This volume contains papers describing and discussing language change in the Austronesian languages of eastern Indonesia and Taiwan. The issues discussed include the unusual development of verbal infixes in the Cendrawasih Bay languages, dialect variations, patterns of borrowing and language contact in Taiwan and in Flores-Alor-Pantar languages of Indonesia, diachronic and synchronic aspects of voice systems of Sulawesi languages, and the reconstruction of Proto Austronesian personal pronouns. This volume should be of interest to Austronesianists and historical linguists.
Preface and Acknowledgements

The 12th International Conference on Austronesian Linguistics was held in Denpasar-Bali in July 2012. The organisers are publishing a series of compilations of papers based on specific topics, and the present volume is one of the planned four volumes containing papers that describe and discuss language change in the Austronesian languages of eastern Indonesia and Taiwan. All papers have been peer-reviewed and revised before publication.

The editors would like to thank our colleagues who happily acted as referees: Kunio Nishiyama, Hsiu-chuan Liao, Sander Adelaar, René van den Berg, Arthur Holmer, David Gil, and Loren Billings. We also thank Bryce Kositz and Vida Mastrika for their editorial help and technical assistance in preparing the manuscript for publication.

About the editors

MALCOLM ROSS is an Emeritus Professor in Linguistics in the College of Asia and the Pacific at the Australian National University. His research area is the reconstruction of the histories of the Austronesian languages of New Guinea and the Pacific and of the Papuan languages of New Guinea. Until recently he also worked on the early history of Austronesian, focusing on the Formosan languages. A special interest is how one distinguishes between the outcomes of contact and of inheritance, and between the outcomes of bilingually induced change and language shift. He has published numerous articles and edited several books on the Austronesian language family.

I WAYAN ARKA is a Fellow at the Department of Linguistics, School of Culture, History and Language, College of Asia and the Pacific ANU and a senior Lecturer at Udayana University Bali (English Department and Graduate Program in Linguistics). His interests are in descriptive, theoretical and typological aspects of Austronesian and Papuan languages of Indonesia. His publications include, among others, books and journal articles on Balinese, Indonesian, Rongga, and Marori. He is currently working on Marori as part of the ARC-funded project on the Languages of Southern New Guinea (2011-2015) and on his Humboldt project on the languages of Flores (2012-2014).
# Contents

<table>
<thead>
<tr>
<th></th>
<th>Author(s)</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Emily A. Gasser</td>
<td>1 - 17</td>
</tr>
<tr>
<td></td>
<td><em>The development of verbal infixation in Cenderawasih Bay</em></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Laura C. Robinson</td>
<td>19 - 33</td>
</tr>
<tr>
<td></td>
<td><em>The Alor-Pantar (Papuan) languages and Austronesian contact in East Nusantara</em></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Philippe Grangé</td>
<td>35 - 50</td>
</tr>
<tr>
<td></td>
<td><em>The Lamaholot dialect chain (East Flores, Indonesia)</em></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>David Mead and Joanna Smith</td>
<td>51 - 78</td>
</tr>
<tr>
<td></td>
<td><em>The voice systems of Wotu, Barang-barang and Wolio: Synchronic and diachronic perspectives</em></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Yuko Kitada</td>
<td>79 - 92</td>
</tr>
<tr>
<td></td>
<td><em>The etymology of the sociative-progress circumfix in Suwawa (Gorontalo-Mongondowic)</em></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Amy Pei-jung Lee</td>
<td>93 - 107</td>
</tr>
<tr>
<td></td>
<td><em>Contact-induced sub-dialects in Toda Seediq</em></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Malcolm Ross</td>
<td>109 - 122</td>
</tr>
<tr>
<td></td>
<td><em>Revising the reconstruction of early Austronesian personal pronouns</em></td>
<td></td>
</tr>
</tbody>
</table>
1 The development of verbal infixation in Cenderawasih Bay

EMILY A. GASSE

1 Introduction

Verbal agreement in Wamesa, an Austronesian language of West Papua, Indonesia, is in most cases expressed by the addition of a prefix to the verb root. In the 2nd- and 3rd-person singular, however, this affix surfaces instead as an infix inserted after the root-initial consonant. This placement leads to cases of vowel hiatus which could easily be avoided if prefixation were to apply rather than infixation, a pattern which is dispreferred both cross-linguistically and language-internally. In synchronic terms, this can be difficult to account for non-stipulatively. The diachronic story, however, has more explanatory power. While the Cenderawasih Bay languages are, as a family, severely underdocumented, existing evidence suggests that articulatory and perceptual pressures overcame typological markedness, causing agreement prefixes to migrate into infixal position in an ancestor language of Wamesa. More specifically, coarticulation of the prefix-final vowel and the root-initial consonant, reinforced by improved discrimination of contrasts in a prominent position, led eventually to full metathesis (Blevins & Garrett (2004)’s Perceptual Metathesis), instantiated synchronically as infixation. This did not apply to the 1st-person singular or 3rd-person plural non-human agreement prefixes, despite their similar shape, as they were not present in the language in an eligible form at the time metathesis took place.

Bermúdez-Otero (2006) notes that a theory of local, myopic sound change such as that set forth by Ohala (1992) and espoused here “predicts that phonologization is blind: it is driven by local phonetic properties and operates without regard for its global effects on the phonological system”, leading, for example, to vowel hiatus. This is precisely what I argue happened here. Certain phonetic properties of the affixed verb in an earlier form of the language, in this case higher degrees of coarticulation of the prefix vowel and root-initial consonant, which originally were simply part of the normal range of variation, were phonologized so that they themselves became the target pronunciation. This led to ambiguity as to the linear order of the vowel and consonant, already strongly coarticulated, and created a situation of metathesis. Each or these changes were locally improving in terms of the specific segments which they affected, but when taken in the larger context of the word created a more marked structure.

This paper is organized as follows. In §2 I describe the pattern in question and the source of the data used here. §3.1 gives an overview of the Cenderawasih Bay languages, of which Wamesa is a part, and the distribution of infixation patterns within them, and an overview of the available data sources is given in §3.2. An argument for a single common source for these patterns is given in §3.3, with examples from several languages. §3.5
Emily A. Gasser

gives evidence that the metathesis which led to infixation was specific to the 2nd and 3rd person singular agreement markers, and §3.6 discusses the articulatory motivations for change and the path taken to produce the modern forms, with additional evidence from the Mixtecan language Trique. §3.8 and §3.7 discuss the first person singular and 3rd person plural non-human agreement markers. §4 concludes.

2 Wamesa Verbal Agreement: Background and Data

Wamesa ([wad], also called Wandamen) is a member of the South Halmahera-West New Guinea (SHWNG) branch of Austronesian (Lewis, Simons & Fennig 2013), with approximately 8,000 speakers in the south-eastern Bird’s Head of New Guinea (Henning et al. 1991). Wamesa verbs are inflected to agree with their subject in person, number, and, in the 3rd person plural, animacy. Verbal agreement morphology is found throughout the SHWNG languages of the Bird’s Head. Wamesa is typical of these languages in making an inclusive/exclusive distinction in the first person non-singular and in differentiating between singular, dual, and plural subjects; some languages use distinct agreement markers for trial subjects as well (Anceaux 1961). Though Wamesa has a set of morphologically transparent trial pronouns, the verbal agreement markers for trial subjects are identical to those used for the plural and will not be treated separately here.

The Wamesa data used here was gathered during fieldwork conducted over a period of seven months in the city of Manokwari in West Papua, Indonesia, and the nearby villages of Windesi and Sombokoro, focusing on the Windesi dialect of Wamesa. This data largely agrees with that given in previous published descriptions of Wamesa (Cowan 1955; Anceaux 1961; Ramar et al. 1983; Henning et al. 1991). The differences lie largely in the cluster-reduction patterns, or lack thereof, in the dual and plural, as well as a possible raising of /a/ to [e] following an infixed [i] in some sources. It is unclear whether these discrepancies are a matter of dialect, language change over the intervening years, or problems of interpretation and transcription of sources, particularly in the case of Cowan.

Wamesa verbal agreement markers usually appear as prefixes. The exception is in the 2nd and 3rd person singular, where the marker appears as a CV prefix, [bu-] and [di-] respectively, on vowel-initial verb roots, and as the -V- infix [-u-] or [-i-] on consonant-initial roots. Infixation is exceptionless and entirely predictable based on the shape of the verb root. The full inflectional paradigms for two Wamesa verbs are given below in (1). Wamesa is an ideal language in which to study this pattern, as it is the only language of the subgroup so far described in which infixation occurs regularly and on all verb roots.

In Wamesa, an unstressed high vowel [u] or [i] often surfaces as a glide when adjacent to another vowel, depending on speech rate, register, and other factors. The high vowels and glides are allophones in Wamesa; they are never contrastive. The infixed vowels in the 2nd and 3rd person singular are subject to the same variable reduction in unstressed positions. Stress in Wamesa is most often penultimate and falls on the infixed vowel on monosyllabic verb roots, such as [ˈri:a] ‘3sg-go’, from the root /ral/. Stressed vowels are never reduced to glides.

1 As Dalrymple & Mofu (2012) point out for the closely related language Dusner, this pattern provides a counterexample to Greenberg’s (1966) generalization that gender distinctions made in the plural pronouns of a language will also be made in the singular pronouns.

2 Cowan refers to the language as ‘Windesi’ in his work, which is the name of the dialect spoken by my consultant, but not all of the sources he used in compiling his description name the dialect from which they are drawn; Henning et al. (1991) and Ramar et al. (1983) explicitly describe the Wondama dialect, which differs phonologically in several respects from Windesi.
The development of verbal infixation in Cenderawasih Bay

(1) **Wamesa Verbal Agreement Paradigm**

<table>
<thead>
<tr>
<th></th>
<th>prefix</th>
<th>-api ‘to eat’</th>
<th>-pera ‘to cut’</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Singular</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1sg</td>
<td>/i-/</td>
<td>j-api</td>
<td>i-pera</td>
</tr>
<tr>
<td>2sg</td>
<td>/bu-/</td>
<td>bu-api</td>
<td>p&lt;u&gt;u&gt;era</td>
</tr>
<tr>
<td>3sg</td>
<td>/di-/</td>
<td>di-api</td>
<td>p&lt;i&gt;era</td>
</tr>
<tr>
<td><strong>Dual</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1du incl</td>
<td>/tur-/</td>
<td>tur-api</td>
<td>tu-pera</td>
</tr>
<tr>
<td>1du excl</td>
<td>/amur-/</td>
<td>amur-api</td>
<td>mu-pera</td>
</tr>
<tr>
<td>2du</td>
<td>/mur-/</td>
<td>mur-api</td>
<td>mu-pera</td>
</tr>
<tr>
<td>3du</td>
<td>/sur-/</td>
<td>sur-api</td>
<td>su-pera</td>
</tr>
<tr>
<td><strong>Plural</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1pl incl</td>
<td>/tat-/</td>
<td>tat-api</td>
<td>ta-pera</td>
</tr>
<tr>
<td>1pl excl</td>
<td>/amat-/</td>
<td>amat-api</td>
<td>ama-pera</td>
</tr>
<tr>
<td>2pl</td>
<td>/met-/</td>
<td>met-api</td>
<td>me-pera</td>
</tr>
<tr>
<td>3pl hum</td>
<td>/set-/</td>
<td>set-api</td>
<td>se-pera</td>
</tr>
<tr>
<td>3pl nh</td>
<td>/si-/</td>
<td>si-api</td>
<td>si-pera</td>
</tr>
</tbody>
</table>

This infixation is unexpected for several reasons. First, infixation creates marked syllable structures without any concomitant reduction of markedness either in terms of cross-linguistic tendencies or language-specific patterns, or any increase in faithfulness. The unattested form *dipera* for ‘3sg cuts’, with prefixation rather than infixation of the agreement marker, is phonotactically less marked than attested *piera*; where the former consists of simple CV syllables throughout, the latter shows a more marked structure, as well as the interruption of the stem by other morphological material. Second, infixation does not appear across the board for all CV-agreement affixes. The 3rd person plural non-human marker, /si-/ , the only other CV agreement marker, always surfaces as a prefix, never as an infix, as does the first person singular marker /i-/ . This suggests that /bu-/ and /di-/ specifically are picked out for infixation, whether with a diacritic, targeted constraints, or some other mechanism, where the other affixes in the paradigm are not.

3 Historical Approaches

3.1 The Distribution of Infixation

According to Blust (1993), verbs in Proto-Central-Eastern Malayo-Polynesian (PCEMP) showed agreement with their subject and direct object by means of a set of pronominal clitics. This is still the case in many CEMP languages, as Klamer (2002b) points out. Within CEMP, the Cenderawasih Bay (CB) languages show these clitics have been morphologized into affixes, and a subset of these, including Wamesa, Dusner (Dalrymple & Mofu 2012), Ambai (Silzer 1983), Biak (van den Heuvel 2006), Wooi (Sawaki In preparation), Iresim, Yaur, and Yeretuar (David Kamholz 2014, p.c.), and a number of others (Anceaux 1961; Silzer 1983) show analogous patterns of infixation. Though the low-level relationships between the Cenderawasih Bay languages have not yet

---

3 This is in fact the expected and attested surface form for underlying /di-it-pera/ ‘3sg uses (it) to cut’, where the applicative marker /it-/ is obscured through cluster reduction and coalescence of identical vowels.
been determined in any detail, the majority of those mentioned above are generally classified as Yapen or Biakic (Lewis et al. 2013). A family tree of the languages in question, based on the classifications in Lewis et al. (2013), is presented in Figure 1. Subgroup names are in small caps; infixing language names are in italics, and non-infixing language names are in plain text. The lists of languages within each family are not necessarily exhaustive; only those languages mentioned in this paper are included.

It appears that a daughter language of Proto-Cenderawasih Bay existed which was a common ancestor of the modern infixing languages. In this proto-language, which I will call Proto-Biak-Yapen (PBY) after the two largest subgroups of its daughter languages, the current pattern of infixation arose. I argue that this happened via a long-term process of metathesis of the affix-final vowel and the root-initial consonant, motivated by increased acoustic discrimination, followed by a reanalysis of the location of the vowel and a simplification of the resulting cluster.

![Figure 1: Eastern Malayo-Polynesian and the Distribution of Infixation](image)

### 3.2 A Note on Data Sources

The Cenderawasih Bay languages, located as they are in a remote and politically volatile province of Indonesia, are seriously under-documented. Therefore the amount and reliability of the available data varies greatly language to language. The Ethnologue (Lewis et al. 2013) lists 32 CB languages in total. For some of these, such as Meoswar, there is very little data at all beyond a few wordlists appearing in early collections, and no way of

---

4 More reconstruction is needed to establish PBY and its development; preliminary comparison of the modern languages suggests a number of possible lexical innovations, including the words for ‘tongue’, ‘feather’, ‘to live’, ‘smoke’, ‘lightning’, and a collection of animal terms. Kamholz (2014) posits the same subgroup on the basis of phonological changes and calls it Proto-Cenderawasih Bay.
telling directly whether infixation exists in the language. Anceaux (1961) gives comparative wordlists and verbal paradigms, some more complete than others, for 13 languages. This data was collected in the early 20th century by various linguists, explorers, missionaries, and Dutch colonial administrators. Greenhill, Blust & Gray (2008) provide 210-item basic wordlists for nine CB languages, drawn from a number of sources. Ambai and Biak are each the subject of a full reference grammar submitted as a dissertation (Silzer 1983; van den Heuvel 2006). Silzer’s Ambai grammar also includes a table of agreement affixes from 13 languages. Dusner is the subject of a recent sketch grammar (Dalrymple & Mofu 2012). Yusuf Sawaki provided me with a sketch of the agreement marking system of Wooi, the subject of his fieldwork and upcoming dissertation, and David Kamholz provided verbal paradigms for Moor, Iresim, Yaur, and Yeretaur. David Gil contributed data on Roon. All Wamesa data used here comes from my own fieldwork.

