that there is a maximum of one High syllable per morpheme, and one realisation of High tone per word; but not obligatory, in that toneless words are attested, including toneless monosyllables. Utterance-medially, High syllables are realised with high pitch, and toneless syllables are realised with low pitch. Utterance-finally, there is a HL% boundary tone, the Low component of which is only realised on heavy syllables, i.e., syllables with a vowel plus sonorant coda. This boundary tone leads to utterance-final realisations that are acoustically and perceptually very similar to those described above for the High and Rise syllables in the Laganyan dialect of Maʿya: heavy High syllables are realised with high falling pitch, and heavy toneless syllables are realised with rise-fall pitch. This similarity will be returned to in section 4.2, where it will be used as evidence for contact between speakers of Ambel and speakers of Laganyan Maʿya.

Work on the tone system of Metsam Ambel is still preliminary. However, there are two or possibly three underlying tones in this dialect. Syllables can be sorted into one of three groups, depending on utterance-medial realisation: those realised with high, rising, and low pitch. For the purposes of this paper, I will work from the assumption that high-pitched syllables are underlyingly High (/$^{3}$/), syllables realised with utterance-medial rising pitch underlyingly bear Rise tone (/$^{12}$/), and low-pitched syllables are underlyingly toneless.[6] While Metnyo Ambel does not have lexical stress, at present it is unclear whether Metsam Ambel only has a lexical tone system (as in Metnyo Ambel); or whether it combines lexical stress and lexical tone (as in Maʿya). Like Metnyo Ambel, the Metsam dialect has an utterance-final HL% boundary tone, again with the Low component only realised on heavy syllables. Utterance-finally, heavy High syllables are realised with high falling pitch; and both heavy Rise and toneless syllables are realised with rise-fall pitch.

Ongoing comparative work between the two dialects shows that the tone system of proto-Ambel (pA) was identical with the present-day tone system of Metsam Ambel: pA, like Metsam Ambel, had a three-way contrast between *High, *Rise, and *toneless syllables (Arnold submitted). The utterance-final HL% boundary tone found in both dialects is also reconstructed to proto-Ambel. In this way, the reconstructed pA system is typologically identical with the present-day Maʿya system, which also distinguishes High, Rise, and toneless syllables; in addition, the Laganyan Maʿya dialect has an utterance-final fall boundary tone. This point will be discussed in more detail in section 4.2.

In Arnold (submitted), the development of the pA tone system is described in detail. The proto-forms reconstructed in that study will be used as comparanda for the remainder of this paper. Only monosyllables have so far been reconstructed in pA: this is due to the lack of clarity mentioned above regarding the presence of lexical stress in

[5] The palatal glide /j/ will henceforth be transcribed *y* for all three languages.

[6] In Arnold (2018a, 2018b, submitted), Ambel tone is transcribed with diacritics on the vowel: High tone is transcribed *á*, and Rise tone is transcribed *ă*. However, as tone in Maʿya and Matbat is conventionally transcribed with superscript numerals, to facilitate comparison Ambel tone is also transcribed in this way in this paper.

Metsam Ambel. By restricting the comparanda to monosyllables, which are intrinsically stressed, we avoid erroneously analysed forms entering the data.

### 2.4 Summary

The properties of the word-prosodic systems discussed in this section are given in table 1. Only the tonal dialects of Maʿya are included in this table.

**Table 1. Summary of the word-prosodic systems of Maʿya, Matbat, and Ambel**

| | **Maʿya** | | | **Matbat** | **Ambel** | | |
|---|---|---|---|---|---|---|---|
| | *Salawati* | *Misool* | *Laganyan* | | *Metnyo* | *Metsam* | *proto-Ambel* |
| Lexical stress? | ✓ | ✓ | ✓ | × | × | ? | ? |
| Tonemes | High /$^{3}$/<br>Rise /$^{12}$/ | High /$^{3}$/<br>Low /$^{1}$/ | High /$^{3}$/<br>Rise /$^{12}$/ | High /$^{3}$/<br>Low Rise /$^{12}$/<br>Low /$^{1}$/<br>Low Fall /$^{21}$/<br>Extra-High Fall /$^{41}$/<br>Rise-Fall /$^{121}$/ | High /$^{3}$/ | High /$^{3}$/<br>Rise /$^{12}$/ | *High /$^{3}$/<br>*Rise /$^{12}$/ |
| Epenthetic final /-o/? | ✓ | ✓ | ✓ | ✓ | × | × | × |
| Utterance-final fall boundary tone? | × | × | ✓ | × | ✓ | ✓ | ✓ |

## 3. Tone and inheritance in Raja Ampat

In this section, the hypothesis that the tone systems of Maʿya, Matbat, and Ambel were inherited from a common ancestor will be considered. The aim of this section is primarily to determine at what point in the history of the three languages tone was first innovated; and secondarily to contribute to the subgrouping of the RASH languages. As I will show below, there is evidence that the tone systems of Maʿya and Matbat have been inherited from a common ancestor, to the exclusion of Ambel, which developed tone independently.

In the following sections, monosyllabic cognates will be compared pairwise for the three Raja Ampat languages.[7] The Maʿya and Ambel cognates will be compared and discussed in section 3.1; the Matbat and Ambel cognates in section 3.2; and the Maʿya and Matbat cognates in section 3.3. The full list of cognates for all three languages, alphabetised by meaning, is provided in the Appendix.

[7] The present study is limited to monosyllabic cognates because, as described in section 2.3, only monosyllabic forms have been reconstructed in proto-Ambel.

### 3.1 Maʿya and Ambel

The monosyllabic cognates in Salawati Maʿya and proto-Ambel are given in table 2. 35 monosyllabic cognates were identified in the two languages.[8] To facilitate discussion, the data are divided into cognate sets: group A.1 (Maʿya High :: pA High); group A.2 (Maʿya High :: pA toneless); group A.3 (Maʿya High :: pA Rise); group A.4 (Maʿya Rise :: pA High); group A.5 (Maʿya Rise :: pA toneless); group A.6 (Maʿya toneless :: pA High); and group A.7 (Maʿya toneless :: pA Low).[9]

**Table 2. Monosyllabic cognates in Salawati Maʿya and proto-Ambel**

| | | Salawati Maʿya | Proto-Ambel |
|---|---|---|---|
| *Group A.1: Maʿya High :: proto-Ambel High* | | | |
| 1. | ‘eight’ | ʿwa$^{3}$l | *wa$^{3}$l |
| 2. | ‘mountain’ | ʿye$^{3}$l | *i$^{3}$l |
| 3. | ‘sea turtle’ | ʿfe$^{3}$n | *fi$^{3}$n |
| 4, | ‘three’ | ʿto$^{3}$l | *tu$^{3}$l |
| *Group A.2: Maʿya High :: proto-Ambel toneless* | | | |
| 5. | ‘enter’ | ʿsu$^{3}$n | *sun |
| 6. | ‘fish’ | ʿdo$^{3}$n | *dun |
| 7. | ‘five’ | ʿli$^{3}$m | *lim |
| 8. | ‘good’ | ʿfi$^{3}$ | *fi |
| 9. | ‘kill’ | ʿbu$^{3}$n | *bun |
| 10. | ‘rise, ascend’ | ʿsa$^{3}$ | *sa |
| 11. | ‘two’ | ʿlu$^{3}$ | *lu |
| 12. | ‘village’ | ʿpnu$^{3}$ | *nu ‘house’ |
| 13. | ‘white’ | ʿbu$^{3}$s | *bus |
| 14. | ‘woman’ | ʿpi$^{3}$n | *bin |
| *Group A.3: Maʿya High :: proto-Ambel Rise* | | | |
| 15. | ‘louse’ | ʿu$^{3}$t | *o$^{12}$wt |
| 16. | ‘sago’ | ʿbi$^{3}$ | *bi$^{12}$ |
| *Group A.4: Maʿya Rise :: proto-Ambel High* | | | |
| 17. | ‘die’ | ʿma$^{12}$t | *mna$^{3}$t |
| 18. | ‘four’ | ʿfa$^{12}$t | *fa$^{3}$t |
| 19. | ‘ground, earth’ | ʿba$^{12}$t | *ba$^{3}$t |

[8] Cognacy judgements are taken from Kamholz (2014, n.d.), supplemented by the author.

[9] Although it is transcribed in Remijsen (2001a), Maʿya epenthetic final *-o* is not represented in this paper, as it is not underlyingly present. For some meanings, data are not available for Salawati Maʿya; in these cases, data from Misool Maʿya are given, as the primary difference with regards to monosyllabic words in the two dialects is in the realisation of tone, rather than tonal category (see section 2.1; in Remijsen 2001a, 2001b, the transcription of the Salawati Rise toneme and the Misool cognate Low toneme is identical, i.e., <$^{12}$>).

| | | Salawati Maˈya | Proto-Ambel |
|---|---|---|---|
| 20. | ‘man’ | ˈma$^{12}$n (Misool) | *ma$^{3}$n |
| 21. | ‘mother’ | ˈne$^{12}$n | *ne$^{3}$n |
| 22. | ‘see’ | ˈ-e$^{12}$m | *e$^{3}$m |
| 23. | ‘swim’ | ˈ-a$^{12}$s (Misool) | *la$^{3}$ |
| *Group A.5: Maˈya Rise :: proto-Ambel toneless* | | | |
| 24. | ‘betel leaf’ | ˈnya$^{12}$n | *nyan |
| 25. | ‘canoe’ | ˈwa$^{12}$k | *wan |
| 26. | ‘fire’ | ˈla$^{12}$p | *lap |
| 27. | ‘full’ | ˈfo$^{12}$n | *fon |
| 28. | ‘rice’ | ˈfa$^{12}$s | *fa |
| 29. | ‘sand’ | ˈle$^{12}$n | *layn |
| 30. | ‘snake’ | ˈko$^{12}$k | *kok |
| *Group A.6: Maˈya toneless :: proto-Ambel High* | | | |
| 31. | ‘person’ | ˈmat | *me$^{3}$t |
| 32. | ‘tree, wood’ | ˈai | *a$^{3}$y |
| 33. | ‘walk’ | ˈdak (Misool) | *ta$^{3}$n |
| *Group A.7: Maˈya toneless :: proto-Ambel toneless* | | | |
| 34. | ‘give’ | ˈbe (Misool) | *bi |
| 35. | ‘know’ | -ˈun (Misool) | *un |

The typological similarity between the tone systems of Maˈya and pA was noted above: both systems contrast High, Rise, and toneless syllables. Despite this, table 2 shows that no systematic tonal correspondences can be identified between the two languages. High monosyllables in Maˈya have pA cognates with High (A.1), Rise (A.2), and no tone (A.3); Rise monosyllables in Maˈya have cognates with High (A.4) or no tone (A.5); and toneless monosyllables in Maˈya have cognates with High (A.6) or no tone (A.7). Nor can any segmental features be identified that might provide a conditioning environment. For example, the cognate of Maˈya *ˈma$^{12}$n* ‘man, male’, with Rise tone, is pA *ma$^{3}$n, with High tone; while the cognate of the near-minimal Maˈya *ˈnya$^{12}$n* ‘betel leaf’, with Rise tone, is the toneless pA *nyan. Similarly, the cognate of Maˈya *ˈbi$^{3}$* ‘sago’, with High tone, is pA *bi$^{12}$, with Rise tone; however, the cognate of Maˈya *ˈbu$^{3}$n* ‘kill’, also with High tone, is the toneless pA *bun.