### 3.3 Cognacy

The first step is to establish that the agreement prefixes in the infixing languages did indeed descend from a common ancestor. In this case the strangeness of the pattern works in our favor - infixation is cross-linguistically relatively rare to begin with (Yu 2007), and the chances that it would arise independently across this many closely-related languages, and only in the 2nd and 3rd person singular, are vanishingly slim. Furthermore, the affixes in question closely resemble each other in all of these languages. Silzer’s (1983) table of full pronouns and the singular verbal agreement markers as they appear with non-infixing and infixing verb roots, respectively, in 13 Cenderawasih Bay languages is reproduced in (2). Data from Roon (Gil 2010) has been added in the final row.⁵

<table>
<thead>
<tr>
<th>Language</th>
<th>1sg Affix</th>
<th>Pron.</th>
<th>2sg Affix</th>
<th>Pron.</th>
<th>3sg Affix</th>
<th>Pron.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wooi</td>
<td>yau</td>
<td>y-/i-</td>
<td>bu-/u-</td>
<td>i</td>
<td>ty-/i-</td>
<td></td>
</tr>
<tr>
<td>Munggui</td>
<td>yau</td>
<td>y-/i-</td>
<td>w-/u-</td>
<td>i</td>
<td>ty-/i-</td>
<td></td>
</tr>
<tr>
<td>Pom</td>
<td>yau</td>
<td>y-/i-</td>
<td>w-/u-</td>
<td>i</td>
<td>di-/i-</td>
<td></td>
</tr>
<tr>
<td>Papuma</td>
<td>yau</td>
<td>y/e-</td>
<td>w-/u-</td>
<td>i</td>
<td>t/-i</td>
<td></td>
</tr>
<tr>
<td>Busami</td>
<td>yau</td>
<td>ya-</td>
<td>w-/u-</td>
<td>i</td>
<td>s/-i</td>
<td></td>
</tr>
<tr>
<td>Wamesa</td>
<td>yau</td>
<td>y-/i-</td>
<td>bu-/u-</td>
<td>i</td>
<td>di-/i-</td>
<td></td>
</tr>
<tr>
<td>Ansus</td>
<td>yau</td>
<td>y/e-</td>
<td>bu-/u-</td>
<td>i</td>
<td>d/-i</td>
<td></td>
</tr>
<tr>
<td>Serui</td>
<td>yau</td>
<td>y-/i-</td>
<td>wau</td>
<td>i</td>
<td>d/-i</td>
<td></td>
</tr>
<tr>
<td>Ambai</td>
<td>yau</td>
<td>y-/i-</td>
<td>bu-/u-</td>
<td>i</td>
<td>d/-i</td>
<td></td>
</tr>
<tr>
<td>Wabo</td>
<td>aya</td>
<td>ay/-a-</td>
<td>awa</td>
<td>b/-o-</td>
<td>d/-i</td>
<td></td>
</tr>
<tr>
<td>Kurudu</td>
<td>aya</td>
<td>ay/-a-</td>
<td>awa</td>
<td>b/-u-</td>
<td>d/-i</td>
<td></td>
</tr>
<tr>
<td>Biak</td>
<td>aya</td>
<td>y/-ya-</td>
<td>w-/u-</td>
<td>i</td>
<td>d/-i</td>
<td></td>
</tr>
<tr>
<td>Waropen</td>
<td>ya</td>
<td>y/-ya-</td>
<td>auo</td>
<td>a/-a-</td>
<td>i(y)/i-</td>
<td></td>
</tr>
<tr>
<td>Roon</td>
<td>ya</td>
<td>y(a)/-i</td>
<td>aw</td>
<td>w(a)/u</td>
<td>(t)i</td>
<td>t/-i-</td>
</tr>
</tbody>
</table>

Table (3) below provides illustrative examples of these affixes as they are used in the CB languages other than Wamesa. These tables show verb conjugations in the singular from three infixing and one non-infixing language.

---

⁵ Some language names here are altered slightly to reflect current practice and for consistency. Original transcriptions using <y> to reflect the high front glide [j] and <ng> for the velar nasal [ŋ] are preserved.

⁶ Roon also has an animacy distinction in the third person singular. Animate forms are in the table; the inanimate forms are non-infixing.
(3)

a. **Dusner (Dalrymple & Mofu 2012)**

<table>
<thead>
<tr>
<th></th>
<th><em>ors</em> ‘to stand’</th>
<th><em>man</em> ‘to see’</th>
</tr>
</thead>
<tbody>
<tr>
<td>1sg</td>
<td>y-ors</td>
<td>man</td>
</tr>
<tr>
<td>2sg</td>
<td>w-ors</td>
<td>m&lt;u&gt;an</td>
</tr>
<tr>
<td>3sg</td>
<td>ndi-ors</td>
<td>m&lt;i&gt;an</td>
</tr>
</tbody>
</table>

b. **Iresim (Kamholz p.c.)**

<table>
<thead>
<tr>
<th></th>
<th><em>areki</em> ‘to see’</th>
<th><em>ra</em> ‘to go’</th>
</tr>
</thead>
<tbody>
<tr>
<td>1sg</td>
<td>ne-j-areki</td>
<td>ne-ra</td>
</tr>
<tr>
<td>2sg</td>
<td>a-gu-areki</td>
<td>a-r&lt;u&gt;a</td>
</tr>
<tr>
<td>3sg</td>
<td>i-di-areki</td>
<td>i-r&lt;i&gt;a</td>
</tr>
</tbody>
</table>

c. **Wooi (Sawaki p.c.)**

<table>
<thead>
<tr>
<th></th>
<th><em>ihang</em> ‘to structure’</th>
<th><em>ra</em> ‘to go’</th>
</tr>
</thead>
<tbody>
<tr>
<td>1sg</td>
<td>y-ihang</td>
<td>ra</td>
</tr>
<tr>
<td>2sg</td>
<td>bu-ihang</td>
<td>r&lt;u&gt;a</td>
</tr>
<tr>
<td>3sg</td>
<td>t-ihang</td>
<td>r&lt;i&gt;a</td>
</tr>
</tbody>
</table>

d. **Non-Infixing: Waropen (Anceaux 1961)**

<table>
<thead>
<tr>
<th></th>
<th><em>ano</em> ‘to eat’</th>
<th><em>ra</em> ‘to go’</th>
</tr>
</thead>
<tbody>
<tr>
<td>1sg</td>
<td>y-ano</td>
<td>ya-ra</td>
</tr>
<tr>
<td>2sg</td>
<td>au-ano</td>
<td>a-ra</td>
</tr>
<tr>
<td>3sg</td>
<td>iy-ano</td>
<td>i-ra</td>
</tr>
</tbody>
</table>

These affixes and their interactions with the verb roots bear a striking resemblance to one another cross-linguistically. Of these languages, all but Waropen display infixation of the 2nd and 3rd person singular on at least some consonant-initial verb roots. Based on these similarities and the known relatedness of the languages, that these forms and their behavior are inherited features can hardly be doubted.

3.4 **Historical Paths to Infixation**

In his book on the subject, Alan Yu (2007) cites five possible historical sources of infixation: 1) pre-existing infixation; 2) entrapment of an affix between two formerly independent, now fused morphemes; 3) mutation of reduplication, in which later changes render the reduplication opaque; 4) morphological excrescence, predicated on the accidental similarity between internal syllables of unrelated words; and 5) metathesis. Of these five, only one, metathesis, can possibly have given rise to the Cenderawasih Bay pattern.

Strictly speaking, infixation in Wamesa is attributable to a pre-existing infixation pattern in its ancestor language, Proto-Biak-Yapen. This does not, however, explain its emergence in PBY itself, which is our focus here. It is impractical to suggest that Infixation was inherited from any more distant ancestor than PBY. If it were innovated in Proto-Cenderawasih Bay, for example, we must explain not only how infixation emerged but also why the process was reversed and prefixation restored in so many of the daughter

---

7 Some languages, such as Biak (van den Heuvel 2006), show extremely complex patterns of agreement marking overall, and only a subset of the possible surface forms are presented here. The additional variations are suggested to be subsequent developments in the language.
The development of verbal infixation in Cenderawasih Bay languages. The problem is only exacerbated by placing infixation’s source higher and higher up the family tree, as it then must have been lost in all branches but one, PBY. Far more likely than repeated undoing of the process at each split of the tree is that infixation was innovated in the one branch in which it appears, Biak-Yapen.

This leads us back to the question of how infixation arose. None of the explanations other than metathesis are applicable to the PBY case. The root-initial consonant after which the infixed vowel surfaces was not originally an independent morpheme, as comparison with cognates across the family attests. Nor is there any reason to believe that reduplication was involved, particularly as the infixed vowel is constant regardless of the other vowels in the root. And it is clear from the larger paradigms that the infixes did not arise from an accidental resemblance of unrelated forms but from the prefixes which continue to surface as such on vowel-initial roots.

Instead, infixation in PBY was a classic case of what Blevins & Garrett (1998) call Perceptual Metathesis (PM). PM occurs when the acoustic cues associated with a given segment have a particularly long duration. This persistence creates ambiguity regarding the origin of the cues in the word, leading to reinterpretation of the origin of the cues in question in a non-historical position. Blevins & Garrett list a number of features whose acoustic cues persist across a CV or VC domain, and which are thus predicted to participate in metathesis, including laterality, rhoticity, aspiration, glottalization, pharyngealization, and, relevantly for PBY, palatalization and labialization. In their view, metathesis is an extreme example of the case where loss of a high vowel leads to secondary articulation of the (formerly) adjacent consonant. Examples of metathesis of high vowels comes from genetically diverse languages. As discussed by the authors, Greek shows metathesis of a high front vowel over consonants, with earlier *VCi surfacing as VjC in the modern language, where the intervening C is a coronal. The Bantoid languages Aghem and Noni do the same for labialization; a reconstructed class 3 noun prefix *u in the proto-language is realized here as an infixed glide [w], with the round vowel prefix persisting on one language as [o] and lost in the other. PBY and its daughter languages combine these two patterns, with the persistent cues from both the *i and the *u leading to metathesis with the following consonant.

3.5 Metathesis as a Morpheme-Specific Process

A further comparison of nine Cenderawasih Bay languages plus PCEMP\(^9\) suggests that the metathesis of high vowels with a following consonant which yielded the modern infixation patterns was a morpheme-specific change, or at the very least did not occur within monomorphemic forms, even those of the appropriate phonemic shape. Basic vocabulary wordlists from the Austronesian Basic Vocabulary Database (Greenhill et al. 2008) were used to compare four infixing languages (Numfor, Ambai, Marau, and Wamesa) and five non-infixing ones (Moor, As, Biga, Minyaifuin, and Waropen). Of the 210 lexical items examined, only one cognate set stands out as a possible case of metathesis of a high vowel within a morpheme: in non-infixing Biga, the word for ‘male’ appears as wa-man, with a bilabial glide before the [m], but this form appears as mua and muang, with the corresponding vowel [u] after the [m], in the infixing languages Wamesa and Ambai, respectively. This however, is not a regular pattern: Biga wa-bin ‘woman, female’ is cognate with Ambai vivi, and Wamesa vavi, with no corresponding [u] in the root. The origin of the [u] in the Wamesa and Ambai words for ‘male’ is unclear.

\(^8\) putting aside cases of ablaut, as in Yaur, which cause changes in the vowel but are attributable to assimilation of neighbouring vowels, not reduplication.

\(^9\) as reconstructed by Blust (1993).
Further counter-evidence for a language-wide metathesis process comes from the word for ‘kill’, reconstructed by Blust as *bunuq for PCEMP. Were metathesis possible in within a morpheme this would appear to be an ideal form for it to take place in, as the first syllable has precisely the same shape as the 2nd person singular prefix in most of the PBY daughter languages. With metathesis, we would expect the *u of the first syllable to coalesce with that in the second syllable and initial *b to disappear due to cluster reduction, yielding something like nu(ː) in the modern infixing languages. The non-infixing languages should retain the initial syllable, with a modern form along the lines of bunu. In reality, while the modern cognates for the non-infixing languages do appear as pun (Minyaifun), -bun (As and Biga), muna (Waropen) and muná Moor), the word remains unreduced in the infixing ones, appearing as muni (Ambai), (ma)mun (Numfor), and mun (Wamesa). The final *q of the PCEMP form drops in all of the modern languages, and the now-final *u of the second syllable surfaces alternately as u, a, i, or Ø following regular patterns. The initial consonant is retained in all languages, nasalizing or devoicing in some, and the *u of the first syllable remains unchanged in all instances. Another word of similar shape, CEMP *bulan ‘moon’, surfaces in infixing Marau as hura, also failing to show any evidence of metathesis. The same is true of all other comparable words in these languages. This is strong evidence against metathesis occurring as a general process in Proto-Biak-Yapen.

This does not, however, preclude the occurrence of metathesis over a morpheme boundary. The account which follows is based largely on the timing relationships between the final vowel of the affix and the initial consonant of the verb root, and the timing of a heteromorphemic VC sequence may be quite different than that of a tautomorphemic one. Several authors (Ladeboged 1992; Byrd 1994; Fougeron & Steriade 1997) have claimed that intergestural timing is specified in the lexicon at the level of either the phonological word or the morpheme. Cho (1998a, b, 2001), using instrumental articulatory studies, provides evidence that timing is encoded in the lexicon at the morphemic level. In the former experiment, he examines the degree of variability in the timing of hetero- versus tautomorphemicic consonant clusters and [pi] sequences in Korean, comparing their timing in lexicalized versus non-lexicalized compounds. In both cases, he finds significantly greater deviations in the relative timing of the two segments in the heteromorphemic cases than the tautomorphemic ones, suggesting that within a morpheme, the relative timing of articulatory gestures is specified in the lexicon and therefore subject to less variation than the heteromorphemic sequences, which are not so specified.

In the second study, Cho investigates the amount of gestural overlap between /l/ or /n/ and a following /i/ both within a morpheme and across a morpheme boundary but within a single word. His findings from the earlier experiment were confirmed here, in that the variation across a boundary was again greater than that within a morpheme. Additionally, Cho finds here that the temporal overlap of the gestures is actually greater between morphemes than within them, to a statistically significant extent. This holds both for /l/, which only palatalizes in Korean before a heteromorphemic /i/, and for /n/, which palatalizes before any /i/, though the difference is greater for /l/. Based on the increased gestural overlap in cases not specified for timing in the lexicon, Cho concludes that the preference of the grammar is actually for higher levels of overlap, and therefore more efficient transmission of cues.