On the basis of the data presented in table 2, we can conclude that the tone systems of Maˈya and Ambel have not been inherited from a common ancestor. The two languages have developed tone independently from one another.

### 3.2 Matbat and Ambel

The monosyllabic cognates between Matbat and proto-Ambel are given in table 3. In total, 35 monosyllabic cognates were identified. As in the previous section, the data are divided into cognate sets: group B.1 (Matbat High :: pA High); group B.2 (Matbat High :: pA toneless); group B.3 (Matbat High :: pA Rise); group B.4 (Matbat Rise :: pA High); group B.5 (Matbat Low :: pA toneless); group B.6 (Matbat Fall :: pA toneless); and group B.7 (Matbat Fall :: pA High).

**Table 3. Monosyllabic cognates in Matbat and proto-Ambel**

| | | Matbat | Proto-Ambel |
|---|---|---|---|
| *Group B.1: Matbat High :: proto-Ambel High* | | | |
| 1. | 'eight' | -wa$^{3}$l | *wa$^{3}$l |
| 2. | 'four' | fa$^{3}$t | *fa$^{3}$t |
| 3. | 'ground, earth' | ba$^{3}$t | *ba$^{3}$t |
| 4. | 'mother' | ne$^{3}$n | *ne$^{3}$n |
| 5. | 'mountain' | he$^{3}$l | *i$^{3}$l |
| 6. | 'person' | ma$^{3}$t | *me$^{3}$t |
| 7. | 'sea turtle' | fe$^{3}$n | *fi$^{3}$n |
| 8. | 'see' | -ɛ$^{3}$ŋ | *e$^{3}$m |
| 9. | 'swim' | la$^{3}$s | *la$^{3}$ |
| 10. | 'three' | to$^{3}$l | *tu$^{3}$l |
| 11. | 'tree, wood' | ha$^{3}$y | *a$^{3}$y |
| *Group B.2: Matbat High :: proto-Ambel toneless* | | | |
| 12. | 'canoe' | wa$^{3}$ŋ | *wan |
| 13. | 'enter' | hu$^{3}$ŋ | *sun |
| 14. | 'fire' | ya$^{3}$p | *lap |
| 15. | 'five' | li$^{3}$m | *lim |
| 16. | 'full' | fo$^{3}$n | *fon |
| 17. | 'good' | fi$^{3}$ | *fi |
| 18. | 'kill' | bu$^{3}$n | *bun |
| 19. | 'rice' | fa$^{3}$s | *fa |
| 20. | 'rise, ascend' | ha$^{3}$ | *sa |
| 21. | 'sand' | ye$^{3}$n | *layn |
| 22. | 'snake' | ko$^{3}$k | *kok |
| 23. | 'two' | lu$^{3}$ | *lu |
| 24. | 'village' | nu$^{3}$ | *nu 'house' |
| 25. | 'white' | bu$^{3}$ | *bus |
| 26. | 'woman' | (wa$^{1}$t)bi$^{3}$n 'kind of mangrove'[10] | *bin |
| *Group B.3: Matbat High :: proto-Ambel Rise* | | | |
| 27. | 'louse' | wu$^{3}$t | *o$^{12}$wt |
| *Group B.4: Matbat Rise :: proto-Ambel High* | | | |
| 28. | 'die' | ma$^{12}$t | *mna$^{3}$t |
| 29. | 'green/blue' | bla$^{12}$w | *bya$^{3}$w |

[10] Compare Ambel *pray bin* 'kind of mangrove', literally 'female mangrove'.

| | | Matbat | Proto-Ambel |
|---|---|---|---|
| *Group B.5: Matbat Low :: Ambel toneless* | | | |
| 30. | ‘betel leaf’ | na$^{1}$n | *nyan |
| 31. | ‘needle’ | la$^{1}$m | *yam |
| 32. | ‘night’ | ka$^{1}$m | *gam |
| *Group B.6: Matbat Fall :: Ambel toneless* | | | |
| 33. | ‘give’ | be$^{21}$ | *bi |
| 34. | ‘know’ | -u$^{21}$n | *un |
| *Group B.7: Matbat Fall :: Ambel High* | | | |
| 35. | ‘man, male’ | (wa$^{3}$y)ma$^{21}$n | *ma$^{3}$n |

Similar to the comparison of the Maʿya and Ambel data in the previous section, no tonal correspondences between Matbat and Ambel can be identified. Cognates of Matbat High monosyllables are High (B.1), toneless (B.2), or Rise (B.3) in pA; cognates of Matbat Fall monosyllables are toneless (B.6) and High (B.7) in pA. Similarly, there are no obvious segmental environments that might condition tonal splits or mergers in one or both of the languages. For example, the cognate of Matbat *wa$^{3}$l* ‘eight’, with High tone, is pA *wa$^{3}$l, with High tone, whereas the cognate of the near-minimal Matbat *wa$^{3}$ŋ* ‘canoe’, also with High tone, is the toneless pA *wan. Similarly, the cognate of Matbat *ne$^{3}$n* ‘mother’, with High tone, is pA *ne$^{3}$n, with High tone, while the cognate of Matbat *ye$^{3}$n* ‘sand’ is the toneless pA *layn.

Based on these data, we can again conclude that the tone systems of Ambel and Matbat have not descended from a common ancestor, but have developed independently.

### 3.3 Matbat and Maʿya

The monosyllabic cognates in Matbat and Salawati Maʿya are given in table 4. 41 monosyllabic cognates were identified. As above, the data are divided into groups: group C.1 (Matbat Extra-High :: Maʿya Rise); group C.2 (Matbat High :: Maʿya High); group C.3 (Matbat High :: Maʿya Rise); group C.4 (Matbat High :: Maʿya toneless); group C.5 (Matbat Rise :: Maʿya Rise); group C.6 (Matbat Low :: Maʿya Rise); group C.7 (Matbat Low Fall :: Maʿya toneless); and group C.8 (Matbat Low Fall :: Maʿya Rise).

**Table 4. Monosyllabic cognates in Matbat and Salawati Maʿya**

| | | Matbat | Salawati Maʿya |
|---|---|---|---|
| *Group C.1: Matbat Extra-High :: Maʿya Rise* | | | |
| 1. | ‘hear’ | no$^{41}$ŋ | ʿdo$^{12}$n |
| *Group C.2: Matbat High :: Maʿya High* | | | |
| 2. | ‘breast’ | su$^{3}$ | ʿsu$^{3}$s |
| 3. | ‘come’ | bo$^{3}$t | ʿbo$^{3}$t |
| 4. | ‘eight’ | -wa$^{3}$l | ʿwa$^{3}$l |
| 5. | ‘enter’ | hu$^{3}$ŋ | ʿsu$^{3}$n |

| | | Matbat | Salawati Maˈya |
|---|---|---|---|
| 6. | ‘five’ | li$^{3}$m | ˈli$^{3}$m |
| 7. | ‘good’ | fi$^{3}$ | ˈfi$^{3}$ |
| 8. | ‘kill’ | bu$^{3}$n | ˈbu$^{3}$n |
| 9. | ‘louse’ | wu$^{3}$t | ˈu$^{3}$t |
| 10. | ‘mountain’ | he$^{3}$l | ˈye$^{3}$l |
| 11. | ‘rise, ascend’ | ha$^{3}$ | ˈsa$^{3}$ |
| 12. | ‘sea turtle’ | fe$^{3}$n | ˈfe$^{3}$n |
| 13. | ‘seawards’ | lo$^{3}$w | ˈlo$^{3}$l |
| 14. | ‘three’ | to$^{3}$l | ˈto$^{3}$l |
| 15. | ‘two’ | lu$^{3}$ | ˈlu$^{3}$ |
| 16. | ‘village’ | nu$^{3}$ | ˈpnu$^{3}$ |
| 17. | ‘white’ | bu$^{3}$ | ˈbu$^{3}$s |
| 18. | ‘woman’ | (wa$^{1}$t)bi$^{3}$n ‘kind of mangrove’ | ˈpi$^{3}$n |
| *Group C.3: Matbat High :: Maˈya Rise* | | | |
| 19. | ‘canoe’ | wa$^{3}$ŋ | ˈwa$^{12}$k |
| 20. | ‘fire’ | ya$^{3}$p | ˈla$^{12}$p |
| 21. | ‘four’ | fa$^{3}$t | ˈfa$^{12}$t |
| 22. | ‘full’ | fo$^{3}$n | ˈfo$^{12}$n |
| 23. | ‘ground, earth’ | ba$^{3}$t | ˈba$^{12}$t |
| 24. | ‘mother’ | ne$^{3}$n | ˈne$^{12}$n |
| 25. | ‘rice’ | fa$^{3}$s | ˈfa$^{12}$s |
| 26. | ‘sand’ | ye$^{3}$n | ˈle$^{12}$n |
| 27. | ‘see’ | -ɛ$^{3}$ŋ | ˈ-e$^{12}$m |
| 28. | ‘snake’ | ko$^{3}$k | ˈko$^{12}$k |
| 29. | ‘swim’ | la$^{3}$s | ˈ-a$^{12}$s (Misool) |
| 30. | ‘under’ | (pa)pa$^{3}$p | ˈpa$^{12}$p |
| *Group C.4: Matbat High :: Maˈya toneless* | | | |
| 31. | ‘person’ | ma$^{3}$t | ˈmat |
| *Group C.5: Matbat Rise :: Maˈya Rise* | | | |
| 32. | ‘die’ | ma$^{12}$t | ˈma$^{12}$t |
| 33. | ‘much’ | to$^{12}$ | ˈmo$^{12}$t |
| *Group C.6: Matbat Low :: Maˈya Rise* | | | |
| 34. | ‘shoot’ | -a$^{1}$n | ˈfa$^{12}$n |
| 35. | ‘betel leaf’ | na$^{1}$n | ˈnya$^{12}$n |
| *Group C.7: Matbat Low Fall :: Maˈya toneless* | | | |
| 36. | ‘eat (tr.)’ | -a$^{21}$ | ˈ-a |
| 37. | ‘give’ | be$^{21}$ | ˈbe (Misool) |

| | | Matbat | Salawati Maˈya |
|---|---|---|---|
| 38. | ‘know’ | -u$^{21}$n | ˈ-un (Misool) |
| 39. | ‘mouth’ | ga$^{21}$l | ˈgal |
| *Group C.8: Matbat Low Fall :: Maˈya Rise* | | | |
| 40. | ‘egg’ | to$^{21}$l | ˈto$^{12}$l |
| 41. | ‘man’ | (wa$^{3}$y)ma$^{21}$n | ˈma$^{12}$n (Misool) |

At first glance, it does not appear that the Matbat and Maˈya data correspond in any meaningful way. However, some patterns can be identified. These patterns provide evidence that the tone systems of the two languages have been inherited from a common source.