Extending these facts to Wamesa, we can say that change occurred only in the derived environment precisely because it was derived,10 and therefore unspecified in the lexicon with regards to the timing relationships and subject to greater variability, and possibly

10 Cho’s findings have broader implications as well, providing a historical/articulatory motivation for the various patterns of synchronic phonology which fail to apply morpheme-internally.
The development of verbal infixation in Cenderawasih Bay

greater default overlap. The order of vowel and consonant in Wamesa are reversed from that shown in the diagrams in (4), but the pattern is the same. As discussed in greater detail below, increased overlap of the articulatory gestures of the affixal high vowel and the root-initial consonant - whether caused by increased variability of realization, a preference in the grammar for higher overlap unless otherwise specified, or a combination of the two - led over time to metathesis of the segments. Within a root, the timing is specified in the lexicon, and therefore less variable and less conducive to metathesis.

3.6 Articulatory Motivations for Infixation

Where metathesis did occur, it was driven by an evolving set of articulatory and perceptual motivations. As Ohala (1992) points out, “sound change is not teleological”; it proceeds in a narrowly local fashion without regard for its effects on the phonological (or morphological) system as a whole, such that improvement in one area can lead to confusion in another, triggering further change. In this case, improved discrimination of segments encouraged greater coarticulation of the affixal vowel and root-initial consonant, which led in turn to ambiguity of the source of the high vowel’s cues, and therefore to metathesis.

I suggest that metathesis here was driven initially by a subset of verb roots and subsequently generalized to the rest of the lexicon. Perceptual metathesis can occur over a wide range of segments, but cues for labialization and palatalization will pass more strongly over some classes of segments than others. As demonstrated by the Greek case above, palatalization will pass most readily over apical segments. In Aghem and Noni, metathesis occurred over all initial segments, but labial cues are most effectively passed over velar consonants (Silverman 2006) It is in these environments that metathesis was most favored, and these cases which drove the process in the rest of the lexicon.

The sequence of changes, in brief, was this: The first step towards metathesis was increased coarticulation of the high vowel with the following consonant, leading to rounding or palatalization of that consonant. As coarticulation continued to increase, the secondary articulations of the consonant developed into a full offglide. In these forms, the segmental source of the palatalization or labialization was ambiguous, and was reanalyzed as originating solely after the consonant. This process may have been reinforced by the unstressed nature of the agreement affixes - Blevins & Garrett (2004) discuss a type of ‘compensatory metathesis’ in which the features of an unstressed vowel move into a neighboring stressed syllable - but as the stress patterns of PBY are as yet unknown this remains speculative.

Pressure from the apical-initial roots, which showed metathesis of /i/, and velar-initial forms, with metathesis of /u/, caused the paradigm to regularize (to a certain extent) and allow metathesis on roots with any initial consonant. This metathesis created an illegal cluster, which was then simplified to comply with the phonotactic requirements of PBY. The outcome of this sequence of events was the modern pattern of infixation.

---

11 It is plausible that PBY had largely penultimate stress and disyllabic roots, but this has not been definitively shown. If this is the case, the agreement affixes would often appear in pre-stress position, and compensatory metathesis may well be a factor. PCEMP verb roots are reconstructed almost exclusively as disyllabic; the CB languages favor disyllabic roots but show a fair amount of diversity nonetheless. Stress throughout Austronesian is predominantly penultimate (Klamer 2002a; Blust 2009), but information on stress in the languages of West Papua is scarce. In Wamesa, stress on verb roots is largely penultimate but not exclusively so; Biak also shows a preference for penultimate stress but is similarly mixed.
3.6.1 Offglide Formation

For the 2nd person singular, I suggest that metathesis was most strongly motivated on velar-initial stems. To support this claim, let us first turn to a similar case in Trique, a Mixtecan language of Mexico. As discussed by Silverman (2006), Trique has sequences of the form [ukʷa] and [uta], but not [uka] or [utʷa]. (Trique has very few labial consonants.) The voiced series of stops follows the same pattern. In other words, velars in the environment /u_a/ are always show a labial offglide, while alveolars never do. The [ukʷa] and [ugʷa] sequences in Trique are reconstructable as simply [uka] and [uga] (Longacre 1957), and appear as such in other Mixtecan languages. For example, the Trique word [ʒukʷa] ‘snake’ descends from Proto-Mixtecan (PM) *ʒu ‘animal’ and *kɔ ‘snake, lizard’. PM *(kɔ) and *[ka] merged into [ka], followed by compounding and labialization to yield the modern form (Silverman 2006).

In the production of any word, adjacent segments will be coarticulated to a greater or lesser degree. Silverman argues that in Proto-Trique, coarticulation of the [u] with the following velar enhanced the perceptual contrast between an [uKa] sequence (where K is either velar stop) and an [uTa] one, leading to improved lexical discrimination. Again, plain [uKa] sequences are not perceptually problematic - due to certain properties of the response of the auditory nerve, post-vocalic stops are some of the most perceptible segments in a word (Wright 2004) - but more rounded [uKʷa] sequences are slightly better. Silverman presents experimental data to support this claim.

Skou, as described by Donohue (2003, 2008), provides an example of the length to which a language will go to improve contrasts in its person agreement system - as existing contrasts were eroded by change over time, new mechanisms of person marking were repeatedly added, so that the modern language has four possible loci of agreement on the verb. The initial reinforcement of segmental contrast in this environment in PBY is a far less extreme example of the same phenomenon.

Velars are a particularly good candidate for coarticulation with an [u] vowel. Articulatorily, as discussed by Silverman, the tongue shape necessary to produce an [u] brings the dorsum farther back in the oral cavity, bringing it closer to the target of the [k] gesture and making the distance traveled by the dorsum in order to achieve closure for the stop relatively short. This reduces the time needed to achieve the target closure, leaving less time to unround the lips if the velar is to be non-round. Further, since the velar and labial rounding gestures make use of different articulators, there is no physical impediment to coarticulation. From the perceptual side, it has previously been shown that labial and velar articulations have mutually-reinforcing effects on the acoustic signal, improving ease of discrimination on the part of listeners, most relevantly those children acquiring the language (Jakobson, Fant & Halle 1961).

If greater coarticulation leads to better lexical discrimination, productions of an item with more coarticulation will be correctly interpreted a higher percentage of the time. It has repeatedly been shown that speakers match the frequency of different variants in their own productions to those in the language they hear with remarkable accuracy (see for example Labov 1994; Coleman & Pierrehumbert 1997; Zuraw 2000; Albright & Hayes 2003; and Liberman 2002, among many others). Thus if more strongly coarticulated productions of [uKa] are less often misheard as, say, [uta], over time they can be expected to make up an increasingly large proportion of actual tokens of [uKa]. Further, as the target pronunciation of the velar becomes rounder, outlying productions will also become rounder, and if these continue to improve lexical discrimination, as Silverman argues they do, gestural overlap of the [u] and the [K] will continue to increase, leading to an evolution in the target production of the sequence from original *[uKa] through intermediate *[uK̹a], with a rounded velar, finally to modern Trique [uKʷa], with a labial offglide.
It bears emphasizing that this path of change does not require any altruism (in the sense of Kingston 2002) or effort towards clear speech on the part of the speaker. Rather, of the range of degrees of coarticulation naturally produced by a speaker, certain productions are more often correctly identified by the listener and therefore slightly overrepresented compared to their actual proportion of tokens produced. Given what Hayes et al. (2009) call the Law of Frequency Matching, these more distinct, more coarticulated productions will thus be produced slightly more often, in a self-perpetuating cycle leading incrementally to the emergence of the full offglide.

This same story can be applied to Proto-Biak-Yapen as well. While in Trique the [uK] was morpheme-internal, in PBY it straddled the morpheme boundary between the 2nd person singular prefix *bu- and a velar-initial verb root. Therefore we can posit an intermediate stage in which the 2nd person singular verb marking included both the bu-prefix and labialization of a root-initial velar.

An analogous change took place in the 3rd person singular di-, where coarticulation of the [i] of the prefix led over time to a full [ʲ] offglide on the following consonant. This may have been reinforced by analogy with the offglide in the 2nd person. Though palatalization is most often caused by a following high vowel, Bettoni-Techio & Koerich (2010) for example has shown that for Brazilian learners of English, a preceding high vowel was sufficient to cause palatalization of word-final alveolar stops. The result is similar to the Greek case of *VCi → VjC described by Blevins & Garrett (1998), but with the order of the segments reversed.

### 3.6.2 Full Metathesis and Spread by Analogy

The Trique change has not proceeded past this stage, but in PBY it continued on. At this point, the language was in the second of Blevins & Garrett’s three proposed stages of diachronic metathesis, with the features of the vowel spreading over the features of the consonant and creating ambiguity as to their source. Increased overlap with the root-initial consonant would have already shortened the duration of the affixal high vowel; this, combined with the presence of the offglide after the root-initial consonant, means that the intrinsically persistent acoustic cues for rounding and palatalization will have been present over a lengthy stretch of the affixed word, and their source in the linear order of segments unclear. Learners of PBY misparsed these cues as originating after the consonant rather than before it, as historically was the case, causing full VC metathesis.

Around the same time as metathesis, the pattern was regularized to include all verb roots in the language, not just the apical- and velar-initial ones. The analogical pressure came from two sides, with metathesis on apical roots after /i/ and on velar-initial roots after /u/, leaving only the bilabial-initial roots entirely unaffected before regularization. Further, given the persistence of these vowels’ cues over all segments, magnified by the increased overlap of gestures over the morpheme boundary, even the initially unaffected segments would have been in an environment prone to metathesis, making this expansion unsurprising. The move towards distinctiveness for apical- velar-initial roots which set this change into motion was overridden by the gain in paradigm uniformity.

Some evidence for this posited intermediate stage before the spread of infixation throughout the lexicon comes from Roon (Gil 2010), a Biakic language spoken on Roon Island, just north of the Wandamen Peninsula and adjacent to the Wamesa-speaking area. Consonant-initial verbs in Roon fall arbitrarily into two conjugation classes, one of which undergoes infixation, while the other does not. Apical- and velar-initial roots can be found in both classes; if infixation originally spread cleanly throughout these environments, it has since been undone in some cases by subsequent changes. Labial-initial infixed verbs, however, are rare. Though a small handful are attested in the modern form of the language,
they are seriously underrepresented in that class (David Gil p.c.). This suggests that infixation never spread to labials in Roon as it did in languages such as Wamesa, and that only later developments in the language led to the inclusion of those few labial-initial items we do find in the infixing class.

3.6.3 Cluster Simplification

One final step is needed to take us to the forms attested in the modern infixing languages of Cenderawasih Bay. The metathesis of the prefix-final vowel and stem-initial consonant described above does not account for what happened to the prefix-initial [b] (2nd person) and [d] (3rd person). In most of the languages involved, consonant clusters are dispreferred to a greater or lesser degree. In Ambai, for example, only homorganic NC clusters are allowed, and only word-medially (Silzer 1983). Wamesa is slightly more permissive, allowing stop-glide sequences as well. The complex clusters of Biak, Numfor, and Dusner appear to have developed at a later stage, at least in part as the result of vowel deletion in certain environments in Proto-Biakic.12 If we assume, as seems warranted, that the phonotactic structure of PBY resembled that of its non-Biakic daughter languages, then the cluster formed by the prefix-initial stop and the root-initial one after metathesis of the vowel would be illegal. The root-initial consonant was preserved in the cluster simplification, due to the prominence of its position (Beckman 1998; Jun 1995). This final step brings us to the modern state of affairs, with infixation on consonant-initial roots in the 2nd and 3rd person singular and prefixation everywhere else.

3.7 1st Person Singular i- ~j-

The scenario described above brings up the question of why metathesis only occurred with the 2nd and 3rd person singular affixes and not the 1st person singular, which surfaces alternately as [i-] and [j-] in Wamesa, and the 3rd person singular non-human, [si]. Both of these affixes end in a high vowel or glide in their modern forms and would appear to be subject to the same pressures, yet neither can appear as an infix in any of the languages in question. The answer here is that neither affix existed in a form with such a final high vowel at the point at which metathesis applied, and therefore neither was subject to it.

Returning to the table of affixes in (2) above, we see that the 1st person singular prefix surfaces as [ya-] in at least some environments in Busami and Biak;13 the same is true of Dusner (Dalrymple & Mofu 2012). In Wabo and Kurudu, the allomorph which appears with consonant-initial verb roots is simply [a-]. These languages are distributed across the Biakic and Yapen subgroups, so we cannot posit the addition of an [a] to the affix as an innovation within a single branch of CB. In fact, the Biakic languages have lost the vowel

12 Some representative cognate sets are as follows. Moor is a non-infixing CB language. Ambai and Wamesa are members of the (infixing) Yapen branch of CB. Numfor and Biak are members of the (infixing) Biakic branch of CB. (Velar [k] in Biakic languages corresponds regularly to [t] in other CB languages and PCEMP.)

<table>
<thead>
<tr>
<th>Gloss</th>
<th>PCEMP</th>
<th>Moor</th>
<th>Ambai</th>
<th>Wamesa</th>
<th>Numfor</th>
<th>Biak</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘back’</td>
<td>-</td>
<td>kuru</td>
<td>karu</td>
<td>kru[ri]</td>
<td>sne</td>
<td>-</td>
</tr>
<tr>
<td>‘belly’</td>
<td>-</td>
<td>ine warĩʔa (intestines)</td>
<td>ene</td>
<td>sane</td>
<td>sne[ri]</td>
<td>sne</td>
</tr>
<tr>
<td>‘to fear’</td>
<td>*ma-takut</td>
<td>-</td>
<td>matai</td>
<td>matai[t]</td>
<td>mkãk</td>
<td>mkãk</td>
</tr>
<tr>
<td>‘to laugh’</td>
<td>*malip</td>
<td>mariʔa</td>
<td>miri</td>
<td>mari</td>
<td>mbrif</td>
<td>mrif</td>
</tr>
<tr>
<td>‘thick’</td>
<td>*kapal</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>kpor</td>
<td>kpor</td>
</tr>
</tbody>
</table>

(Greenhill et al. 2008; van den Heuvel 2006)

13 as well as Waropen, a non-infixing language.
*a in many initial syllables, as mentioned in footnote 10. More likely, then, is that this prefix can be reconstructed as *ja in PBY (and probably Proto-Cenderawasih Bay, for that matter), with a final low vowel which subsequently dropped in many of the daughter languages.

The loss of /a/ after a high vowel or glide does not appear to have been a regular change across the lexicons of the languages in question. There are instances in the wordlists of *ja and *ia sequences in PCEMP and which surface across the daughter languages as [a] or [ja], reflecting maintenance of the sequence or loss of the *j rather than the low vowel.

However, it does appear in two pronominal forms. The PCEMP 3rd person plural *sida, reduced by the time of PBY to *sia, appears in the infixing languages, including Wamesa, as se- or s-, with loss or coalescence of the *a after an *i. Similarly-shaped PCEMP *s-ia appears across the infixing languages as i '3sg pronoun’, again with loss of the *a. There is evidence cross-linguistically that reduction of clitics and affixes in agreement paradigms can be unpredictable and irregular. Donohue (2003), for example, describes the development of the verbal agreement morphology in Skou, a non-Austronesian language of north-eastern Papua, which is highly irregular both synchronically and diachronically. In particular, the Proto-Skou 1st person singular prefix *ŋ- now surfaces as Ø on most verbs, but three verbs take k-, and one takes n-, with no apparent conditioning environment for the total loss of the segment versus just the loss of nasality versus fronting of the place of articulation.

As the agreement affixes are clearly cognate across languages, the alternative to slightly irregular dropping of the inherited affixal *a is to posit the independent innovation of the same vowel in a number of languages across several branches of CB, clearly a less plausible hypothesis, especially given the tendency of agreement affixes to reduce over time rather than augment. This is further supported by the fact that the agreement prefixes appear to ultimately be descended from the same ancestor forms as the full pronouns, a common pattern cross-linguistically. Across the infixing languages, and in some non-infixing CB languages as well, the first person singular pronoun includes the final [a], appearing as [jau], [ja], or [aja]. The most plausible scenario then is that the full prefixes were reduced in their affixed state, first to *ja- by the point of PCBY, then further to j- in a subset of the modern languages. If this is the case, the high vowel would not have appeared adjacent to the root-initial consonant, and so would not have been in a position to trigger metathesis in PBY.

3.8 3rd Person Plural Non-Human si-

The final piece of the puzzle here is the 3rd person plural non-human marker si-. Though it has a very similar shape, si- is unlike bu- and di- (and like i-) in that it never appears as an infix, regardless of the shape of the root to which it is attached. This is because it was a later innovation in the language and therefore was not present when the process of metathesis was under way. Only a small subset of the Cenderawasih Bay languages have a separate agreement prefix for non-human or lower animate subjects in the 3rd person plural (or any other person/number combination, for that matter), namely Wamesa and the Biakic languages Dusner, Roon, and Biak. While complete paradigms including the plural are only available for a small number of languages, in Anceaux’s (1961) section on the verbal systems of the Yapen languages he sees fit to point it out as an unusual feature of Wamesa. As none of the other Yapen languages appear to show this form, I suggest that it developed independently in Proto-Biakic and Wamesa, perhaps

---

14 In Wamesa at least, [i] and [j] are non-contrastive.
occurring first in one language and later as a contact-induced change in the other. The agreement marker in Wamesa is si-, probably ultimately derived from the same source as the 3rd person plural human marker set-; in Dusner, Roon, and Biak it is non-cognate na-, suggesting an independent innovation. Finally, si- is the only agreement marker in the plural which does not end in /l/. This suggests that, while /l-/ may have been an independent plural number marker at an earlier stage, by the point at which si- was added to the language it had been reanalyzed as simply a part of the main agreement prefix.

4 Conclusion

Infixed is not uncommon in Austronesian languages - see for example the *-um-, *-in- prefixes reconstructed for Proto-Austronesian (Dahl 1976) and their related forms across the modern languages - but infixation which creates vowel hiatus is. Over the development of a subset of the Cenderawasih Bay languages, however, this tendency has been overcome by articulatory and perceptual pressures, yielding a typologically marked pattern which is nonetheless widespread in this group. While more work, both historical and documentary, is needed to establish the internal structure of this subgroup, infixation suggests a common ancestor language of the Yapen languages, the Biakic languages, Iresim, Yaur, and Yetaur, in which the verbal agreement prefixes present in many West New Guinea languages migrated to become infixes on verb roots of the appropriate form. This case provides another example of how locally improving changes can lead to a globally marked outcome.

References


15 The Dusner-speaking village is surrounded by Wamesa territory, and bilingualism between the two was historically common. The three remaining speakers of Dusner all speak Wamesa (Dalrymple & Mofu 2012).
The development of verbal infixation in Cenderawasih Bay


Wright, Richard. 2004. A review of perceptual cues and cue robustness. In Bruce Hayes,


The Alor-Pantar (Papuan) languages and Austronesian contact in East Nusantara

Laura C. Robinson

1 Introduction

Austronesians settled East Nusantara some 3800 years ago (Spriggs 2011), and it is widely believed that these areas were previously populated with people (Bellwood 2007, Spriggs 2011), and those people must have spoken languages that were not Austronesian. However, the area is dominated by Austronesian languages today. It is worth asking, what happened to those original languages? Why did the Timor-Alor-Pantar family survive as the westernmost relic of those earlier populations? What has been the impact of this linguistic colonization? While several authors have suggested typological changes in both Austronesian languages and non-Austronesian languages in the area can be attributed to contact (e.g., Klamer 2002, Ross 2003, Donohue 2004, Klamer et al 2008, Klamer 2012), very little lexical data has been examined. In light of the recent flood of documentation efforts on the Alor-Pantar languages and the recent reconstruction of proto-Alor-Pantar phonology and vocabulary (Holton et al 2012), we can now re-examine the contact situation in light of lexical data. We can even attempt to answer the question: did the arrival of the Austronesians pre-date the breakup of proto-Alor-Pantar?

The Alor-Pantar languages form a discreet subgroup of the Timor-Alor-Pantar (TAP) family (Schapper et al 2012), and they are relatively isolated. There are more than twenty languages in the Alor-Pantar group (Holton et al 2012), and only one Austronesian language is spoken in the regency. That language is Alorese, which is closely related to Lamaholot but should be considered a separate language (Klamer 2011). Alorese is a relative late-comer to the Alor regency, having arrived in the 14th or 15th century (Klamer 2012). Alorese was used as a lingua franca until relatively recently, but today a local form of Malay is used.

Alor, Pantar, and the intervening islands are rugged and mountainous. In many places, water is a scarce commodity. The main occupation is subsistence farming, and the largest crop is corn. Until the Dutch era began at the turn of the 20th century, most speakers of Papuan languages lived in the mountains, and there was much inter-group warfare. Speakers of Alorese, on the other hand, typically lived on the coast and subsisted on fishing. Because of the practice of warfare, villages were on mountaintops and areas that were easily defensible rather than fertile. This fractionalization has led to an extremely dense linguistic diversity, with over twenty languages spoken in an area less than 3000 km² (the land area of Alor regency is comparable to Samoa and significantly smaller than Brunei). Figure 1 shows the location of the Alor-Pantar languages in eastern Indonesia,
while is a map of the languages of Alor and Pantar, with the Austronesian language Alorese in black.

![Figure 1: Location of the Alor-Pantar languages](image)

The history of language contact in Alor and Pantar is very different from the history of language contact in Timor. Because Austronesian languages are a majority in Timor, the non-Austronesian languages of Timor have been much more profoundly affected by language contact than the non-Austronesian languages of Alor and Pantar. Moreover, each group has been in contact with different sets of Austronesian languages. This chapter focuses primarily on the history of language contact in the Alor-Pantar languages, and the Timor languages will only be mentioned where relevant. Data sources consulted for this paper are listed in Appendix A.

### 2 AN borrowings into AP languages

While there has been massive intra-family borrowing (Holton et al 2012), this section is limited to lexical influence from Austronesian languages on the Alor-Pantar branch of the Timor-Alor Pantar family. **Table 1** shows the percentage of Austronesian loanwords on a 200-word Swadesh list\(^1\) for twelve different AP languages. The average is 8%, ranging from just over 4% in

\(^1\) There are fewer than 200 words because some inappropriate items were excluded (‘snow’, ‘freeze’, ‘ice’) and some redundant items were conflated (‘hand’/‘arm’, ‘foot’/‘leg’, ‘in’/‘at’, ‘sleep’/‘lie’).
Western Pantar to over 9% in Blagar and Adang, both of which are spoken in close proximity to Alorese. While 8% may not seem like a large number, it is important to keep in mind that the Swadesh list is designed specifically to include items that are unlikely to be borrowed. To put this in context, English has 42% loans on a 1500-item word list (Grant 2009), but it has just over 15% loans on an expanded Swadesh list, and if we exclude intra-family Germanic loans (i.e., from Old Norse), then we get a figure comparable to the 8% found in the AP languages (Anthony Grant, p.c.).

Table 1: Percent of basic vocabulary borrowed from Austronesian

<table>
<thead>
<tr>
<th>Language</th>
<th>AN Loan percentage</th>
<th>Vocabulary items compared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teiwa</td>
<td>6.9%</td>
<td>188</td>
</tr>
<tr>
<td>Nedebang</td>
<td>6.0%</td>
<td>167</td>
</tr>
<tr>
<td>Kaera</td>
<td>8.4%</td>
<td>190</td>
</tr>
<tr>
<td>Western Pantar</td>
<td>4.2%</td>
<td>192</td>
</tr>
<tr>
<td>Blagar</td>
<td>9.5%</td>
<td>190</td>
</tr>
<tr>
<td>Adang</td>
<td>9.5%</td>
<td>189</td>
</tr>
<tr>
<td>Klon</td>
<td>7.3%</td>
<td>192</td>
</tr>
<tr>
<td>Kui</td>
<td>6.4%</td>
<td>188</td>
</tr>
<tr>
<td>Abui</td>
<td>6.3%</td>
<td>190</td>
</tr>
<tr>
<td>Kamang</td>
<td>6.6%</td>
<td>181</td>
</tr>
<tr>
<td>Sawila</td>
<td>7.3%</td>
<td>177</td>
</tr>
<tr>
<td>Wersing</td>
<td>7.7%</td>
<td>168</td>
</tr>
<tr>
<td>proto-Alor Pantar</td>
<td>9.4%</td>
<td>64</td>
</tr>
</tbody>
</table>

Lexical influence from Alor-Pantar languages into Alorese has been more limited. Though Klamer (2012) estimates 5% of Alorese vocabulary on a 270-item wordlist comes from AP sources, I count just 2.2% (4 items of 185), in part because I have re-classified some of these borrowings as going from Alorese into Alor-Pantar and not vice versa; see section 3.

The Timor branch of the Timor-Alor-Pantar family is more influence by Austronesian languages than the AP languages. Unlike the Alor-Pantar subgroup, which forms a geographical grouping with only minimal incursion by speakers of Alorese, the Timor non-Austronesian languages are surrounded by Austronesian languages, and lexical influence from Austronesian languages has been more profound. Schapper (2010) estimates that over 30% of the Bunaq vocabulary comes from Tetun alone, with another smaller stratum of loans from other Austronesian languages.

---

2 Anthony Grant (p.c.) calculates 15.5% loans on a 223-item Swadesh wordlist of English, “with equal proportions of French and Norse and a few strays.”
2.1 Recent borrowings

The majority of Austronesian loanwords in AP languages appear to be relatively recent, postdating the arrival of Austronesian speakers to the Alor archipelago since the 14th century. Some occur in only one or two languages. Others have a wide distribution, but nevertheless cannot be reconstructed as proto-Alor-Pantar (pAP) loans. They occur in a number of AP languages, but the consonants correspondences are not those that would be expected in cognate vocabulary, as identified by Holton et al (2012), so we assume that they were borrowed multiple times or that they were borrowed after the breakup of proto-Alor-Pantar and subsequently spread to other AP languages via diffusion. Nevertheless, some of them may be fairly old loans, which could potentially be reconstructed to lower-level proto-languages, but since the internal structure of the AP languages has not yet been worked out with any certainty (see Robinson & Holton 2013, Holton & Robinson 2014), we shall not attempt to assign these words to any proto-language. A complete catalogue of such borrowings into AP languages from outside the family is beyond the scope of this chapter, so this section is confined to loans found on a 400-item wordlist collected for most of the AP languages.

Direction of borrowing is not always easy to determine, but an item is assumed to come from Alorese if it is analyzable in Alorese, has cognates in nearby Austronesian languages, or has a clear Austronesian etymology.

The Alorese phrase *birekari* ‘children’ (example (1)) is partially analyzable in Alorese (cf., Alorese *kari* ‘small’), suggesting that the direction of borrowing is from Alorese into Teiwa, which is apparently the only AP language that has this form.

(1) Alorese *birekari* > Teiwa *biar karim* ‘children, young people’

Examples (2) through (10) must come from an Austronesian language because they have identifiable Austronesian etymologies. In the absence of evidence to the contrary, we assume they came into AP languages from Alorese, though they could have come from a different Austronesian source language.

(2) PMP *banua* ‘inhabited land, territory supporting the life of a community’ > Alorese *banna* ‘forest’ (cf., Lamaholot (Ile Ape) *bɔanawa* ‘forest’) > Retta *vana*, Adang *bana*, Kula *banan* ‘forest’

(3) Alorese *kasi* > Teiwa *kasi*, Nede bang *katela*, Kaera *kasi*, Western Pantar *kattelu*, Blagar *kasi*, Kabola *takle*, Klon *kesel*, Kui *matakel*3 ‘papaya’ (cf., Minangkabau *buah klata* ‘papaya; lit., Castilian fruit’)

(4) PMP *kulit* > Lamaholot (Lewotobi) *kuli* (cf., Tukang Besi *kuli*) > Nede bang *kuwe*, Western Pantar *kili*, Teiwa *kuwai*, Kaera *kuwal*, Kabola/Blagar (Bama) *-kol*, Adang *kuli*, Klon *ku*, Kui *kuil*, Kamang *-kul*, Abui *kul* ‘skin, bark’4 (cf., Alorese *kamang*)

(5) PMP *muntay* > Alorese *muda* (also Lamaholot, Lewolema) > Teiwa *muud*, Kaera *mud*, Western Pantar *muri*, Adang *mud*, Klon *muud*, Abui *mur*, Kamang *mu* ‘citrus tree/fruit’

(6) PMP *pitu* > A lorese *pitto* (cf., Palu’e *bitu*>) > Blagar *bititu*, Retta *bititi(toga)*, Kabola *wuitto*, Adang *itito* ‘seven’

3 This means ‘sweet potato’ in Klon, and is also part of a compound for ‘sweet potato’ in Kaera and Blagar. Since both papayas and sweet potatoes are introduced items, the etymology ‘Castilian’ seems likely.

4 Note Wersing klut is a separate borrowing, probably from Malay.
(7) proto-Central Eastern Malayo Polynesian (pCEMP) *upi > Alorese pui (cf., Kedang puiʔ) > Nedebang puya, Western Pantar puyan, Teiwa/Kaera pui, Blagar ḥupu, Adang-pu, Kabola ḥupu⁵ ‘to blow’

(8) PMP *susu > Alorese tuhu (cf., Kedang tu, Lamaholot (Lamalera) tuo) > Kaera tuu, Adang to ‘breast’⁶

(9) PMP *tali > Alorese tale (cf., Malay tali) > Teiwa tar⁷ ‘rope’

(10) PMP *zalan > Alorese (some dialects) larang > Sawila lurang ‘path, road’

Examples (11) through (14) also have identifiable Austronesian etymologies, but they seem to come from a recent ancestor of Alorese rather than modern Alorese. Example (11) and (14) appear to have been borrowed into (at least some of) the AP languages before Alorese lenited *s > h, while examples (12) through (14) must have been borrowed before Alorese lost final /l/.

(11) pAN *basbas > pre-Alorese *bese⁸ (cf., modern Alorese behe) > Adang beh, Kabola biisi ‘to hit’

(12) PMP *batuR ‘weave’ > pre-Alorese *batul (cf., modern Alorese batu, Kedang batur) ‘needle’ > Nedebang batu, Teiwa beti, Kaera baatti, Blagar batul, Adang batung ‘needle’

(13) PMP *dumpul > pre-Alorese *kumbul (cf., modern Alorese kumbu) > Kaera kumai, Blagar (Dolabang) kumal, Blagar (Nuhawala) kumbul, Kabola kumu, Adang ñume, Klon kun ‘blunt, dull’

(14) PMP *kawil > proto-Lamaholot *kawil > Nedebang kawil, Western Pantar kawa, Kaera, Blagar kawil, Wersing awil ‘fish hook’ (cf., Malay matakail, Alorese kafi, Lamaholot (Lamalera) kawi, Buru kawil)

(15) proto-Central Malay Polynesian (pCMP) *sora > pre-Alorese *soru (cf., modern Alorese horu) > Kaera sroto, Blagar horota, Retta haruata, Kaboal saroito, Adang horot, Klon horot, Kui serot, Kula swa, Kirmang surot, Sawila sorra, Wersing sor ‘sew’

For examples (16) through (22), no Austronesian etymology has yet been identified, but these words are also assumed be Austronesian in origin because they occur in Alorese and have cognates in other nearby Austronesian languages. Although it is possible they were borrowed into Austronesian from an ancestor of the AP languages, we assume this is not the case due to the limited distribution of these items within the AP languages.