Let us begin with the cognates in groups C.2, C.3, and C.4. In these groups, we see that High monosyllables in Matbat are cognate with High (C.2), Rise (C.3), or toneless (C.4) monosyllables in Maˈya (e.g., Matbat *wu$^{3}$t* ‘louse’ :: Maˈya *ˈu$^{3}$t*; Matbat *ne$^{3}$n* ‘mother’ :: Maˈya *ˈne$^{12}$n*; Matbat *ma$^{3}$t* ‘person’ :: Maˈya *ˈmat*). Leaving group C.4 to one side for the moment, the correlation between Matbat High and Maˈya High and Rise is striking. If we assume inheritance as the explanation for the tone of the extant forms, two hypotheses arise: there was a split in Maˈya (i.e., proto-Maˈya-Matbat *High > Maˈya High, Rise), or there was a merger in Matbat (i.e., pMM *High, *Rise > Matbat High). From the available data, we find evidence that the monosyllables in groups C.2 and C.3 originally bore *High tone, and that there has been a split in Maˈya, conditioned by vowel height.

The evidence for this split is as follows. First, in group C.3 (Matbat High :: Maˈya Rise), none of the 12 cognates have a close vowel nucleus /i/ or /u/ in either of the languages. In other words, where Matbat High corresponds to Maˈya Rise, all of the cognates contain the vowel /e/, /a/, or /o/ (e.g., Matbat *ne$^{3}$n* ‘mother’ :: Maˈya *ˈne$^{12}$n*; Matbat *wa$^{3}$ŋ* ‘canoe’ :: Maˈya *ˈwa$^{12}$k*; Matbat *ko$^{3}$k* ‘snake’ :: Maˈya *ˈko$^{12}$k*). Second, a small majority of the monosyllables in group C.2 contain the close vowels /i/ or /u/ in both languages (e.g., Matbat *li$^{3}$m* ‘five’ :: Maˈya *ˈli$^{3}$m*; Matbat *lu$^{3}$* ‘two’ :: Maˈya *ˈlu$^{3}$*; 10/17 cognates, or 59%). This correlation between vowel height and tone in Maˈya suggests that, for the monosyllables in groups C.2 and C.3, we can reconstruct *High tone to pMM. This *High tone split in Maˈya: syllables with close vowel nuclei *i or *u remained High, while those with open vowel nuclei *e, *a, or *o developed Rise.

However, seven of the 17 cognates in group C.2 contain the open vowels /e/, /a/, or /o/, and thus constitute exceptions to the stated conditions. But if we look at data from other RASH languages, including other dialects of Maˈya, we find evidence to support reconstruction of at least three of these cognates with a close vowel *i or *u further up the family tree. These three cognates, with the cognate forms in the other RASH languages for which data are available, are given in table 5. Cognates with a close vowel nucleus are highlighted in bold.

**Table 5. Other RASH cognates for group C.2 (cognates with a close vowel nucleus highlighted in bold)**

| Matbat | Salawati Maˈya | Other RASH cognates |
|---|---|---|

| | | | |
|---|---|---|---|
| 'come' | bo$^{3}$t | ˈbo$^{3}$t | Biga *bot*, Fiawat ***but***, Kawe Maˈya ***but***, Laganyan Maˈya ***but***, Wauyai Maˈya ***but*** |
| 'mountain' | he$^{3}$l | ˈye$^{3}$l | Proto-Ambel ****i$^{3}$l***, Biga *yel*, Kawe Maˈya ***yil***, Laganyan ***yil***, Wauyai Maˈya ***yil*** |
| 'sea turtle' | fe$^{3}$n | ˈfe$^{3}$n | Proto-Ambel ****fi$^{3}$n***, Biga ***fin***, Buli *fen*, Fiawat ***fin***, Gane *fen*, Kawe Maˈya ***fin***, Laganyan Maˈya ***fin***, Sawai *fɛn*, Taba *hen* |

Based on the data in table 5, it is possible that the ancestor vowel for these three cognates sets can also be reconstructed as a close vowel, i.e., *i in the case of the 'mountain' and 'sea turtle' sets, and *u in the case of the 'come' set, and that these vowels have subsequently lowered in Matbat and Salawati Maˈya.[11] A systematic reconstruction of proto-RASH phonology is required to confirm this hypothesis.

However, even if the three cognates in table 5 can be reconstructed with *i or *u, several unexplained reflexes remain: the remaining four cognates in group C.2 (4. 'eight', 11. 'rise, ascend', 13. 'seawards', and 14. 'three'), all of which would be predicted by the conditions to bear Rise tone in Maˈya; and the one cognate in group C.4 (31. 'person') which would also be predicted to bear Rise tone. Despite these exceptions, a conservative count of the reflexes shows that the stated conditions account for 22 of the 30 cognates in groups C.2, C.3, and C.4 (73.3%); a Fisher's exact test shows that p=0.001, so this distribution is highly unlikely to be due to chance. If further investigations confirm that the cognates in table 5 can be reconstructed with close vowels in proto-Maˈya-Matbat, this number rises to 25 of the 30 cognates (83.3%; p=0.00004).[12]

[11] As pointed out by an anonymous reviewer, the words for 'sea turtle' are reflexes of proto-Malayo-Polynesian (PMP) *peñu, with a non-high vowel. The cognate forms in the South Halmahera (SH) branch of RASH (i.e., Buli, Gane, Sawai, and Taba) all have a non-high vowel *e* or *ɛ*, suggesting a proto-RASH reconstruction with *e; elsewhere in SHWNG, PMP *e is also retained (e.g., Waropen *eni*). It is therefore possible that the Matbat and Salawati Maˈya vowels are retentions of the PMP vowel, and that the high vowels in the other non-SH RASH languages are innovations. While the subgrouping of the non-SH RASH langauges is at present unclear, the minimum possible number of times that a high vowel was innovated in this scenario is two: once in a putative ancestor to Ambel, Biga, and Fiawat; and once in an ancestor of Kawe and Laganyan Maˈya. If the languages can be subgrouped in this way, then the hypothesis that Matbat and Salawati Maˈya retain proto-RASH *e is more likely than the hypothesis that they have both innovated *e*: the former requires fewer innovations (two) than the latter (three – one innovation of *fen > *fin in a common ancestor to the non-SH RASH languages, and an independent innovation of *fin > *fe$^{3}$n* in both Matbat and Salawati Maˈya). However, if further research shows that Ambel, Biga, and Fiawat do not share a common ancestor to the exclusion of Maˈya and Matbat, more innovations are required, making the retention hypothesis as likely as, or less likely than, the *fin > *fe$^{3}$n* in Matbat and Salawati Maˈya hypothesis.

[12] Some comments can be made about the exceptions found in group C.2. As more data come in from other RASH languages, 13. 'seawards' may also turn out to be reconstructable with a close vowel – at present, the only other known RASH cognates of Matbat *lo$^{3}$w* 'seawards' and Maˈya *ˈlo$^{3}$l* are Buli *lau* 'seaside' and Metnyo Ambel *lu$^{3}$l* 'seawards direction', both of which contain a close vowel. For the remaining three, however, there is no evidence that a close vowel can be reconstructed: the Matbat and Maˈya forms for 4. 'eight', for example, are descended from proto-Austronesian (PAN) *walu 'eight (of non-humans)', and SHWNG cognates include proto-Ambel *wa$^{3}$l, Biak *wār*, Sawai *pɛ-wal*, and Taba *-wal*; the forms for 11. 'rise, ascend' descend from PAN *sakat 'rise, climb up', and only one other SHWNG cognate has been identified, proto-Ambel *sa; and the forms for 14. 'three' descend from proto-

Tone splits conditioned by vowel height are very rare cross-linguistically (Kingston 2011). However, several cases of tonal developments conditioned by vowel height have now been described in Eastern Malayo-Polynesian languages (i.e., SHWNG and Oceanic; see Blust 2005, 2017 for a discussion of unusually-conditioned phonological developments unexpectedly clustering within genetic groupings elsewhere in Austronesian). Within Raja Ampat, Arnold (submitted) describes how vowel height conditioned a tone split in the Metnyo dialect of Ambel: proto-Ambel toneless monosyllables split in Metnyo, with those with an open vowel nucleus developing High tone, and those with a close vowel remaining toneless. Kamholz (2014:106–114) describes how word-final *a* triggered a tone shift from *High > Low on either the second mora of the penult or on word-final syllables in Yerisiam, a SHWNG language spoken near Cenderawasih Bay, on the other side of the Bird's Head from Raja Ampat. Finally, in Cèmuhî, a tonal Oceanic language spoken in New Caledonia, the reflex of the proto-sequences *aqa, *ao, and *oa, all with low vowels, is Low-toned *à* (Rivierre 2001).