(16) proto-Lamaholot *doro (cf., modern Alorese doho, Lamaholot (Lewolema, Lamalera) doru) > Blagar doho, Rettadoro ‘rub’

(17) Alorese balolo (cf., Lamaholot (Lewolema) belo, Lamaholot (Lamalera) belolo) > Blagar blolu, Retta balolu ‘tall’

⁵ Although this form occurs in a relatively large number of languages, the distribution is confined to the languages of Pantar and the adjacent Straits. It is not found in any of the languages of the main part of Alor island.

⁶ Note that Kui has -sə ‘breast’, which is likely a more recent Malay loan.

⁷ Note also Kaera sil, Blagar sal, Adang hei, Abui tila, Kui sel ‘rope’, which suggest pAP *sil and is probably unrelated.

⁸ Reconstructions labeled ‘pre-Alorese’ or ‘proto-Lamaholot’ are my tentative ad hoc reconstructions for the purposes of this paper. Much more work is needed on the history of these languages.
(18) Alorese *bappa (cf., Kedang bapa) > Blagar (Bama) -bapa, Adang bab ‘grandparent’

(19) pre-Alorese *kiki > Alorese kae (cf., Kedang keke, Tetun kīʔik) > Kaera kiki, Blagar kiki, Adang kaʔai ‘small’

(20) Alorese *kubang (cf., Kedang kubang) > Blagar kubang, Wersing kabang ‘heart’

(21) Alorese *kubang (cf., Kedang kubang) > Blagar kubang, Wersing kabang ‘heart’

(22) Alorese *kubang (cf., Kedang kubang) > Blagar kubang, Wersing kabang ‘heart’

(23) Alorese bapa (cf., Kedang bapa) > Blagar, Retta bapa ‘crocodile’

(24) Alorese *dola (cf., Bima doro) > Adang dol, Hamap doi, Klón/Kui dol ‘mountain’

(25) Alorese *doli (cf., Tetun tudik, Ngadha tuɗi, Sika tudi; Bunaq has tudiʔ from Tetun) > Adang dur, Kabola dur, Klón duur, Kui dur ‘knife’

(26) Alorese *hapo ‘wipe’ (cf., Lamaholot (Lamalera) hapu ‘erase’) > Nedeblang api, Teiwa/Kaera ap, Blagar (Nuhawala) hapo ‘wipe’

(27) Alorese *kalita (cf., Lamaholot (Ile Ape) prita) > Teiwa klita ‘grubby, dirty’, Blagar klitak, Retta karita ‘dirty’

(28) Alorese *tobang (cf., Kedang obang, Buru tobe) > Kaera -tobung, Blagar tobang, Retta-tomba ‘push’

(29) Alorese *kondo (cf., Makassarese kondó) > Blagar kondó ‘clothing’

Examples (23) through (29) are assumed to come from Alorese because they have potential cognates in nearby Austronesian languages. In these cases, the direction of borrowing is less certain, and they could actually be borrowings from AP into Alorese.

(23) Alorese *bapa (cf., Kedang bapa) > Blagar, Retta bapa ‘crocodile’

(24) Alorese *dola (cf., Bima doro) > Adang dol, Hamap doi, Klón/Kui dol ‘mountain’

(25) Alorese *doli (cf., Tetun tudik, Ngadha tuɗi, Sika tudi; Bunaq has tudiʔ from Tetun) > Adang dur, Kabola dur, Klón duur, Kui dur ‘knife’

(26) Alorese *hapo ‘wipe’ (cf., Lamaholot (Lamalera) hapu ‘erase’) > Nedeblang api, Teiwa/Kaera ap, Blagar (Nuhawala) hapo ‘wipe’

(27) Alorese *kalita (cf., Lamaholot (Ile Ape) prita) > Teiwa klita ‘grubby, dirty’, Blagar klitak, Retta karita ‘dirty’

(28) Alorese *tobang (cf., Kedang obang, Buru tobe) > Kaera -tobung, Blagar tobang, Retta-tomba ‘push’

(29) Alorese *kondo (cf., Makassarese kondó) > Blagar kondó ‘clothing’

Example (30) is borrowed into AP from Alorese, but is ultimately of Dutch origin.

(30) Dutch rekonen > Alorese rekeng > Blagar (Nuhawala) rekeng ‘to count’

Examples (31) through (35) seem to be borrowed from Malay.

(31) PMP *huaji > Malay adik > Adang diʔ, Kamang idika ‘younger sibling’ (cf., Alorese aring)

(32) Malay *baju > Klón/Kui bad, Sawila/Wersing badu ‘shirt, clothes’ (cf., Alorese kondó)

(33) Malay *panen ‘to harvest’ > Kui panen ‘garden’ (cf., Alorese nihha)

(34) PMP *tulung > Malay tolong > Abui tulung, Kamang tolon, Sawila tulong ‘help’ (cf., Alorese sambo)

(35) Malay *tawon > Nedeblang toʔon, Western Pantar taʔan, Kaera tawuŋ, Blagar taŋ ‘bee’ (cf., Alorese tabuŋ, Lamaholot (Lewolema) tewuan)

9 Potentially from PMP *baba ‘father’.
10 Note that these probably represent independent borrowings, as Blagar and Wersing are geographically separated, and the medial /b/ segments do not regularly correspond.
11 Klamer (2011) suggests that this is a loan from Makassarese into Blagar, which was subsequently borrowed into Alorese from Blagar.
12 Ultimately from Persian (Jones 2008:31).
Examples (36) through (38) seem to be borrowed from Malay, but could also have been borrowed from another Austronesian source.

(36) PMP *buni > Malay sem-bunyi (cf., Kedang boni, Wolio bunî) > Nedebang uni, Western Pantar unning, Retta bunîj, Adang füning, Abui bunua, Wersing woing ‘hide’ (cf., Alorese dafu)

(37) PMP *beli ‘price, brideprice’ > Malay beli > Nedebang ali, Kaera walo, Blagar beli, Retta beli, Adang fel, Abui bel, Kamang baila, Wersing ali ‘buy’13 (cf., Alorese hope)

(38) PMP *pusaj > Old Javanese pusəə > Malay pusar (dialectal) pusar14 > Blagar (Bama) -pusal, Retta/Blagar -pual, Kabola -pusu, Adang -puhei, Klon -puh ‘navel’ (cf., Alorese kapuhor)

Examples (39) through (42) have been borrowed from either Malay or Alorese, or both.

(39) PMP *bunga > Malaybunga, Alorese bunga > Kaera buum, Blagar, Retta buma, Adang bung, Kabola bung, Klon buum, Kui bungan ‘flower’ (cf., Makassarese bunga, Lamaholot, Lewotobi bunga)

(40) Malay sala(h), pre-Alorese *sala (cf., modern Alorese hala) > Teiwa -sai, Kaera -saing, Retta/Blagar -hala, Kabola -sala, Adang -hal, Abui -sala ‘be wrong, make a mistake’

(41) PMP *Ratus > Alorese ratu, Malay ratu > Teiwa raru, Nedebang rat, Kaera raru, Western Pantar raru, Blagar raru, Adang rau ‘hundred’

(42) PMP *tektek > Alorese take, Malay tokek > Teiwa takok, Nedebang taka (raab), Kaera tek, Western Pantar take, Blagar teke, Adang teko, Kui takok, Abui tekok, Kamang takkee ‘gecko’

Example (43) is a Portuguese loan. Some of the AP languages show a form which looks like modern Alorese, but Adang and Kabola (which are quite closely related to each other) show a form which has either been borrowed directly from Portuguese or reflects an earlier version of Alorese that retained /s/.

(43) Portuguese espada ‘sword’ > Alorese peda > Nedebang peda, Western Pantar pera, Teiwa/Kaera peed, Blagar peda, Adang sapad, Kabola spada, Klon ped, Kui peda, Sawila piida, Wersing pede ‘machete’

Example (44) appears to an Austronesian borrowing, but a source language has not yet been identified. It is likely to have been borrowed before various sound changes affected the Austronesian languages of the region, but without a more detailed understanding of the history of these languages (especially of Alorese and the various dialects of Lamaholot), it is difficult to say with certainty.

---

13 There are similar forms in many of the AP languages with the meanings ‘sell’, ‘price’, and ‘brideprice’. Often the various words are differentiated by non-productive vowel mutations. For example, Adang has fel ‘buy’, fali ‘price’, fail ‘sell’.

14 Although both Standard Indonesian and Alor Malay have pusat, many dialects of Malay have pusar(e.g., Makassar, Ternate) or puser(e.g., Jakarta, Denpasar) (David Gil, p.c.).
Example (45) actually shows regular sound correspondences in the AP languages and could be reconstructed on that basis as pAP *batar, ‘corn, maize’ (with the only irregularity being the loss of final syllable in Klon), but maize is a new world crop that was only introduced to the region in the 16th century, so assigning this word to pAP, which is presumably much earlier than the 16th century, is problematic. On the other hand, Antoinette Schapper (p.c.), notes that “[t]he lexical item batar was first used in Tetun in reference to sorghum and comes from Old Malay batari ‘sorghum’. Today in Tetun batar is typically taken as referring to maize, but it can also be used in reference to sorghum....” In Alorese, the term is generally defined as ‘maize’, but can also refer to ‘rice’ and Philippe Grangé (p.c.) suggests that both meanings are present in the Witihama dialect of Lamaholot spoken in northeast Adonara. If the word originally referred more generally to grain crops, it could be an earlier loan than its modern glosses would lead us to believe.

(45) Tetun batar > Nedebang baata, Western Panar batte, Teiwa, Kaera, Blagar, Kui batar, Adang bati?, Abui fati, Kamang patei, Sawila patar, Wersing peter ‘maize’ (cf., Alorese fata ‘corn, cooked rice’, Lamaholot (Lamalera) wata?, Kedang water ‘corn, also food’)

The possibility of reconstructing pAP *batar ‘corn, maize’ raises an issue in determining the age of loanwords. Namely, the various sound changes from the proto-language to the modern languages are not necessarily all of the same antiquity. Proto-Alor-Pantar *t, for example, remains unchanged in all the daughter languages, so is not diagnostic of age (see Holton et al 2012 and Holton & Robinson 2014 for details on the individual sound changes).

2.2 Ancient borrowings

There are twelve Austronesian borrowings ((46) through (57)) which we can reconstruct to the level of proto-Alor-Pantar.

(46) PMP *babi > Makassarese bawi > pAP *bai ‘pig’ (cf. Alorese fafe, Malay babi)
(47) PMP *baliung > Alorese bali(ng) > pAP *balin ‘axe’
(48) PMP *bituka > Alorese tuka > pAP *-tok ‘belly’
(49) PMP *buaq ‘fruit, betel nut’ > Tetun bua ‘betel nut’ > pAP *bui ‘betel nut’ (cf., Alorese ufa, Lamaholot (Lewolema) wua?, Lamaholot (Lamalera) fua ‘betel nut’)
(50) pCMP *mai > Makassarese mai, Tetun mai > pAP *mai, (cf., Alorese beta, Malay mari) ‘to come’

Note that Schapper et al (To appear) reconstruct proto-Timor-Alor-Pantar *iser, but irregular correspondences in the Alor-Pantar languages suggest that this is a more recent loan.

Proto-Alor-Pantar reconstructions can be found in Holton & Robinson (2014), supplemented by my own additional reconstructions, particularly since Holton & Robinson (2014) purposefully excludes reconstructions that are clearly borrowings from Austronesian.

Pawley (n.d.) reconstructs proto-Trans-New-Guinea *me- ‘come’ and considers the AP forms to be descended from this, but we consider the Austronesian etymology more likely (see Robinson & Holton (2012), Holton & Robinson (2014) on the wider genealogical affiliations of the Timor-Alor-Pantar languages.)
The Alo-Pantar (Papuan) languages and Austronesian contact in East Nusantara

(51) PMP *patung > Alorese patung, Malay betung > pAP *petun ‘bamboo (large, thick species)’
(52) PMP *u(R)sah > Malay rusa > Lamaholot (Lewolema) rusa (cf., Alorese ruha) > pAP *(a)rus-i (also Tetun rusa)\(^{18}\)
(53) PMP *taqun > Alorese tun > pAP *tun ‘year’
(54) PMP *wani > Lamaholot (Lamalera) wane, Tetun wani > pAP *wani ‘bee’

We can also reconstruct pAP *mugul ‘banana’, which has apparently related forms in Austronesian languages, such as Alorese muko. Denham & Donohue (2009) suggest that a term *muku ‘banana’ spread west from New Guinea prior to the arrival of Austronesians with the expansion of banana plants subsequent to their domestication in New Guinea 7000 BP. If that is true, then this is not a borrowing from Austronesian, but a borrowing from another Papuan language into both pAP and local Austronesian languages.

(55) pAP *mogol, cf., Alorese/Lamaholot muko ‘banana’

Examples (56) and (57) have identifiable Austronesian etymologies, but potential source languages have yet to be identified because the pAP forms preserve phonemes or distinctions from PMP that are lost in the modern Austronesian languages of the region.

(56) PMP *qasiRa > pAP *asir ‘salt’ (cf. Alorese/Lamaholot (Lamalera) sia)
(57) PMP *takaw > pAP *taqaw (cf., Alorese tamaku) ‘steal’

In (56), the pAP form preserves the liquid as well as the first /a/ vowel, both of which are lost in Alorese. Loss of the prepenultimate vowel is characteristic of most Central Malayo Polynesian languages (Blust 1993). In (57), we see that pAP preserves the final diphthong that was neutralized in most Central Malayo Polynesian languages (see Blust 1993). These two forms show that an Austronesian language preserving final diphthongs and prepenultimate vowels must have been in place prior to the subsequent diffusion of these sound changes to the majority of CMP languages.

2.3 Discussion

The borrowings discussed in the preceding sections seem to fall in a range of semantic domains, as shown in Table 2.

\(^{18}\) The Javan Rusa (Rusa Timorensis) is native to Java and Bali, but was introduced to this region “in antiquity” (Grubb 2005: 670).
### Table 2: Loans by semantic domain

<table>
<thead>
<tr>
<th>semantic domain</th>
<th>items</th>
<th>number</th>
</tr>
</thead>
<tbody>
<tr>
<td>numbers</td>
<td>‘hundred’, ‘seven’</td>
<td>2</td>
</tr>
<tr>
<td>nouns (35)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>people</td>
<td>‘children’, ‘grandparent’, ‘younger sibling’</td>
<td>3</td>
</tr>
</tbody>
</table>

We note, however, an interesting pattern in the age of loans. In addition to a number of recent loans from both Alorese and Malay, we also find more ancient loans that reconstruct to the level of pAP, including loanwords that preserve phonological distinctions which were subsequently lost in the Austronesian languages of the region. This suggests that the Austronesians arrived before the breakup of proto-Alor-Pantar.

### 3 AP borrowings into AN languages

There are far fewer loans from AP languages into Austronesian languages. In example (58), we see that the Baranusa dialect of Alorese has tor ‘path, road’ from the neighboring AP language Western Pantar ya tor ‘main road’ (Klamer 2012), while other dialects of Alorese have larang < PMP *zalan (see example (9)). We can easily identify Western Pantar as the source language in this case because ya means ‘path’ in Western Pantar, while ya tor is ‘main road’, so Alorese has borrowed the modifier instead of the head.

(58) Western Pantar ya tor ‘main road’ > Alorese (Baranusa) tor ‘path’

Examples (59) through (61) appear to be loans from one AP language into Alorese, though in these cases, the direction of borrowing is difficult to determine because of the limited distribution of the word in both AP and Austronesian.

(59) Teiwa haʔa (part of a well-ordered paradigm in Teiwa but not in Alorese) > Alorese haʔa ‘this’
(60) Western Pantar kolang ‘roll’ > Alorese lakong ‘turn over’
(61) Western Pantar lambing ‘wash clothes’ > Alorese lambing ‘wash vegetables’

There are a number of borrowings from into Alorese that could be from any of the AP languages, as the items are reconstructable to pAP.

(62) pAP *is(i) > Alorese ihi ‘fruit’\(^{20}\)

\(^{19}\) In the AP languages, properties, states, and numbers are frequently verbs.
The Alor-Pantar (Papuan) languages and Austronesian contact in East Nusantara

There are three potential ancient loans into the ancestor of the modern AN languages of the region. Example (65) is somewhat speculative, as the formal similarity is not perfect. Examples (66) and (67) are more solid, but, again, the direction of borrowing is less certain.

(65) pAP *lebur > proto-Lamaholot *(b)ebel ‘tongue’
(66) pAP *kir > proto-Lamaholot/Alorese *kiri ‘comb’ (cf., Alorese/Lamaholot (Lewolema) kiri, Lamaholot (Lamalera) giri)
(67) pAP *bunaq > proto-Lamaholot *banuʔ ‘smoke’

Klammer (2012: 92) also considers (68) through (75) to be loans from AP into Alorese, but we have not included these. For examples (68) through (73), we suggest that the direction of borrowing is the reverse (i.e., from Alorese into AP and not vice versa). In example (68), we show that the Alorese has an Austronesian etymology, though it is possible that pAP borrowed the form in antiquity, and Alorese and Kedang subsequently borrowed the forms from AP languages. Klammer (2011) proposes a similar scenario for (72), suggesting that Blagar borrowed the form from Makassares, and Alorese subsequently borrowed it from Blagar. Such scenarios are certainly possible, but working them out requires a more detailed understanding of the history of Alorese, Lamaholot, and their closest Austronesian relatives.

(68) PMP *baliung > Alorese bali(ng) (cf., Kedang baliʔ) > pAP *balin ‘axe’
(69) proto-Lamaholot *doro (cf., modern Alorese doho, Lamaholot (Lewolema, Lamalera) doru) > Blagar doho, Retta doro ‘rub’
(70) Alorese dur › Adang dur, Kabola dur, Klon dur, Kui dur ‘knife’
(71) Alorese kalita (cf., Lamaholot (Ile Ape) prita) > Teiwa klita ‘grubby, dirty’, Blagar klitak, Retta karita ‘dirty’
(72) Alorese kondyo > Blagar kondo ‘clothing’
(73) Alorese tobang (cf., Kedang obang, Buru tobe) > Kaera -tobung, Blagar tobang, Retta -tomba ‘push’

Klammer (2012) further considers (74) and (75) to be loans from AP into Alorese. While these are potential loans, we believe that in the case of (74), the formal similarity is not strong enough (note that (64) the pAP uvular is usually borrowed as Alorese /k/). In (75), the difference in the glosses, while not an implausible semantic change, makes the comparison less certain.

(74) Teiwa kalok, Kaera xolo, Blagar (Bama) ɣolo; Alorese ele (cf., Lamaholot, Lewoeleng ilo) ‘wet’
(75) Alorese kari ‘small’; Teiwa/Kaera/Blagar kira ‘thin’

20 This is potentially an earlier loan into pAP from an Austronesian source. Note PMP *isiʔ or *hesi ‘flesh (of humans, animals, fruits, tubers), contents’. Reflexes are found with this meaning in Alorese, Kedang, Lamaholot, and other Austronesian languages of the region (see example (44)), but only the AP languages and Alorese have a form like this with the meaning ‘fruit’. Note that the forms for ‘fruit’ and ‘meat’ are distinct in Alorese and in most AP languages.
4 Discussion

Contact between speakers of Lamaholot/Alorese and AP languages has left a significant amount of vocabulary in the AP languages (comparable to that of French on the vocabulary of English), but Lamaholot/Alorese, in contrast, have borrowed very little vocabulary from AP languages. The patterns of loanwords suggest that the first contact between Austronesian speakers and AP speakers happened before the breakup of proto-Alor-Pantar and before some widespread sound changes that subsequently affected most Central Malayo Polynesian languages.

We are still left with the question posed in the introduction: why did pre-Austronesian speakers from the Philippines to Flores switch to speaking Austronesian languages in antiquity, but in the Alor archipelago, speakers of non-Austronesian languages retained their original tongues? What was different about the Alor archipelago?

I suggest that the difference may have been the presence of agriculture or some other effective subsistence strategy among the ancestors of AP speakers. Agriculture was developed in New Guinea 10,000 years ago and may have spread west prior to the Austronesian arrival (Donohue & Denham 2010). Proto-Alor-Pantar had a word for ‘banana’, and it is possible that speakers of pAP were already cultivating bananas before the arrival of the Austronesians.

When the Austronesians arrived in the Timor region around 3800BP (Spriggs 2011), they may have been technologically superior to the indigenous people of Flores, who switched to speaking Austronesian languages. But, when they arrived in Alor and Pantar, they may have encountered an agricultural people. The agricultural people of Alor and Pantar would have had roughly equal status with the Austronesians, and thus resisted switching to speaking Austronesian languages. Only much later, when the ancestors of the Alorese speakers came to the archipelago, did shifting to an Austronesian language become more common.
### Appendix A: Data sources

#### Table 1: Sources for AP languages

<table>
<thead>
<tr>
<th>Language</th>
<th>ISO 639-3 code</th>
<th>Dialect/Location</th>
<th>No. items</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Atengmelang</td>
<td>~400</td>
<td>Schapper fieldnotes</td>
</tr>
<tr>
<td>Adang</td>
<td>adn</td>
<td>Pitungbang</td>
<td>920</td>
<td>Robinson fieldnotes</td>
</tr>
<tr>
<td>Blagar²¹</td>
<td>beu</td>
<td>Dolabang</td>
<td>~400</td>
<td>Robinson fieldnotes</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>~300</td>
<td>Steinhauer fieldnotes</td>
</tr>
<tr>
<td>Kabola</td>
<td>klz</td>
<td>Monbang</td>
<td>~400</td>
<td>Robinson fieldnotes</td>
</tr>
<tr>
<td>Kamang</td>
<td>woi</td>
<td>Bukapiting</td>
<td>~1800</td>
<td>Schapper fieldnotes, Schapper &amp; Manimau (2011)</td>
</tr>
<tr>
<td>Kaera</td>
<td>-</td>
<td>Abangiwang</td>
<td>890</td>
<td>Klamer fieldnotes, Klamer (2014)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>~400</td>
<td>Robinson fieldnotes</td>
</tr>
<tr>
<td>Klon</td>
<td>kyo</td>
<td>Bring</td>
<td>1900</td>
<td>Baird 2008, Baird fieldnotes</td>
</tr>
<tr>
<td>Kui</td>
<td>kvd</td>
<td>Moru</td>
<td>~400</td>
<td>Holton fieldnotes</td>
</tr>
<tr>
<td>Kula</td>
<td>tpg</td>
<td>Lantoka</td>
<td>~400</td>
<td>Williams fieldnotes</td>
</tr>
<tr>
<td>Nedebang</td>
<td>nec</td>
<td>Balungada</td>
<td>~400</td>
<td>Robinson fieldnotes</td>
</tr>
<tr>
<td>pAP</td>
<td></td>
<td></td>
<td>~130</td>
<td>Holton &amp; Robinson (2014), my reconstructions</td>
</tr>
<tr>
<td>Retta</td>
<td>ret</td>
<td>Bogakele</td>
<td>~400</td>
<td>Robinson fieldnotes</td>
</tr>
<tr>
<td>Sawila</td>
<td>swt</td>
<td>Lalamana</td>
<td>~1800</td>
<td>Kratochvil fieldnotes</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>~400</td>
<td>Robinson fieldnotes</td>
</tr>
<tr>
<td>Wersing</td>
<td>wrs</td>
<td>Kolana</td>
<td>432</td>
<td>Holton fieldnotes</td>
</tr>
<tr>
<td>Western Pantar</td>
<td>lev</td>
<td>Tubbe</td>
<td>2500</td>
<td>Holton &amp; Lamma Koly (2008)</td>
</tr>
</tbody>
</table>

²¹ Blagar exhibits significant dialect variation with respect to the consonants. Unless otherwise noted, the data cited in this paper are from the Dolabang dialect.
Table 2: Data sources for other languages

<table>
<thead>
<tr>
<th>Language</th>
<th>Family</th>
<th>Location</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alorese</td>
<td>AN</td>
<td>Alor and Pantar</td>
<td>Robinson fieldnotes, Klamer fieldnotes, Klamer (2011)</td>
</tr>
<tr>
<td>Bunaq</td>
<td>TAP</td>
<td>Timor</td>
<td>Schapper (2010)</td>
</tr>
<tr>
<td>Buru</td>
<td>AN</td>
<td>Buru Island, Maluku</td>
<td>Tryon (1995)</td>
</tr>
<tr>
<td>Kedang</td>
<td>AN</td>
<td>Lembata</td>
<td>Samely &amp; Barnes (2013)</td>
</tr>
<tr>
<td>Lamaholot (Lewotobi)</td>
<td>AN</td>
<td>Flores</td>
<td>Nagaya (2013)</td>
</tr>
<tr>
<td>Lamaholot (Lewolema)</td>
<td>AN</td>
<td>Flores</td>
<td>Pampus (2001)</td>
</tr>
<tr>
<td>Lamaholot (Lamalera)</td>
<td>AN</td>
<td>Lembata</td>
<td>Keraf (1978)</td>
</tr>
<tr>
<td>Lamaholot (Ile Ape)</td>
<td>AN</td>
<td>Lembata</td>
<td>Keraf (1978)</td>
</tr>
<tr>
<td>Makassarese</td>
<td>AN</td>
<td>Sulawesi</td>
<td>Cense &amp; Abdoerrahim (1979)</td>
</tr>
<tr>
<td>Ngadha</td>
<td>AN</td>
<td>Flores</td>
<td>Tryon (1995)</td>
</tr>
<tr>
<td>pAN, PMP</td>
<td>AN</td>
<td>Palu’e Island, Flores</td>
<td>Donohue fieldnotes</td>
</tr>
<tr>
<td>Palu’e</td>
<td>AN</td>
<td>Palu’e Island, Flores</td>
<td>Blust (2010)</td>
</tr>
<tr>
<td>pCMP, pCEMP</td>
<td>AN</td>
<td>Flores</td>
<td>Blust (1993)</td>
</tr>
<tr>
<td>Sika</td>
<td>AN</td>
<td>Flores</td>
<td>Tryon (1995)</td>
</tr>
<tr>
<td>Tetun</td>
<td>AN</td>
<td>Timor</td>
<td>Williams-van Klinken (2008)</td>
</tr>
<tr>
<td>Tukang Besi</td>
<td>AN</td>
<td>Sulawesi</td>
<td>Donohue (1999)</td>
</tr>
<tr>
<td>Wolio</td>
<td>AN</td>
<td>Sulawesi</td>
<td>Tryon (1995)</td>
</tr>
</tbody>
</table>

References


Spriggs, Matthew, 2011, Archaeology and the Austronesian expansion: Where are we now? Antiquity 85, 510-528.


The Lamaholot dialect chain
(East Flores, Indonesia)

PHILIPPE GRANGÉ

1 Introduction

The Lamaholot linguistic area consists of a dialect chain in which speakers of neighbouring dialects can understand each other, while the most separated dialects may be considered different languages. To date, only some dialects have been described (Keraf 1978, Pampus 2001, Nishiyama & Kelen 2007, Nagaya 2009, 2010). In this paper, I group these Lamaholot dialects into three clusters according to the geography of the islands, from west to east: Southeast Flores, Adonara and Lembata. Two small areas have a less well-defined position: Tanjung Bunga (the Cape of Flores), which is close to Adonara, and western Solor, close to Southeast Flores. Lembata has been divided into three sub-groups by Keraf (1978), but I will not discuss this sub-grouping.

In this paper, I examine phonological and lexical data in different Lamaholot dialects and subdialects. I discuss some syntactic features, principally semantic alignment (split-intransitivity) and agreement rules. I argue that the Lamaholot language originated from the islands of Adonara and/or Tanjung Flores, then spread westwards to the southeastern tip of Flores and eastwards to Lembata (except for the Kédang language area at the northeastern tip of Lembata). I discuss typological issues and finally propose some hypotheses about the emergence of the Lamaholot dialect chain.

2 The linguistic area of Lamaholot

Lamaholot is an Austronesian language spoken on the eastern tip of Flores Island and three nearby islands (Solor, Adonara, Lembata). There are approximately 200,000 Lamaholot speakers in a chain of highly diversified dialects. Keraf (1978) has conducted the only global survey of Lamaholot dialects. On the basis of a lexical study, he determined a minimum of 33 languages or dialects, among which common vocabulary using the Swadesh list was as low as 44%, with a maximum of 89%. His doctoral research focused on his native Lamaholot dialect of Lamaléra, Lembata Island.

Since this study, a few linguists who have research interests in Lamaholot have each selected and studied just one of the dialects; this is already a significant task. James Fox (pers.comm.) suggested that, like in the parable of the ‘Blind Men and the Elephant’, each

---

1 Keraf (1978) identified as many as 33 Lamaholot dialects, a number that may be overstated. His grouping of three dialect clusters differs from mine. Keraf considers that Lembata Island is divided into three clusters (namely East Lamaholot, Central Lamaholot and West Lamaholot), while the islands of Adonara, Solor, and the eastern tip of Flores belong to the West Lamaholot cluster.
linguist only understands a tiny fraction of the complete Lamaholot dialect chain, so data is still too scarce to propose a comprehensive review of this language. As a result, we get an incomplete picture of this dialect chain, and little data on the history of this language.

My position is no different: my research field lies in the eastern part of the Adonara Island. This densely inhabited region has two dialects, Dulhi and Kiwangona in Keraf (1978), that are regarded as variants of the same dialect, and tentatively labelled the ‘East Adonara dialect’. Within this dialect area variations are few and mostly of a phonological nature.