For the remaining cognate sets in table 4, the data are too few to draw any firm conclusions. However, two preliminary hypotheses are suggested. First, note that all of the Matbat monosyllables bearing Extra-High, Rise, and Low in the data (i.e., those in groups C.1, C.5, and C.6) correspond to Rise monosyllables in Maʿya: for example, Matbat *no*$^{41}$*ŋ* 'hear' :: Maʿya *ˈdo*$^{12}$*n*; Matbat *ma*$^{12}$*t* 'die' :: Maʿya *ma*$^{12}$*t*; and Matbat *na*$^{1}$*n* 'betel leaf' :: Maya *ˈnya*$^{12}$*n*. There is only a total of five cognates in these three groups – surely too few for us to conclude anything. However, that these patterns are exceptionless in the available data suggests the preliminary hypothesis that they are descended from a single source: possibly from something akin to the Maʿya Rise, which then split in Matbat; or possibly from something more like the three separate tones in Matbat, which merged in Maʿya. Confirmation of this hypothesis awaits further data and analysis.

The second preliminary hypothesis concerns four of the six Matbat Low Fall monosyllables in the data (those in group C.7), which correspond to toneless monosyllables in Maʿya: for example, Matbat *ga*$^{21}$*l* 'mouth' :: Maʿya *ˈgal*. Phonetically, stressed toneless syllables in Salawati Maʿya are realised with falling pitch (Remijsen 2001a, 2001b). While again the data are too few for us to be confident, this correlation suggests that the monosyllables in C.7 are reflexes of forms which, in an ancestral language, were realised with falling pitch (either because they bore Fall tone, or for some phonetic or other prosodic reason). However, this pattern is not exceptionless: two Matbat Low Fall monosyllables (those in group C.8) are cognate with Rise monosyllables in Maʿya. The tone of these cognates is unexplained, and requires further investigation.[13]

---

Central-Eastern Malayo-Polynesian *təlu 'three', with many SHWNG cognates including Ansus *toru*, Biga *tol*, Dusner *tori*, Kowiai *tor*, Sawai *pɛ-tel*, Taba *-tol*, and Wandamen *toru*, all of which have open vowel nuclei (although note proto-Ambel *tu$^{3}$l 'three', with a close vowel). As suggested by an anonymous reviewer, the High tone of 4. 'eight' and 14. 'three' in Maʿya may be due to contamination: in the case of *ˈto*$^{3}$*l* 'three', the preceding numeral *ˈlu*$^{3}$ 'two' has High tone (although the following numeral, *ˈfa*$^{12}$*t* 'four', bears Rise); and in the case of *ˈwa*$^{3}$*l* 'eight', both the preceding and following numerals are High (*ˈfi*$^{3}$*t* 'seven', *ˈsi*$^{3}$ 'nine').

[13] It is noteworthy that the Matbat cognate of Maʿya *ˈma*$^{12}$*n* 'man' is only preserved in the compound *wa*$^{3}$*yma*$^{21}$*n*; it is possible that derived forms undergo some kind of prosodic alternation in Matbat.

The data presented above support the hypothesis that the tone systems of Maʿya and Matbat have descended from a common ancestor, called here proto-Maʿya-Matbat (pMM).[14] There is weak support for the hypotheses that the tonal correlations between Matbat Extra-High, Rise, and Low monosyllables and Maʿya Rise monosyllables (i.e., the data in groups C.1, C.5, and C.6), and between Matbat Low Fall and Maʿya toneless monosyllables (i.e., group C.7), are due to descent from a common source. There is stronger evidence to suggest that the Matbat High monosyllables and Maʿya High and Rise monosyllables in groups C.2 and C.3 have descended from monosyllables which bore pMM *High, and that this *High split in Maʿya, with pMM *High monosyllables with close vowels *i or *u developing High tone (group C.2), and pMM *High monosyllables with open vowels *e, *a, or *o tending to develop Rise tone (group C.3). While there are some unexplained developments, in total these conditions account for the tonal specification of 31 of the 41 cognates (75.61%) in table 4, rising to 34/41 (82.93%) of the cognates if we are able to reconstruct the cognates in table 5 with close vowels.

Further support for the hypothesis that the prosodic systems of Maʿya and Matbat descend from a common ancestor comes from the presence of epenthetic final *-o* in the two languages, described in sections 2.1 and 2.2. This segment occurs in similar contexts in the two languages: sentence-finally on words with final toneless, stressed syllables (realised phonetically with falling pitch) in Maʿya; and utterance-finally on words with final Low Fall in Matbat. Remijsen (2007:25) suggests that the two languages may share this feature due to contact. Kamholz (2014:137), however, considers this epenthetic *-o* to have been inherited from a common ancestor, identifying it as one of the two innovations that define a Maʿya-Matbat subgroup within RASH. Taken with the tonal correspondences discussed above, this latter hypothesis seems to be more likely: both the tone systems, and epenthetic *-o*, were inherited from pMM.

## 4. Mechanisms for tonogenesis

Now that we have established that tone was inherited from a common ancestor into Maʿya and Matbat, and that it was innovated separately in Ambel, mechanisms for tonogenesis in pMM and pA can be explored. Independent, spontaneous innovation without external influence will be discussed in section 4.1, and contact-induced change will be addressed in section 4.2.

### 4.1 Independent innovation

One possible source of a tone system is the through the spontaneous phonemicisation of an earlier phonetic pitch difference conditioned by segmental features; for example, through the transfer of laryngeal features of an onset voicing contrast to the following vowel (Hombert, Ohala & Ewan 1979). Kamholz (2014:96–101) explores independent innovation as a possible mechanism for tonogenesis in Maʿya and Matbat.[15] As Ambel

[14] Further investigations may show that other less well-described Raja Ampat languages can also be classified with Maʿya and Matbat, in which case this name should be reconsidered.

[15] Based on the conclusion in the previous section, that that Maʿya and Matbat likely inherited their tone systems from a common ancestor, ideally we would want to examine reconstructed data from proto-Maʿya-Matbat. Unfortunately, no such reconstructions are yet available. Nonetheless, if Maʿya and Matbat have descended from a common ancestor, as hypothesised here, we should expect to see segmental predictors in the daughter languages.

has only recently been discovered to be tonal (Arnold 2018a, 2018b), no previous studies have looked at whether tone patterns with segmental features in ancestral forms. Therefore, following a summary of the findings presented in Kamholz (2014:96–101), I address the question of whether there is any evidence for independent innovation in proto-Ambel.

Kamholz (2014:96–101) considers words with Austronesian etymologies in Maʿya and Matbat. He finds 53 words with Austronesian etymologies for Maʿya, and 33 for Matbat. He then attempts to determine whether there are any systematic patterns between the segmental form of the reconstructed Austronesian form, and the tonal specifications on the Maʿya and Matbat reflexes.

On the basis of the Maʿya data, Kamholz concludes that "there are no very convincing segmental predictors of tone" (2014:97). However, he notes two correlations between segment and tone: (1) syllable onset *q correlates with Rise in four out of five examples in the 53 reflexes (e.g., PMP *qasu 'smoke' > *ʾla$^{12}$s*; exception PMP *t<in>aqi 'intestines' > *ʾna(o)* 'belly'); and (2) word-final *R correlates with Rise in three examples with no exceptions (e.g., PMP *qitəluR 'egg' > *ʾto$^{3}$l*; this correlation was also noted in Remijsen 2001a:120). Kamholz notes that there are no obvious phonetic motivations for the Maʿya predictors; I would also add that the number of reflexes is too low to draw any firm conclusions regarding whether tone developed from the onset *q or word-final *R.

In Matbat, only one potential predictor of tone is found: onset *p correlates with High tone in five examples in the 33 reflexes, with no exceptions (e.g., PMP *pəñu 'sea turtle' > *fe$^{3}$n*). However, Kamholz notes that, while the voiceless stop *p predictor is a well-known source for the development of high tone (Hombert, Ohala & Ewan 1979), there are again too few examples of this pattern to be confident that this was the origin of High tone in Matbat. In addition, even if the segment *p were the origin of High tone in Matbat, the origin of the five tones other than High from segmental predictors is unclear.

For these reasons, it does not appear likely that pMM independently innovated tone. However, a reconstruction of pMM is required before we reject this hypothesis completely.

Turning now to the Ambel data, proto-Ambel words with Austronesian etymologies were identified, using the data in Kamholz (n.d.), supplemented by additional data from Arnold (2018a). 22 pA words with Austronesian etymologies were identified; these are presented in table 6.[16]

**Table 6. Proto-Ambel words with Austronesian etymologies**

| | Austronesian source | | Proto-Ambel |
|---|---|---|---|
| *Sources of proto-Ambel High tone* | | | |
| 1. | PCEMP | *lakaw 'to be in motion; go, walk' | *do$^{3}$k 'come' |
| 2. | PCEMP | *matay 'die' | *ma$^{3}$t |

[16] The Austronesian etymologies in Kamholz (n.d.) are taken from Blust & Trussel (2010–), supplemented by Blust (1978, 1993, 1999). Once again, as only monosyllables have been reconstructed to proto-Ambel, this discussion is limited to monosyllabic words with Austronesian etymologies. The following abbreviations are used in this table: 'PAN' = Proto-Austronesian, 'PMP' = Proto-Malayo-Polynesian, 'PCEMP' = Proto-Central-Eastern Malayo-Polynesian, 'PEMP' = Proto-Eastern Malayo-Polynesian.

| | Austronesian source | | Proto-Ambel |
|---|---|---|---|
| 3. | PEMP | *pat ‘four’ | *fa$^{3}$t |
| 4. | PMP | *peñu ‘the green turtle, *Chelonia mydas*’ | *fi$^{3}$n ‘sea turtle’ |
| 5. | PCEMP | *ma-Ruqanay ‘male, man’ | *ma$^{3}$n |
| 6. | PCEMP | *kayu ‘wood’ | *a$^{3}$y ‘tree, wood’ |
| *Sources of proto-Ambel Rise tone* | | | |
| 7. | PCEMP | *kutu ‘head louse’ | *u$^{12}$t ‘louse’ |
| 8. | PMP | *Rambia ‘sago palm’ | *be$^{12}$y ‘sago palm, uncooked sago’ |
| *Sources of proto-Ambel toneless syllables* | | | |
| 9. | PEMP | *api ‘fire’ | *lap |
| 10. | PCEMP | *(ba-)b<in>ay ‘woman’ | *bin ‘female, woman’ |
| 11. | PMP | *banua ‘inhabited land, territory supporting the life of a community’ | *nu ‘(village >) house’ |
| 12. | PEMP | *boRe ‘give’ | *bi |
| 13. | PCEMP | *bunuq ‘kill’ | *bun ‘kill, hit’ |
| 14. | PCEMP | *dua ‘two’ | *lu |
| 15. | PCEMP | *lima ‘five’ | *lim |
| 16. | PCEMP | *ma-pia ‘good’ | *fi |
| 17. | PMP | *ma-penuq ‘full’ | *fon |
| 18. | PMP | *pajey ‘rice plant’ | *fa ‘rice’ |
| 19. | PAN | *sakat ‘rise, climb up’ | *sa ‘rise, ascend’ |
| 20. | PCEMP | *waiR ‘fresh water’ | *we ‘water’ |
| 21. | PEMP | *waŋka ‘canoe’ | *wan |

As with the Maʻya and Matbat data, there are no obvious segmental predictors of tone in pA. The best candidate is word-initial *b, which correlates with tonelessness in four examples, with no exceptions (e.g., PEMP *boRe ‘give’ > pA *bi; PCEMP *bunuq ‘kill’ > pA *bun ‘kill, hit’). There is a phonetic motivation for this correlation: onset voiced stops tend to lead to the development of low tone (Hombert, Ohala & Ewan 1979).[17] However, as above, there are too few data for this correlation to be significant. In addition, the origin of pA *Rise is unexplained.