![Figure 1: The Lamaholot dialect chain, encompassing three dialect clusters: Southeast Flores, Tanjung Flores + Adonara + East Solor and Lembata.](image)

The Lamaholot dialect chain extends across an area from Sikka in the east to Kédang in the west. Sikka, Lamaholot and Kédang are undoubtedly related, but their hypothetical ancestor, Proto-Flores-Lembata, has still to be reconstructed. Considering lexical data, Doyle (2010) has convincingly argued that Kédang and Sikka separated from Lamaholot earlier than the start of dialect differentiation within Lamaholot itself.

This paper focuses on the Lamaholot area, which displays a complicated linguistic ‘elephant’ in Fox’s words. This region is relatively large and encompasses, totally or partially, four islands. It is unlikely that Lamaholot emerged simultaneously on all these islands. Positing that all Lamaholot dialects originated from one language, my hypothesis is that Lamaholot expanded from its place of origin, undergoing local diversification to give the present-day dialects. This language did not undergo separation, except in pockets on the coast of Alor where Lamaholot speakers settled (Klamer 2011).

Geographically speaking, the East Adonara dialect is positioned at the centre of the Lamaholot linguistic area. My hypothesis is that from this central geographical position (still the most fertile and populated area) Lamaholot extended westwards and eastwards. I will examine this hypothesis by comparing the East Adonara dialect to other dialects through a survey of phonological, lexical and syntactic data.

---

2 By foot and by boat, the journey to cross the Lamaholot linguistic area would take at least three days. The islands are mountainous and the seas have treacherous streams. History has kept no traces of a unique kingdom which would have ruled over the whole area encompassing the Lamaholot dialect chain; warfare and mistrust between various communities prevailed throughout history, according to oral traditions and colonial archives (Barnes 1987, 2005). The Raja of Larantuka (eastern tip of Flores) and Adonara or Sagu (north coast of Adonara Island) were the main leaders, although sporadic wars, vendettas and volatile alliances kept this region under continuous unrest. The Dutch finally brought down the powers of these Rajas around 1904.

3 Glossing: AGR: agreement; PERF: perfective; PFCT: perfect; IMPF: imperfective; DET: determiner; GEN: genitive; RED: reduplication; SG: singular; PL: plural; PRN: proper noun; A: subject of a transitive
3 Phonology: interpreting some innovations

A simple hypothesis will be tested: if an innovation has emerged in the East Flores dialects (and not in the central dialects of Adonara-Solor), the same innovation is unlikely to appear in the peripheral eastern dialects in Lembata, and vice versa. The PCMP (Proto-Central-Malayo-Polynesian) reconstructions are Robert Blust’s, quoted from Greenhill et al. (2008).

3.1 The fate of *w: the conservative Adonara dialect

Doyle (2010: 23) postulates that PMP (Proto-Malayo-Polynesian) *b and *w merged in Proto-Flores-Lembata as /w/ in some words, mostly in initial position. The Adonara and East Flores dialects have retained this /w/, while it has evolved into /f/ in Lembata (Lamaléra and other dialects), probably through a fortition from *w to *v, which can be represented as *w > *v > f. Indeed, Kédang has retained an initial /v/ for the cognates of these words, recalling a probable earlier stage of the Lembata dialects\(^4\).

<table>
<thead>
<tr>
<th>1)</th>
<th>East Flores + Adonara</th>
<th>Lembata (Lamaléra)</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>wulé, wuli(^a)</td>
<td>fuli</td>
<td>‘neck’</td>
<td></td>
</tr>
<tr>
<td>wɔwɑ̃</td>
<td>ʃɑfɔ̃</td>
<td>‘mouth’</td>
<td></td>
</tr>
<tr>
<td>wua, wuɑ̃</td>
<td>fʊɑ̃</td>
<td>‘fruit’</td>
<td></td>
</tr>
<tr>
<td>wulan</td>
<td>fʊlan</td>
<td>‘moon’</td>
<td></td>
</tr>
<tr>
<td>əwɑ̃</td>
<td>ʃɑ̃</td>
<td>‘animal’</td>
<td></td>
</tr>
</tbody>
</table>

In PCMP *bua? ‘fruit’ and *bulan ‘moon’ have been reconstructed. Thus, the evolution can be retraced from the PCMP *b to the Lembata *b > *w > *v > f. In addition, the western and central dialect clusters (Adonara and East Flores) seem more conservative: *b > w.

3.2 The *l shift to r: the innovative Adonara dialect

The shift *l > r is restricted to Tanjung Flores and Adonara:

<table>
<thead>
<tr>
<th>South East Flores</th>
<th>Tanjung Flores</th>
<th>Adonara</th>
<th>Lembata</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>laræ</td>
<td>ræær</td>
<td>ræær, ræær(^6)</td>
<td>laræ</td>
<td>‘road’</td>
</tr>
<tr>
<td>laron</td>
<td>ræœ(^a)</td>
<td>ræœ, ræœ(^)</td>
<td>laro</td>
<td>‘day’</td>
</tr>
<tr>
<td>lara</td>
<td>ræ⁢</td>
<td>ræœ, lara</td>
<td>lara</td>
<td>‘sun’</td>
</tr>
<tr>
<td>balara</td>
<td>bœrærar</td>
<td>bœræra</td>
<td>balaræ</td>
<td>‘sick, ill’</td>
</tr>
</tbody>
</table>

All these examples share the same consonant context. Furthermore, ‘day’ and ‘sun’ obviously share the same etymon. In Sikka, the cognates are lalang ‘road’, lerong [lərɔŋ] structure; O: object; S: subject of an intransitive structure; S\(_A\): agent-like S; S\(_P\): patient-like S; NP: nominal phrase; PP: prepositional phrase.

\(^4\) There is an inaccuracy in the data gathered by Doyle (2010: 23): there is no /v/ in Léwoléma (East Flores) according to Pampus’ dictionary and my own data.
'day', and in Kédang lala ‘road’, loyo ‘day’. Therefore, the *l was obviously common to the Flores–Lembata group before it split into Proto-Sikka, Proto-Lamaholot and Proto-Kédang. In East Flores, the shift is restricted to the medial position only: *l > r / V_V. In Adonara, Léwoléma and Tanjung Bunga (Tanjung), the shift *l > r has also occurred in initial position: *l > r / #_.

The Tanjung Flores dialects share the same feature as the Adonara–Solor dialect. This is not surprising considering that this area is very close to Adonara Island. More problematic is the fact that the Adonara–Tanjung Flores dialects seem to be more innovative than their western and eastern neighbours. Interestingly, the peripheral dialects (eastern and western Lamaholot dialect clusters) are similar. In addition, if we compare the neighbouring languages, comparing Sikka to the west and Kédang to the east, we observe some stems with striking resemblances. The following table summarises these features, taking raran ‘road’ in Adonara–Tanjung Flores dialects as an example.

<table>
<thead>
<tr>
<th>Sikka</th>
<th>Lamaholot dialect chain (dialect clusters)</th>
<th>Kédang</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Southeast Flores</td>
<td>Adonara, Flores, Tanjung Eastern, Lembata</td>
</tr>
<tr>
<td>lalan</td>
<td>lara</td>
<td>larã</td>
</tr>
<tr>
<td></td>
<td></td>
<td>lala</td>
</tr>
</tbody>
</table>

We find this general scenario:
- PCMP *zalan > Proto-Flores-Lembata *lalan. Sikka and Kédang retained this form, although Kédang lost the final nasal or nasalisation.
- Proto-Lamaholot *lalan > *laran, which can be represented as *l > r / V_V.
- The peripheral dialects (in Southeast Flores and Lembata) remained at this step, Southeast Flores lost the final nasal, while the Adonara-Tanjung Flores dialects shifted further: *l > r / #_, for example, *laran > rarã.

Regarding the *l > r / #_ shift, Solor is split in two, e.g. West Solor (Ritaébang) [larã] ‘road’, following the Southeast Flores trend, as opposed to East Solor (Lamakéra) [rarã], similar to Adonara. There are some other examples of west–east dialectal splits within the island of Solor, showing that East Solor consistently corresponds to Adonara. A comparable split between the two areas in Solor has been noted by Rappoport (2010), regarding the traditional ‘diphonic singing’ song style.

In this particular case, the two peripheral areas (Southeast Flores and Lembata) prove more conservative than the Adonara–Tanjung Flores dialects. The Adonara–anjung Flores dialect cluster evolved independently, but these innovations do not contradict the hypothesis of Adonara as the homeland of Lamaholot, because the peripheral dialects remain similar.

3.3 The lack of innovations in peripheral dialects

Some innovations appear in only one of the three Lamaholot dialect clusters. No evidence was found of any innovation being shared by the peripheral dialects, but not by the central dialect (Adonara–Tanjung Bunga). In other words, if a lexeme in Adonara–Tanjung Bunga is conservative (close to the PCMP etymon), the innovations in the peripheral dialects (both Southeast Flores and Lembata) will be different from each other.
In the following example, Adonara and Lembata are closer to the PCMP etymon *taʔun ‘year’. A shift appears in the Southeast Flores dialects: *t > s / #.

<table>
<thead>
<tr>
<th></th>
<th>Tanjung Flores and Southeast Flores</th>
<th>Adonara</th>
<th>Lembata</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘year’</td>
<td>sũ, suŋ</td>
<td>tun, tuʔun</td>
<td>tun, tuŋ, tunən</td>
</tr>
</tbody>
</table>

The Southeast Flores lexeme is not borrowed from Sikka [ʔliwaŋ] or [ʔliwa] ‘year’, while Kédang is similar to Adonara and Lembata tun. Therefore, it seems reasonable to say that Proto-Lamaholot was *tun.

In the following example, it is unclear where the innovation appeared, because the lexeme is totally different from the reconstructed PCMP *ma-basəʔ ‘wet’. However, the eastern and central dialects merge, while some lexemes from Southeast Flores dialects seem to have shifted initial *n- to d-.

<table>
<thead>
<tr>
<th></th>
<th>Southeast Flores</th>
<th>Adonara, Tanjung Flores</th>
<th>Lembata</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘wet’</td>
<td>dəəmãŋ, dəmeʔ?</td>
<td>nəmʔe</td>
<td>nəmãŋ, sənəbe</td>
</tr>
</tbody>
</table>

Unfortunately, no data are available regarding the reconstruction of this word in Proto-Lamaholot. Furthermore, no insights are gained from Sikka gema ‘wet’ or Kédang ritaʔ ‘wet’.

A shift may be apparent in Western and Central dialect clusters, while the figure is more varied in the Eastern clusters (Lembata).

<table>
<thead>
<tr>
<th></th>
<th>Southeast Flores</th>
<th>Adonara, Tanjung Flores</th>
<th>Lembata</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘new’</td>
<td>wuʔun, wuʔuŋ</td>
<td>wuʔun, wuʔuŋ</td>
<td>wu, wun, fu, wərun, fərun</td>
</tr>
</tbody>
</table>

We find Sikka wərun and Kédang wərun ‘new’. Given the PCMP etymon *baʔəru/*baʔəru ‘new’, a Proto-Lamaholot *wərun is plausible.

It seems that part of Lembata adopted an innovation common to all Lamaholot dialects, *r > ? / V_V, but loss of the glottal stop then entailed the merger of the vowels (syneresis or coalescence); hence wuʔun > wun and the regular shift described above *w > *v > f, for instance *wun > wu > fu. However, some Lembata dialects turn out to be more conservative and have retained the probable Proto-Lamaholot *wərun, and in some cases applied the regular change *w > *v > f, for instance *wərun > fərun.

Another example of unshared innovation is the general tendency to create lexical metathesis, for instance, in Lembata.5 There is an excrescence of a f, between the stem and the person agreement suffix, as in -nək ‘1sg.agr’ (Adonara dialects) > -fkən ‘1sg.agr’ (Lembata dialects), for instance pana ‘walk’, go panafkã ‘I walk’.

---

5 Many other metatheses in Alorese have been reported by Klamer (2011) and also in a very tiny sub-dialect of Adonara, on the northeastern coast of Adonara, which seems to originate from a ‘returning’ settlement from Lembata.
Note that the Adonara dialect cluster is not always more conservative than its neighbours. But in the case of an innovation, the innovation is not shared by the neighbouring dialect clusters.

In summary:

- If Adonara proves to be more conservative, an innovation is never shared between the peripheral dialects of Southeast Flores and Lembata.
- If only Adonara presents an innovation, a more conservative phonology will be shared by Southeast Flores and Lembata.
- Adonara innovations are often shared in common with one of the peripheral dialect clusters, as the innovation had spread eastwards (to Lembata) or westwards (to Flores).

This is coherent with the central geographical position of Adonara in the dialect chain. This, however, is not sufficient evidence for stating that Adonara is the Lamaholot place of origin; the lexicon must also be examined and compared.

### 4 Lexicon: the central link of the dialect chain

Comparing the lexicons of the Lamaholot dialects is a challenge. Only Keraf (1978) has carried out this task by comparing a revised Swadesh list of 200 words in 33 Lamaholot dialects (a number potentially overestimated). Moreover, a word in dialect A may also be apparent in dialect B, but it may have a different meaning, or be judged as archaic, formal, colloquial, vulgar, or even bounded to certain registers, for instance ritual speech.

I limit this section to the observation of a few lexemes, selected because in at least one dialect cluster they are not cognate with their synonyms in other dialect clusters. These lexical data are displayed in the table (6) below. Apart from the regular phoneme shifts mentioned in the previous section, my data show examples of innovations that are not shared by all the dialect clusters.

<table>
<thead>
<tr>
<th>6)</th>
<th>Dialect clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Southeast Flores</td>
</tr>
<tr>
<td>to hear</td>
<td>bain, baïŋ</td>
</tr>
<tr>
<td>leaf</td>
<td>ləpə</td>
</tr>
<tr>
<td>skin</td>
<td>kuli, kulet</td>
</tr>
<tr>
<td>meat, flesh</td>
<td>matan, manaŋeŋ</td>
</tr>
<tr>
<td>rotten</td>
<td>òwaŋ, bɔdwa</td>
</tr>
<tr>
<td>what?</td>
<td>a, aa</td>
</tr>
<tr>
<td>dull, blunt</td>
<td>məmaŋən, məno</td>
</tr>
<tr>
<td>old</td>
<td>tua, tuʔun</td>
</tr>
<tr>
<td>in, inside</td>
<td>tə one, onoonŋ, si one, pe onən</td>
</tr>
<tr>
<td>to burn</td>
<td>tuno, sərəu, deo</td>
</tr>
</tbody>
</table>

---

6 For the stems displayed in table (6), the corresponding PCMP reconstructions, quoted from Greenhill et al. (2008) are as follows: *dəŋəR ‘to hear’; *daun ‘leaf’; *kulet ‘skin’; *isi ‘meat, flesh’; *buRuʔ (or) busuk ‘rotten’; *apa (or) *sapa ‘what?’; *dumpul ‘dull, blunt’; *tuʔa ‘old’; *dalaŋ ‘inside’; *tuno ‘burn’.