Based on the observations in this section, there is no strong evidence to conclude that tone developed independently in either pMM or pA. It therefore seems likely that contact played at least some role in the development of tone in Raja Ampat. We now turn to a discussion of this hypothesis.

[17] Recall from section 2.3 that utterance-medial toneless syllables are realised with low pitch in both Metsam and Metnyo Ambel.

### 4.2 Contact-induced innovation

Contact is often identified as the trigger for tonogenesis and tonal development (e.g., Matisoff 1973). It was shown in the previous section that independent tonogenesis looks unlikely in either pMM or pA, leaving contact as the most likely explanation for tonogenesis in both proto-languages. There are three contact scenarios that could have given rise to the situation seen today:

> **Hypothesis 1:** Proto-Maʿya-Matbat innovated tone through contact with a tonal Papuan language, and proto-Ambel subsequently innovated tone through contact with pMM or one of its descendants;
>
> **Hypothesis 2:** Proto-Ambel innovated tone through contact with a tonal Papuan language, and pMM subsequently innovated tone through contact with pA or one of its descendants;
>
> **Hypothesis 3:** Both pMM and pA separately innovated tone through contact with a tonal Papuan language or languages.

In this section, I will argue that, while there are no Papuan languages spoken in Raja Ampat today, pMM developed tone through contact with a now-extinct Papuan substrate. I will also show that there is not enough evidence to identify the source of the pA tone system. I will, however, show that there is evidence for convergence of the pA system and the tone system of the Laganyan dialect of Maʿya, post-tonogenesis.

In order to present the evidence in favour of this analysis, an overview of the ways in which different contact scenarios may induce tonogenesis in a language is first required. Ratliff (2002) is a catalogue of the attested outcomes from various contact situations in East and Southeast Asia between languages with and without tone (Donor atonal, Borrower tonal; Donor and Borrower both tonal; Donor tonal, Borrower atonal; Donor and Borrower both atonal). While not a model of language contact *per se*, this paper provides a useful starting point for speculating about the kind of contact that led to tonogenesis in pMM and pA. As tone cannot be reconstructed higher than either pA or pMM, it is likely that the RASH languages were atonal when they arrived in Raja Ampat. As we are interested in the result of contact with a tonal language, only the scenario in which the Donor is tonal and the Borrower is atonal is relevant here.

In situations where the Donor is tonal and the Borrower is atonal, Ratliff describes four different outcomes, depending on the intensity of contact. First, if contact is minimal, an atonal Borrower will borrow words from a tonal Donor without tone (e.g., Chinese loans in English). If contact is more intense, individual words may be borrowed with their tones intact, as in the borrowing of Northern Thai numerals along with their tones into the previously atonal Mon-Khmer language Mal (Filbeck 1972). Mal also displays the third strategy for integrating tonal loans into a formerly atonal language: the assignation of a special 'loan tone' to loanwords. In Mal, this loan tone has a rising pitch profile, which does not correspond with the pitch contours of the Thai words (e.g., Thai *yâak*, Mal *ɲăak* 'difficult'; Thai *khèɛk*, Mal *khĕɛk* 'guest'), and which is also used to mark borrowings from other, atonal Mon-Khmer languages. Finally, if contact is very intense, two systems, ultimately of independent origin, may come to resemble each other closely. This final scenario occurred in the development of tone in Tsat, a Chamic language spoken on Hainan island. Thurgood (1999) reconstructs proto-Cham as atonal. On arrival on Hainan, speakers of Tsat came into contact with dialects of Hainan Min, and two dialects of the Tai-Kadai language Li, all of which were tonal. Contact with these languages stimulated first the development of a register system in Tsat, in which

monosyllables with voiced obstruent onsets developed breathy voice and low pitch, while those with other onsets were realised with modal voice and high pitch. Subsequently, a "full-blown" tone system with five tones developed in Tsat, through the transfer of pitch contours conditioned by final consonants to the vowel nucleus. The resulting system is typologically near-identical with the Tan-chou dialect of Hainan Min, and the two dialects of Li. Importantly, while contact induced the transfer of features of both proto-Cham initials and finals to the vowel in Tsat, tone itself developed through language-internal mechanisms; thus, there are segmental predictors of tone in Tsat.

Let us first consider the role of contact in the development of tone in pMM, through contact either with a Papuan language (hypotheses 1 and 3), or with pA or one of its descendants (hypothesis 2). Of the borrowing scenarios in which the Donor is tonal and the Borrower is atonal, the minimal-contact scenario can be ruled out, as the outcome is toneless loans in the Borrower language; we, however, are seeking a source for tone. The fourth scenario, that of convergence in extreme contact situations, can also be eliminated in these cases, as the tone system develops through language-internal mechanisms. As noted above, in this scenario we would expect to see segmental predictors of tone; but as shown in the previous section, no convincing segmental predictors can be identified in either Maʿya or Matbat.

This leaves two possible scenarios which led to the development of tone in pMM: the borrowing of words from a tonal language with the original tones intact, as in the Northern Thai numerals in Mal; or the borrowing of words with a special 'loan tone', as found elsewhere in Mal. The second scenario seems unlikely, as in this situation we would expect all loans to bear one tone, and all native words to bear another; this is not what we see in either Maʿya or Matbat. Of the four scenarios outlined by Ratliff, this leaves the borrowing of words from a tonal Donor with the original tones intact as the most likely explanation for tonogenesis. As noted by Remijsen (2001a:103–4), there are many words without an Austronesian etymology in Maʿya and especially Matbat, some of which are likely to have been borrowed from a pre-Austronesian substrate.[18] However, this scenario also has its problems, most notably the question of how the tones borrowed with the loan words were extended out to the native vocabulary.

In terms of the identification of the tonal Donor that induced tonogenesis in pMM, we can rule out pA and its descendants as candidates (i.e., hypothesis 2). This is due to the relative complexity of the systems of the languages concerned – it is unclear how the rich system of Matbat, which distinguishes six tones, could have developed through the borrowing of loan words from pA, which only distinguished two tones, without further language-internal developments that would have left traces of conditioning environments.[19] In lieu of further evidence, the most likely scenario is therefore that pMM developed tone through contact with another language that was once spoken in Raja Ampat. As tone is not typically associated with Austronesian languages, but is a fairly common feature of Papuan languages (Donohue 1997; Foley 1986:63–4), we can infer that the tonal Donor was a Papuan language.

---

[18] Aside from some shallow loans from Tidore, an atonal North Halmahera language that was historically dominant throughout Raja Ampat (e.g,. Maʿya *dyou* 'respectful greeting' < Tidore *jou* 'lord'; van Staden 2000), forms similar to these putative borrowings in Maʿya and Matbat have not yet been identified in any of the extant Papuan languages of the area.

[19] It is of course possible that pre-proto-Ambel had a richer tone system, which has subsequently collapsed. However, there is no independent evidence to support this analysis.

The idea that Maʿya and Matbat developed tone through contact with an extinct Papuan substrate is not new: this scenario has also been argued for in Remijsen (2001a:102–4), based on the rich tone system of Matbat, and the unusual combination of lexical tone and lexical stress in Maʿya. Remijsen concludes that these prosodic systems are unlikely to have developed independently. No Papuan languages, tonal or atonal, are spoken in Raja Ampat today (the recent – and atonal – incomers Moi and Duriankari notwithstanding). However, there are several tonal Papuan languages spoken on the Bird's Head of New Guinea, the closest mainland to Raja Ampat: for example, Mpur (Odé 2002a:50–1, 2002b), Abun (Berry & Berry 1999:20–2), Sougb (Reesink 2000, 2002:194–6), Meyah (Gravelle 2002:121–3, 2004:44–54), and Moskana (Gravelle 2010:49–55). It is possible that one or more tonal Papuan languages, possibly genetically related to one or more of these languages, may also once have been spoken in Raja Ampat.[20]

However, the prosodic systems of the tonal Papuan languages spoken on the Bird's Head are much simpler than those found in Maʿya and Matbat: Mpur and Abun are analysed with three tones (respectively: High, Mid, Low; and High/Rise, Low, Fall), Sougb and Moskana with two tones (High, Low), and Meyah with two "phonemic pitch levels" ([+HIGH] and [–HIGH]).[21] If pMM did develop tone through contact with a tonal Papuan language, either this language had a more complex tone system than the extant Papuan languages spoken on the Bird's Head; or it had a simple tone system, akin to the systems of the extant Papuan languages on the Bird's Head, but the tone system of pMM complexified after it was acquired through contact. Similar to the argument against Ambel being the tonal Donor given above, this latter hypothesis can be ruled out: again, we would expect to see evidence of conditioning by segmental features, at least in Matbat with its very complex tone system, as evidence of subsequent language-internal splits in the tone systems. The development of tone in pMM through contact with a substrate with a more complex tone system than those found in the Papuan languages of the Bird's Head is thus the more likely hypothesis.

Let us turn now to the development of tone in proto-Ambel. Returning to the contact scenarios outlined by Ratliff (2002), we can again rule out the minimal-contact situation, in which words are borrowed from a tonal Donor with no tone. The borrowing of words with a special 'loan tone' can also be eliminated, as there is no phonological distinction in the lexicon of pA between loans and native Austronesian words. This leaves two explanations for the development of tone in pA: through the borrowing of loan words with the tones intact; or through the convergence of the prosodic system of pA with a neighboring language.

---

[20] While Mpur and Abun are isolates, Sougb, Meyah, and Moskana together comprise the East Bird's Head family (Voorhoeve 1975, Reesink 2002).

[21] As only two pitches are distinguished in Sougb, Meyah, and Moskana, Reesink (2000, 2002:194–6), Gravelle (2004), and Gravelle (2010) respectively analyse the word-prosodic systems of these languages as "pitch accent" systems. However, an examination of the data presented in the analyses shows that High pitch is not obligatory in Sougb or Moskana, and that High pitch is neither obligatory nor culminative in Meyah (see Arnold 2018b:218, f.n. 21 for details). Hyman (2006, 2009) argues against the classification of word-prosodic systems as "pitch accent" systems, on the grounds that there is no one pitch accent prototype. Instead, word-prosodic systems can be typologised as stress accent systems (if there is at least and at most one metrically prominent syllable per word, i.e., if the marking of prominence is both obligatory and culminative) or tone systems (if pitch is used non-obligatorily and/or non-culminatively). Following this typology, the word-prosodic systems of Sougb, Meyah, and Moskana can be analysed as tone systems.

With regards to the convergence hypothesis, we saw above that in these cases a language is stimulated to tonogenesis using language-internal resources through contact with a tonal language. In this scenario, the prosodic systems of the languages in contact come to resemble each other. As discussed above for Maʿya and Matbat, we would expect to see segmental predictor remnants of the language-internal development of the tone system, which we do not in pA.

However, it is at this point worth returning to the similarities between the tone systems of Ambel and the Laganyan dialect of Maʿya, introduced in section 2.3. While no correspondences could be identified, the tonal phonologies of Maʿya and pA are identical: both distinguish High, Rise, and toneless syllables. The phonetic realisation of tone in Laganyan Maʿya and Metsam Ambel (the more conservative of the two Ambel dialects) is also very similar: both dialects have a HL% boundary tone, such that heavy High syllables are realised with high pitch utterance-medially, and high falling pitch utterance-finally; and heavy Rise syllables, which utterance-medially are realised with low pitch in Laganyan Maʿya and rising pitch in Metsam Ambel, are both realised with a distinctive rise-fall pitch contour utterance-finally. The similarities between the systems are summarised in table 7.

**Table 7. The Metsam Ambel and Laganyan Maʿya tonal phonologies and realisations of tone compared (similarities in boldface)**

| Tone | Syllable weight | Utterance position | Metsam Ambel | Laganyan Maʿya |
|---|---|---|---|---|
| High | Heavy | Medial | **H** | **H** |
| | | Final | **HL** | **HL** |
| | Light | Medial | **H** | **H** |
| | | Final | **H** | **H** |
| Rise | Heavy | Medial | LH | L |
| | | Final | **LHL** | **LHL** |
| | Light | Medial | LH | L |
| | | Final | **LH** | **LH** |
| Toneless | Heavy | Medial | **L** | **L** |
| | | Final | LHL | L |
| | Light | Medial | **L** | **L** |
| | | Final | LH | L |

These similarities are unlikely to be a coincidence. Both Ambel and Laganyan Maʿya are spoken around Mayalibit Bay on Waigeo, and the two groups are in close contact, including mutual bilingualism. It is likely that, once both languages had become tonal, the two systems converged, similar to the Tsat, Hainan Min, and Li systems described in Thurgood (1999). Most likely, the Laganyan Maʿya system became more similar to the Metsam Ambel system – while the Laganyan Maʿya system diverges from the other dialects of Maʿya, the tone system of Metsam Ambel is identical with that reconstructed for pA (see further Arnold submitted). However, as there are no segmental predictors

for tone in pA as there are in Tsat, this scenario still does not provide an explanation as to how tone was originally innovated in pA.

With the elimination of Ratliff's convergence scenario as an explanation for tonogenesis, this leaves one scenario: that pA, like pMM, developed tone through the integration of loanwords from a tonal Donor with the original tones intact. As discussed above, this scenario still leaves us with the unanswered question as to how the borrowed tones subsequently spread through the native lexicon. However, of the four scenarios, it seems the most likely.

There is not at present enough information available for us to confidently identify the Donor of these tonal borrowings into pA. We can tentatively rule out Matbat as a possibility: because Ambel and Matbat are spoken some distance from each other (Ambel on Waigeo, the northernmost island of Raja Ampat, and Matbat on Misool, the southernmost island), and because both groups were traditionally 'land-oriented', it is unlikely that speakers of pA and post-split Matbat were in contact.[22] However, tonogenesis in pA may have been instigated through contact with Maʿya, pMM, or a tonal Papuan substrate. In order to determine the source of pA tone, more work is required to identify the source of loan words in pA. In particular, subgrouping and reconstruction of the RASH languages is necessary to separate out Maʿya loans into Ambel from cognate forms. If pA had developed tone by borrowing loanwords with the tone intact from Maʿya (or pMM), we would expect to see some patterns: specifically, we would expect to see these loans realised with the same pitch in both Maʿya and pA. Whether this prediction is borne out awaits further research.[23]

Our conclusions in this section are shaky at best. While, through a process of elimination, we have concluded that both pMM and pA developed tone by borrowing words with the tonal specifications intact, and that at least in the case of pMM, the tonal Donor was a now-extinct Papuan substrate, several questions remain. First, how did the tones, originally confined to loanwords, spread through the native lexicon in both cases? In particular, what was the mechanism that assigned some native words one tone, and others another? Second, what was the nature of the Papuan substrate – how many languages were spoken in Raja Ampat, and what prosodic features did they have? Finally, when and for how long were speakers of the Austronesian and Papuan languages of Raja Ampat in contact, and what were the sociolinguistic relations between speakers of these languages like? As the Papuan language or languages that

---

[22] Remijsen (2001a:163ff.) contrasts the 'land-oriented' Raja Ampat groups, such as the Matbat and the Ambel, with the 'sea-oriented' Maʿya. Land-oriented groups originally lived in the interior of the islands and cultivated sago. The sea-oriented Maʿya, however, have traditionally lived in coastal areas, and have had wide-ranging trade and political contacts throughout Raja Ampat and beyond.

[23] With regards to Papuan loans: like Maʿya and Matbat, some loans from atonal Tidore can be easily identified in Ambel (e.g., Ambel *tua* 'bed' < Tidore *tua*; Ambel *kapaya* 'papaya' < Tidore *kapaya*; van Staden 2000). Besides these shallow loans, only two other words that are somewhat similar in form and meaning to Ambel words have thus far been identified in any of the Papuan languages of the area. Both are found in Kalamang, an atonal West Bomberai language spoken to the south of the Bird's Head: Ambel *kalabét* 'goanna', Kalamang *kalabet* 'earthworm'; Ambel *go* 'bamboo', Kalamang *gous* 'ibid.' (Visser 2016). These similarities may simply be due to chance. However, if there is a common origin, this suggests that the Donor of these words into Ambel was either Kalamang; a language genetically related to Kalamang; or a language with which Kalamang has also been in contact. It may have been that this unidentified language was a substrate once spoken on Waigeo. Another possibility is that historic trade links between Raja Ampat and the Bomberai Peninsula introduced these words into Ambel (Goodman 2006).

were formerly spoken in Raja Ampat have disappeared without leaving any direct evidence, we may never find satisfactory answers to these questions. However, with regards to the final question, there are several speculations that we can make regarding the timescale and nature of Austronesian-Papuan contact in Raja Ampat. It is to this discussion that we now turn.

## 5. Discussion: Austronesian-Papuan contact in Raja Ampat

Based on the conclusions reached in the preceding sections, we can make two inferences about the contact between Austronesian and Papuan languages in Raja Ampat. The first inference concerns the relative timescale of this contact; and the second inference concerns the nature of the contact.

First, regarding the relative timescale of contact. For contact to have independently affected the prosodic systems of pMM and pA, it must have occurred after the break-up of the most recent common ancestor of Maʿya, Matbat, and Ambel, but before Maʿya and Matbat split. As mentioned at several points throughout this paper, the internal subgrouping of the RASH branch of SHWNG is still unknown: while Kamholz (2014) concludes that pMM constitutes a primary branch of RASH, the data presented in Kamholz (2015) cast doubt on this subbranch. The data in the present paper do support a Maʿya-Matbat subbranch. However, it remains unclear whether Ambel, Maʿya, and Matbat form a separate subbranch of RASH, to the exclusion of, for example, the RASH languages spoken in South Halmahera, or the other RASH languages spoken in and around Raja Ampat; or whether other Raja Ampat languages can be grouped with Maʿya and Matbat, to the exclusion of Ambel, or with Ambel, to the exclusion of Maʿya and Matbat. As recommended above, detailed bottom-up comparative work is required to make progress with this question.

The second inference regards the nature of contact in Raja Ampat. Trudgill (2010:313–314) distinguishes two types of sociolinguistic situation in language contact: scenarios in which there is stable, long-term contact, involving intermarriage and childhood bilingualism; and those in which the contact is short-term, where bilinguals are typically adult second-language learners. Trudgill argues that the former type of contact scenario leads to additive complexification, as the childhood bilinguals in these situations will have learnt the second language before the critical threshold, whereas the latter scenario tends to lead to simplification. As contact between pMM and the putative Papuan substrate led to the addition of a feature in the proto-language – tone – we can infer that the contact scenario was of the former type, i.e., long-term and stable. The same reasoning can be applied to the contact between pA and the tonal Donor language, whether that Donor ultimately turns out to be Papuan or Austronesian.

Unfortunately, not enough is known of Maʿya and Matbat oral history to know how probable this scenario of intermarriage and childhood bilingualism is. In addition, if tonogenesis in pA was stimulated by contact with a Papuan language, the oral history of the Ambel appears to contradict a scenario in which relations between the Ambel and this group were friendly, involving intermarriage. Many Ambel myths and legends are centred around characters known as the *kábyo*, who appear to be an oral record of a pre-Austronesian population group that once lived on Waigeo. The *kábyo* are described in these stories as malevolent spirit beings, who manifest in human form – typically taking the form of someone known to their victim, in order to lure them away from the village to kill and eat them. The language used by the *kábyo* is recorded in a traditional fish-poisoning ritual performed by the Ambel in the settlement of Darumbab, on the north

coast of Waigeo.[24] This language does not appear to be Austronesian, in that it does not contain any recognisable Austronesian roots; hence the speculation that the *kábyo* are a memory of a now-extinct pre-Austronesian population group. Based on the stories, the *kábyo* appear to have been aggressive, and possibly cannibalistic – there certainly does not appear to have been the kind of relationship between the *kábyo* and the Ambel that would have been conducive to the complexification discussed above.

However, there are several possible scenarios that could reconcile the linguistic data with these stories of *kábyo*. First, it is still possible that tonogenesis in pA was not stimulated by contact with a Papuan language, but rather with either pMM or Maʿya. Second, if contact with a Papuan language were the trigger for tonogenesis, it could have been that relations between the Ambel and this group were historically friendly, but subsequently deteriorated. Third, it may have been that there was more than one Papuan group living on Waigeo: one (or more) may have had good relations with the Ambel, and intermarried with them, acting as the tonal Donor; and one (or more), which are recorded in the stories of the *kábyo*, may have been more antagonistic and violent towards the Ambel. Finally, the stories of the *kábyo* may be purely fictional, with no historical basis.

In the cases of tonogenesis in both pMM and pA, the nature of the prosodic system of the Donor is unknown. We can therefore only speculate as to the extent to which the tone system of the Borrower converged with that of the Donor, an outcome that Ratliff (2002) describes as the result of the most intense kind of contact. In the case of Ambel and Laganyan Maʿya, however, similarities both in the underlying phonology of the tone systems, and in the realisation of tone (including the utterance-final boundary tone) are strongly indicative that speakers of these two languages have been in extreme contact. This contact likely began before the break-up of pA; but after proto-Maʿya split into its daughter dialects.

We can make some speculations with regard to the timescale and nature of the contact between speakers of pA and Laganyan Maʿya. There is, in the present day, a religious prohibition on marriage between the two groups: whereas the Laganyan Maʿya are Muslim, the Ambel are Christian.[25] The Laganyan Maʿya were converted to Islam by the Tidore sultanate, for whom they acted as vassals in Raja Ampat (Remijsen 2001a:166). While the precise date of conversion is unknown, Raja Ampat was under the influence of the Tidore sultanate from at least the time of the arrival of Europeans in the archipelago in the early sixteenth century, until the mid-nineteenth century (Huizinga 1998), so it is likely to have occurred at some point in this time period.[26] The Ambel, however, practised traditional religion until their conversion to Christianity by European missionaries in the mid-twentieth century (Arnold 2018a:6). If intermarriage and childhood bilingualism were the sole mechanism for contact-induced change in the two languages, and presuming that the religious prohibition on intermarriage began when the Laganyan Maʿya were converted to Islam, the convergence in the tone systems must have occurred before this conversion (whenever this may have been). An

[24] See https://elar.soas.ac.uk/Record/MPI1064549 and https://elar.soas.ac.uk/Record/MPI1094119 for a documentation of a stylised performance of this ritual, and the associated dance and song.

[25] No such prohibition exists between the Ambel and either the Kawe or Wauyai Maʿya, who are also Christian. Intermarriage between these groups is common.

[26] The Tidore sultanate itself was converted to Islam at some point in the fifteenth century, before the arrival of Europeans in the area (van Staden 2000:13).

alternative explanation for the similarities between the two systems is that the mechanism was not intermarriage, but close political ties and trade. According to the oral history of the Ambel, relations between the Laganyan Maʿya and the Ambel have long been friendly – for example, the Laganyan Maʿya make frequent appearances in myths belonging to several Ambel clans, typically to help the Ambel fight off the various external raiding parties that periodically plagued the archipelago. Furthermore, one clan, the Gaman clan, is split between the two groups, such that some members identify as Laganyan Maʿya, and others identify as Ambel, further testifying to the good relations between the groups. The timescale for this kind of contact is less easily identifiable, in that it appears to have been ongoing for several centuries, up to the present day. These two explanations are of course not mutually exclusive: it may well have been the case that intermarriage occurred up to the point when the Laganyan Maʿya converted to Islam, and was followed by a long period of mutually beneficial and reciprocal political ties between the Laganyan Maʿya and the Ambel.

## 6. Conclusions and future research

Several conclusions have been reached in the course of this paper. First, based on a comparison of monosyllabic cognates in Maʿya and Matbat, it was determined that tone developed in a common ancestor to these two languages. This finding provides support for the subgrouping of Maʿya and Matbat in a single branch of RASH, to the exclusion of Ambel. Latterly, it was argued that, due to the complexity of the prosodic systems of Maʿya and Matbat, and the lack of evidence pointing to spontaneous language-internal innovation, tone first entered pMM as the result of long-term and stable contact, involving intermarriage and childhood bilingualism, with a tonal Papuan substrate with a complex tone system. Ambel, on the other hand, did not inherit its tone system from a common ancestor with Maʿya and Matbat. Instead, it was argued that pA developed tone as the result of contact, but that in this case it is unclear whether the Donor language was also Papuan, pMM, or Maʿya. Again, based on the additive nature of this change, it is likely that the contact between speakers of pA and the tonal Donor was also long-term and stable.

As stated in the title of this paper, these conclusions regarding the origins and development of tone in the Austronesian languages of Raja Ampat are only preliminary. Much more research is required to flesh out the picture presented here. Further avenues for investigation include a fuller analysis of the prosodic system of Metsam Ambel (specifically, the determination of whether Metsam Ambel combines lexical tone with lexical stress), and then a reconstruction of pA polysyllabic forms; a move towards a reconstruction of pMM, followed by a more careful inspection for any potential segmental predictors of tone in the proto-language; and the comparison of polysyllabic cognates in the three languages. In addition, more data are required to analyse the Tomolol dialect of Matbat, as well as the several other RASH languages spoken on and around Raja Ampat, such as As, Biga, and Bata. Preliminary lexical data presented in Kamholz (2016) suggest that at least some of these languages are also tonal. Further data are required, first to confirm this; second, if these languages are tonal, to analyse the tone systems, and to determine how they compare with those of Maʿya, Matbat, and Ambel; third, to investigate whether these languages have inherited their tone systems from the same source as either pMM or pA, thus contributing further to the subgrouping of RASH; and finally, to explore whether data from these languages can provide any further evidence regarding contact and change in Raja Ampat.

## Abbreviations

| | | | |
|---|---|---|---|
| pA | proto-Ambel | PAN | proto-Austronesian |
| pMM | proto-Maʿya-Matbat | PMP | proto-Malayo-Polynesian |
| RASH | Raja Ampat-South Halmahera | PCEMP | proto-Central-Eastern Malayo-Polynesian |
| SHWNG | South Halmahera-West New Guinea | PEMP | proto-Eastern Malayo-Polynesian |

## References

Arnold, Laura. 2018a. A grammar of Ambel: An Austronesian language of Raja Ampat, west New Guinea. PhD diss., the University of Edinburgh.

Arnold, Laura. 2018b. Lexical tone in Metnyo Ambel. *Oceanic Linguistics* 57. 199–220.

Arnold, Laura. Submitted. Highs and lows: Towards reconstructing the word-prosodic system of proto-Ambel.

Berry, Kevin & Christine Berry. 1999. *A description of Abun, a West Papuan language of Irian Jaya*. Canberra: Pacific Linguistics.

Blust, Robert. 1978. Eastern Malayo-Polynesian: A subgrouping argument. In S. A. Wurm & Lois Carrington (eds.), *Second International Conference on Austronesian Linguistics: Proceedings*, fascicle 1. Pacific Linguistics C-61. Canberra: The Australian National University. 181–234.

Blust, Robert. 1993. Central and Central-Eastern Malayo-Polynesian. *Oceanic Linguistics* 32. 241–293.

Blust, Robert. 1999. Subgrouping, circularity and extinction: Some issues in Austronesian comparative linguistics. In *Selected Papers From the 8th International Conference on Austronesian Linguistics*, vol. 1. 31–94.

Blust, Robert. 2005. Must sound change be linguistically motivated? *Diachronica* 22. 219–269.

Blust, Robert. 2017. Odd conditions: Context-sensitive sound change in unexpected contexts. *Journal of Historical Linguistics* 7. 322–371.

Blust, Robert & Stephen Trussel. 2010–. Austronesian Comparative Dictionary. Online database. http://www.trussel2.com/acd/.

de Vries, Lourens. 1998. Some remarks on the linguistic position of the Inanwatan language. In Jelle Miedema, Cecilia Odé, & Rien A. C. Dam (eds.), *Perspectives on the Bird's Head of Irian Jaya, Indonesia. Proceedings of the Conference, Leiden, 13-17 October 1997*. Amsterdam; Atlanta: Rodopi. 643–653.

Donohue, Mark, 1997. Tone systems in New Guinea. *Linguistic Typology* 1. 347–386.

Filbeck, David. 1972. Tone in a dialect of T'in. *Anthropological Linguistics* 14. 111–118.

Foley, William A. 1986. *The Papuan languages of New Guinea*. Cambridge: Cambridge University Press.

Goodman, Thomas E. 2006. The sosolot: An eighteenth century east Indonesia trade network. Ph.D thesis, University of Hawai'i at Mānoa.

Gravelle, Gilles. 2002. Morphosyntactic properties of Meyah word classes. In Ger Reesink (ed.), *Languages of the eastern Bird's Head*. Canberra: Pacific Linguistics. 109–80.

Gravelle, Gilles. 2004. Meyah: An East Bird's Head language of Papua, Indonesia. PhD diss., Vrije Universiteit Amsterdam.

Gravelle, Gloria J. 2010. A grammar of Moskana: An East Bird's Head language of West Papua, Indonesia. PhD diss., Vrije Universiteit Amsterdam.

Hombert, Jean-Marie, John J. Ohala & William G. Ewan. 1979. Phonetic explanations for the development of tones. *Language* 55. 37–58.

Huizinga, F. 1998. Relations between Tidore and the north coast of New Guinea in the nineteenth century. In Jelle Miedema, Cecilia Odé, & Rien A. C. Dam (eds.), *Perspectives on the Bird's Head of Irian Jaya, Indonesia. Proceedings of the Conference, Leiden, 13-17 October 1997*. Amsterdam/ Atlanta: Rodopi. 385–420.

Hyman, Larry M. 2006. Word-prosodic typology. *Phonology* 23. 225–57.

Hyman, Larry M. 2009. How (not) to do phonological typology: The case of pitch accent. *Language Sciences* 31. 213–38.

Kamholz, David. 2014. Austronesians in Papua: Diversification and change in South Halmahera-West New Guinea. Ph.D dissertation, University of California, Berkeley.

Kamholz, David. 2015. The reconstruction of Proto-SHWNG morphology. Presentation given at the 13th International Conference on Austronesian Linguistics, Academia Sinica, July 19. http://ical13.ling.sinica.edu.tw/Full_papers_and_ppts/July_19/G-55.pdf.

Kamholz, David. 2016. Tone and language contact in southern Cenderawasih Bay. Presentation given at the workshop "Contact and substrate in the languages of Wallacea", KITLV Leiden, December 1.

Kamholz, David. 2017. Tone and language contact in southern Cenderawasih Bay. *NUSA: Linguistic studies of languages in and around Indonesia* 62. 7–39.

Kamholz, David. n.d. Lexifier. Online database. http://lexifier.lautgesetz.com.

Kingston, John. 2011. Tonogenesis. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume, & Keren Rice (eds.), *The Blackwell companion to Phonology*, vol IV. Malden, MA: Wiley-Blackwell. 2304–33.

Matisoff, James, 1973. Tonogenesis in Southeast Asia. In Larry Hyman (ed.), *Consonant types and tone*, Southern California Occasional Papers in Linguistics No. 1. Los Angeles: University of Southern Colombia. 71–96.

Odé, Cecilia. 2002a. A sketch of Mpur. In Ger Reesink (ed.), *Languages of the eastern Bird's Head*. Canberra: Pacific Linguistics. 45–107.

Odé, Cecilia. 2002b. *Mpur prosody: An experimental-phonetic analysis with examples from two versions of the Fentora myth*, ELPR Publication series A1-003. Osaka: Endangered Languages of the Pacific Rim.

Polansky, E. A. 1957. Memorie van overdracht van het district Salawati (Onderafdeling Radja Ampat Afdeling Nieuw Guinea). Unpublished ms. in the archive of the Koninklijk Instituut voor Taal-, Land-, en Volkenkunde, Leiden.

Ratliff, Martha. 2002. Timing tonogenesis: Evidence from borrowing. In Patrick Chew (ed.), *Proceedings of the Twenty-Eight Annual Meeting of the Berkeley Linguistics Society: Special Session on Tibeto-Burman and Southeast Asian Linguistics*. Berkeley, CA: Berkeley Linguistics Society. 29–41.

Reesink, Ger. 2000. Sougb. *NUSA: Linguistic studies of Indonesian and other languages in Indonesia* 47. 105–29.

Reesink, Ger. 2002. A grammar sketch of Sougb. In Ger Reesink (ed.), *Languages of the eastern Bird's Head*. Canberra: Pacific Linguistics. 181–275.

Remijsen, Bert. 2001a. *Word-prosodic systems of Raja Ampat languages*. Utrecht: LOT.

Remijsen, Bert. 2001b. Dialectal variation in the lexical tone system of Maˈya. *Language and Speech* 44. 473–499.

Remijsen, Bert. 2002. Lexically contrastive stress accent and lexical tone in Maˈya. In Carlos Gussenhoven & Natasha Warner (eds.), *Laboratory Phonology VII*. Berlin: Mouton de Gruyter. 585–614.

Remijsen, Bert. 2007. Lexical tone in Magey Matbat. In Vincent J. van Heuven & Ellen van Zanten (eds.), *Prosody in Indonesian Languages*. Utrecht: LOT. 9–34.

Remijsen, Bert. 2010. Nouns and verbs in Magey Matbat. In Michael Ewing & Marian Klamer (eds.), *East Nusantara: Typological and areal analyses*. Canberra: Pacific Linguistics. 281–311.

Rivierre, Jean-Claude. 2001. Tonogenesis and the evolution of tonal systems in New Caledonia, the example of Cèmuhî. In Shigeki Kaji (ed.), *Proceedings of the symposium Cross-linguistic studies of tonal phenomena*. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa. 23–42.

Thurgood, Graham. 1999. *From ancient Cham to modern dialects: Two thousand years of language contact and change*. Oceanic Linguistics Special Publication No. 28. Honolulu: University of Hawai'i Press.

Trudgill, Peter. 2010. Contact and sociolinguistic typology. In Raymond Hickey (ed.), *The Handbook of Language Contact*. Chichester; Malden, MA: Wiley-Blackwell. 299–319.

van der Leeden, A. C. 1993. *Maˈya: A language study. A: Phonology*. Jakarta: Lipi-Rul.

van der Leeden, A. C. n.d. Maˈya dictionary, morphology, and syntax. Unfinished Ms.

van Staden, Miriam. 2000. Tidore: A linguistic description of a language of the North Moluccas. PhD diss., Leiden University.

Visser, Eline. 2016. A grammar sketch of Kalamang with a focus on phonetics and phonology. Masters thesis, University of Oslo.

Voorhoeve, C. L. 1975. *Languages of Irian Jaya: Checklist, preliminary classification, language maps, wordlists*. Canberra: Pacific Linguistics.

## Appendix: Monosyllabic cognates in Maʿya, Matbat, and proto-Ambel

Unless otherwise noted, the Maʿya forms are from the Salawati dialect.

| | | Matbat | Maʿya | Proto-Ambel |
|---|---|---|---|---|
| 1. | ‘betel leaf’ | na$^{1}$n | ʿnya$^{12}$n | *nyan |
| 2. | ‘breast’ | su$^{3}$ | ʿsu$^{3}$s | – |
| 3. | ‘canoe’ | wa$^{3}$ŋ | ʿwa$^{12}$k | *wan |
| 4. | ‘come’ | bo$^{3}$t | ʿbo$^{3}$t | – |
| 5. | ‘die’ | ma$^{12}$t | ʿma$^{12}$t | *mna$^{3}$t |
| 6. | ‘eight’ | -wa$^{3}$l | ʿwa$^{3}$l | *wa$^{3}$l |
| 7. | ‘enter’ | hu$^{3}$ŋ | ʿsu$^{3}$n | *sun |
| 8. | ‘fire’ | ya$^{3}$p | ʿla$^{12}$p | *lap |
| 9. | ‘fish’ | – | ʿdo$^{3}$n | *dun |
| 10. | ‘five’ | li$^{3}$m | ʿli$^{3}$m | *lim |
| 11. | ‘four’ | fa$^{3}$t | ʿfa$^{12}$t | *fa$^{3}$t |
| 12. | ‘full’ | fo$^{3}$n | ʿfo$^{12}$n | *fon |
| 13. | ‘give’ | be$^{21}$ | ʿbe (Misool) | *bi |
| 14. | ‘good’ | fi$^{3}$ | ʿfi$^{3}$ | *fi |
| 15. | ‘green/blue’ | bla$^{12}$w | – | *bya$^{3}$w |
| 16. | ‘ground, earth’ | ba$^{3t}$ | ʿba$^{12}$t | *ba$^{3}$t |
| 17. | ‘hear’ | no$^{41}$ŋ | ʿdo$^{12}$n | – |
| 18. | ‘kill’ | bu$^{3}$n | ʿbu$^{3}$n | *bun |
| 19. | ‘know’ | -u$^{21}$n | -ʿun (Misool) | *un |
| 20. | ‘louse’ | wu$^{3}$t | ʿu$^{3}$t | *o$^{12}$wt |
| 21. | ‘man’ | (wa$^{3}$y)ma$^{21}$n | ʿma$^{12}$n (Misool) | *ma$^{3}$n |
| 22. | ‘mother’ | ne$^{3}$n | ʿne$^{12}$n | *ne$^{3}$n |
| 23. | ‘mountain’ | he$^{3}$l | ʿye$^{3}$l | *i$^{3}$l |
| 24. | ‘mouth’ | ga$^{21}$l | ʿgal | – |
| 25. | ‘much’ | to$^{12}$ | ʿmo$^{12}$t | – |
| 26. | ‘needle’ | la$^{1}$m | – | *yam |
| 27. | ‘night’ | ka$^{1}$m | – | *gam |
| 28. | ‘person’ | ma$^{3}$t | ʿmat | *me$^{3}$t |
| 29. | ‘rice’ | fa$^{3}$s | ʿfa$^{12}$s | *fa |
| 30. | ‘rise, ascend’ | ha$^{3}$ | ʿsa$^{3}$ | *sa |
| 31. | ‘sago’ | – | ʿbi$^{3}$ | *bi$^{12}$ |

| | | Matbat | Maˈya | Proto-Ambel |
|---|---|---|---|---|
| 32. | ‘sand’ | ye$^{3}$n | ˈle$^{12}$n | *layn |
| 33. | ‘sea turtle’ | fe$^{3}$n | ˈfe$^{3}$n | *fi$^{3}$n |
| 34. | ‘seawards’ | lo$^{3}$l | ˈlo$^{3}$l | – |
| 35. | ‘see’ | -ɛ$^{3}$ŋ | -ˈe$^{12}$m | *e$^{3}$m |
| 36. | ‘shoot’ | -a$^{1}$n | ˈfa$^{12}$n | – |
| 37. | ‘snake’ | ko$^{3}$k | ˈko$^{12}$k | *kok |
| 38. | ‘swim’ | la$^{3}$s | -ˈa$^{12}$s (Misool) | *la$^{3}$ |
| 39. | ‘three’ | to$^{3}$l | ˈto$^{3}$l | *tu$^{3}$l |
| 40. | ‘tree, wood’ | ha$^{3}$y | ˈai | *a$^{3}$y |
| 41. | ‘two’ | lu$^{3}$ | ˈlu$^{3}$ | *lu |
| 42. | ‘village’ | nu$^{3}$ | ˈpnu$^{3}$ | *nu 'house' |
| 43. | ‘walk’ | – | ˈdak (Misool) | *ta$^{3}$n |
| 44. | ‘white’ | bu$^{3}$ | ˈbu$^{3}$s | *bus |
| 45. | ‘woman’ | (wa$^{1}$t)bi$^{3}$n ‘kind of mangrove’ | ˈpi$^{3}$n | *bin |