7 However, two small areas are difficult to classify: Tanjung Flores and Eastern Solor, which sometimes merge with Adonara and Southeast Flores, respectively. For instance hewa ‘to hunt’ (from ewan ‘animal’?) is similar in Tanjung Flores and Southeast Flores, while in Adonara we have bati, batin ‘to hunt’.
Many of Adonara’s innovations are shared, either by Southeast Flores or by Lembata. Nevertheless, each of the three clusters may show striking difference from the others, as shown in table (7) below; but in this case, the central cluster (Adonara) often proves more conservative. The underlined stems recall a PCMP etymon.\(^8\)

<table>
<thead>
<tr>
<th>7)</th>
<th>Southeast Flores</th>
<th>Adonara &amp; Solor, Tanjung Flores</th>
<th>Lembata</th>
</tr>
</thead>
<tbody>
<tr>
<td>to come</td>
<td>saga, hewo</td>
<td>-ai (2sg: mai), beto</td>
<td>baso</td>
</tr>
<tr>
<td>to hit</td>
<td>barin, bariŋ</td>
<td>palei, tubuk, holə</td>
<td>tado, tubuk</td>
</tr>
<tr>
<td>bird</td>
<td>kolən</td>
<td>manuk, kolən</td>
<td>kukak, kolo</td>
</tr>
<tr>
<td>to blow</td>
<td>bu</td>
<td>puit(^9)</td>
<td>die, didʒi, pur</td>
</tr>
<tr>
<td>good</td>
<td>məa</td>
<td>məla</td>
<td>sare, sənarəŋ</td>
</tr>
<tr>
<td>thick</td>
<td>bate, budʒet</td>
<td>.powerəɾə, powerəɾa^n</td>
<td>fəɾa^n, pəɾəɾe</td>
</tr>
<tr>
<td>back (of body)</td>
<td>kolan, kolaŋ</td>
<td>wohə</td>
<td>uhuk, punuʔu</td>
</tr>
</tbody>
</table>

When the Adonara stem looks to be more conservative (closer to a PCMP etymon), the Southeast Flores and Lembata dialect clusters have innovated in diverging ways. However, the three clusters may also have innovated independently from each other, while none of the stems seems to correspond to a PCMP etymon, as shown by the three last examples in (7).

To sum up, on the one hand, many of Adonara’s innovations are shared, either by Southeast Flores, or by Lembata. On the other hand, Southeast Flores and Lembata normally don’t share the same innovation if Adonara is more conservative (closer to the PCMP stem). This suggests that the Adonara–East Solor–Tanjung Bunga cluster (geographically at the central position in the dialect chain) must be the most conservative dialect, compared to those in the western and eastern clusters.

5 Syntax: the semantic alignment erosion

Considering syntactic features, the Adonara–East Solor–Tanjung Bunga cluster again seems more conservative than the neighbouring dialect clusters. The Adonara dialect in particular presents the most complex inflection system, with attributive agreement, genitive agreement, NP boundary markers and a full paradigm of object pronouns, for example, a relatively sophisticated morphophonology (including consonant epen thesis and vowel alternation). Moreover, the complex semantic alignment system in Adonara can be regarded as a striking particularity.

These morphological features are not completely absent from the Southeast Flores and Lembata dialects: some are similar, some are absent, and some have been lexicalised. In this section, I will only deal with the semantic alignment system (split intransitivity). Outside Adonara, the semantic alignment systems are almost entirely lexicalized. In Lamaléra (Lembata) for instance, the remains of a former semantic alignment system are

\(^8\) The corresponding PCMP reconstructions, quoted from Greenhill et al. (2008), are as follows: *mai ‘to come’; *palu ‘hit’; *manuk ‘bird’; *upi (but *Siup in PAn) ‘to blow’; *ma-pia (or) *diaʔ ‘good’; *upi ‘blow’; *kapal (or) *telu ‘thick’; *mudi ‘back’.

\(^9\) The corresponding PCMP etymon is *upi, and *Siup in PAn ‘to blow’ (Greenhill et al. 2008). In the Adonara cluster, puit ‘to blow’ probably originates, through a metathesis, from *tiup. Metathesis is quite common in the Lamaholot lexicon.
noticeable through the lexical morphology of certain verbs. It seems that the semantic alignment system was once the common rule in Lamaholot, but that it subsequently became highly eroded in the peripheral dialects, only remaining active in Adonara dialects.

5.1 The semantic alignment system: the conservative Adonara dialect

Semantic alignment in Lamaholot consists of a ‘split intransitivity’ system, as defined by Dixon (1994: 70-110). Following Dixon, if we name A the agent of a transitive structure, S the subject of an intransitive structure, and O the patient in a transitive structure, then a split intransitive S will be marked identically to either O or A (or will remain unmarked if this is normally the case for A). Marking S as O leads to the interpretation that the subject of the intransitive sentence plays the semantic role (theta-role) of a patient (hence symbolised $S_P$), rather than that of an agent. Conversely, when S is marked as A (i.e., in Lamaholot, unmarked), playing an agent-like semantic role, it is labelled $S_A$.

The Adonara dialects display a complex split intransitivity system; a Split-S (lexically fixed alignment) and a Fluid-S (fluid alignment) are triggered by a series of contexts, including aspect. The agreement morpheme consists of a suffix that cross-references the subject person$^{10}$. These agreement suffixes are displayed in the following table.

<table>
<thead>
<tr>
<th>Person</th>
<th>Subject pronoun</th>
<th>Adonara dialect agreement suffixes$^{11}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1sg</td>
<td>go</td>
<td>-k ; -nek ; -kek</td>
</tr>
<tr>
<td>2sg</td>
<td>mo</td>
<td>-no ; -ko</td>
</tr>
<tr>
<td>3sg</td>
<td>na</td>
<td>-na ; -ʔ</td>
</tr>
<tr>
<td>1pl.incl</td>
<td>tité</td>
<td>-net ; -ket</td>
</tr>
<tr>
<td>1pl.excl</td>
<td>kamé</td>
<td>-nem ; -kem</td>
</tr>
<tr>
<td>2pl</td>
<td>mio</td>
<td>-né ; -kê</td>
</tr>
<tr>
<td>3pl</td>
<td>ra</td>
<td>-na ; -ra</td>
</tr>
</tbody>
</table>

Semantic alignment is also present in some other East Nusantara languages. Klamer (2008: 249) mentioned nine East Nusantara languages, both Autraloan and non-Austronesian (Papuan-related), that display a semantic alignment system. Indeed, semantic alignment can be found elsewhere in Indonesia, for example, in Acehnese (northern tip of Sumatra).

However, the Adonara Lamaholot dialects prove especially sophisticated, not only because Split-S and Fluid-S coexist, but because under Fluid-S the agreement encompasses a wealth of semantic features. This depends on how the speaker considers the subject referent, whether he has control over the process and is affected by the process, whether the verb is a motion verb (completed or not by a prepositional phrase) and whether a perfective aspect is expressed.

---

$^{10}$ The agreement morphology applied to intransitive verbs in Adonara dialects is comparable to French for a subset of intransitive verbs referring to ‘change of state’ or ‘displacement’, which in the passé composé require the auxiliary être ‘to be’, and must agree with the subject: elle est allée ‘she went’ (-e: feminine agreement morpheme), ils sont descendus ‘they descended’ (-s: plural agreement morpheme).

$^{11}$ In the Adonara sub-dialects of Ilé Boleng and Kiwang Ona, the first consonant of the agreement suffix may differ, for instance /j/ or /t/ instead of /n/.
5.2 Split-S: lexically fixed semantic alignment

Lexically fixed semantic alignment (fixed intransitive split, Split-S) implies that while some intransitive verbs are always marked, others are never marked. Split-S is relevant to verbs expressing feeling or biological functions, or to sentience verbs of perception and proprioception. For instance, so’ot ‘be afraid’ must be suffixed by an agreement morpheme:

8) Go so’ot -ek k-o’on aho.
1SG afraid -1SG 1SG- with dog
‘I am afraid of dogs.’

9) *Go so’ot k-o’on aho.
1SG afraid 1SG- with dog
‘I am afraid of dogs.’

Examples of intransitive verbs (e.g., stative verbs expressing feelings) that always require agreement with their subject include kedok’o- ‘be surprised’, mia- ‘be ashamed’, so’ot- ‘be afraid’, suke- [ sukə] ‘be sad’, beken- ‘be angry’, senan- ‘be happy’, too- ‘be of the same mind, be in sympathy’.

Verbs of biological function verbs, for instance, the verb ewun- ‘to sweat’ must be affixed for agreement:

10) Go ewun -ek. Mo ewun -no nya ewun -na
1SG dream -1SG 2SG dream -2SG 3SG dream -3SG
‘I sweat.’ ‘You sweat.’ ‘He/She sweats.’

Other examples of biological functions include te’ure- ‘to dream’, béle- ‘to grow up’, piè- [ pie ] ‘to shut up, to be silenced’, bekot- ‘to emerge, to appear’, suku- ‘to suffocate, to smother’ and keduhu- ‘to cough’.

Regarding sentience verbs, we must distinguish the verbs of perception (unmarked: S_A) from the verbs of proprioception (marked: S_P). For instance, dénge ‘to hear’ will never be marked for agreement, while hopâ- ‘be tired, out of breath, exhausted’ must agree with the subject.

5.3 Fluid-S: control over the process (S_A), or not (S_P)

Fixed semantic alignment (Split-S) relies mainly on the trait ‘±control’: whether or not the subject referent has control over the process. Unsurprisingly, this trait still plays an important role in fluid semantic alignment (Fluid-S). For an intransitive verb under Fluid-S, the S will be marked (S_P) by a verbal agreement if the referent of S does not control the process.

The speaker can oppose, for instance the verb géka ‘to laugh’, without agreement (the subject retains control over his amusement) or with agreement, signalling the lack of control over the process.

11) Go géka. Go géka -nek
1SG laugh 1SG laugh -1SG
‘I laugh.’ ‘I burst out laughing, I have the giggles.’
This category of Fluid-S may affect verbs like énénén ‘to dance’, goka ‘to fall’, or even verbs of biological function, for example, méké ‘urinate’, over which one may lose control of under certain circumstances.

5.4 Fluid-S: SA unaffected, SP affected

The Fluid-S system also distinguishes between whether the subject’s referent is affected or not by the process. The label ‘affected’ means that the subject’s referent undergoes a change of state, for example, receives a new property or is deprived of a property. The more a subject’s referent is affected, the more it is considered as patient-like.

Generally, this Split-S coincides with the animate-inanimate opposition: an animate referent is more likely to be affected by the process, and vice versa. But if S refers to an animate, usually this referent is affected and the verb agrees, showing the patient-like role of the subject, marked SP.

<table>
<thead>
<tr>
<th>12)</th>
<th>(S unaffected) S ⇒ SA</th>
<th>(S affected) S ⇒ SP</th>
</tr>
</thead>
<tbody>
<tr>
<td>tudak ‘to hinder, to stop’</td>
<td>Go tudak. ‘I stop’</td>
<td>Go tudak...ek. ‘I am stuck, hampered’ (by a lack of money, a ceremonial obligation, etc.)</td>
</tr>
</tbody>
</table>

5.5 Fluid-S: perfect aspect marking

Fluid semantic alignment (Fluid-S) depends on the semantics of S, but it may also be triggered by aspect marking. This feature applies to stative verbs and to dynamic intransitive verbs, especially verbs of displacement or motion. For stative verbs in an attributive structure, an agreement morpheme is compulsory. The –n suffix is the 3sg agreement morpheme, whereas –ka signals in addition a perfect aspect, more precisely a perfect of result, for example:

a. Lolon bete -n.
   leaf be.green -3sg
   ‘The leaves are green.’

   Lolon bete -ka.
   leaf be.green -3sg.pfct
   ‘The leaves have become green.’

---

12 In other East Nusantara languages such as Klom (Klamer 2008:235-237) and Tanglapui (Klamer 2008: 239-243), the semantic alignment relies principally on whether S undergoes, or not, a change of state.

13 Stative verbs may also be used as adjectives within a NP. In this case, they will be affected by a ‘NP boundary marker’ –n, if they are the final morpheme within this NP, regardless of the NP head which is not always a noun, but may be a personal pronoun of 1st, 2nd or 3rd person.

14 Comrie (1976: 56-58) proposes this definition of the ‘perfect of result’: ‘a present state is referred to as being the result of some past situation.’ The last example of Comrie’s section about the ‘perfect of result’ deals with the Mandarin Chinese particle -le, and could also fit in with my description of the Lamaholot 3rd person suffixes -ka, -na or -ra: ‘the verbal particle –le indicates perfective aspect and relative past time reference […] With stative predicates, the force of this particle –le is often to indicate a state resulting from some previous situation, as in dōngxi gui-le ‘things are expensive’ (but with the implication that once they were not, i.e., they have become expensive).’
In (12)b, bete-ka ‘have become green’ expresses a perfect of result, because the state (‘being green’) results from a prior unmentioned process. For dynamic intransitive verbs, the agreement suffix may also express aspect. Displacement verbs,\textsuperscript{15} for instance, can be marked for perfect aspect, as in (14)).

13) \textit{Kamé géré hēti Balawéling.}\\
\smallskip
\textsuperscript{1PL.EXCL} ascend up NPR\\
‘We go upward to Balawéling.’

14) \textit{Kamé géré -kem hēti Balawéling.}\\
\smallskip
\textsuperscript{1PL.EXCL} ascend \textsuperscript{-1PL.EXCL} up NPR\\
‘We have gone upward to Balawéling.’

Unsurprisingly, an imperfect aspect marker like mētė ‘be …-ing’ in not compatible with the agreement suffix indicating the perfect aspect, as in (16). It is the opposite for the perfect aspect marker kaē, which optionally follows an agreeing verb, as in (15)).

15) \textit{Lusi rua (*mētė) beka -ka (kaē).}\\
enagle two (*IMPF) fly -3PL (PFCT)\\
‘Two eagles have flown away (already).’

16) \textit{Lusi rua (mētė) beka. (*kaē)}\\
enagle two (IMPF) fly PFCT\\
‘Two eagles fly (are flying).’

5.6 \textit{Summary of semantic alignment in Lamaholot Adonara dialects}\\
The semantic alignment system in East Adonara can be regarded as a striking particularity, because it combines Split-S and Fluid-S, as outlined in the following table.

<table>
<thead>
<tr>
<th>Fixed and fluid semantic alignment in East Adonara Lamaholot</th>
<th>Fluid semantic alignment (Fluid-S):</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed semantic alignment (Split-S):</td>
<td></td>
</tr>
<tr>
<td>Sentience verbs</td>
<td>S ±control</td>
</tr>
<tr>
<td>Verbs of biological functions</td>
<td>S ±affected</td>
</tr>
<tr>
<td>Feeling and psychological states</td>
<td>Aspect ±perfect</td>
</tr>
</tbody>
</table>

The main trait on which the semantic alignment rests is ‘±control’. It is manifest when we consider the Split-S and the Fluid-S rules. This could be related to a former East Adonara \textit{état de langue} where the ±control feature accounted for an all-embracing fluid alignment system, as is observed currently in Kédang (a neighbouring language at the eastern end of the Lamaholot dialect chain).

It appears that in the Lamaholot dialects westward from Adonara (Southeast Flores) and in the eastern dialects (on Lembata Island), the semantic alignment systems are highly eroded. For instance, the verb for ‘eat’ is the only verb that inflects for person and number, and agrees with S and A arguments in Lewotobi dialect (Nagaya 2010: note 3). It is also

\textsuperscript{15} In addition, there is fixed semantic alignment between dynamic intransitive verbs of ‘movement’ and ‘displacement’. Each of these verb categories splits further, depending on the presence or absence of a locative prepositional phrase, but this is not discussed in this paper.
the only documented occurrence of $S_P$ agreement (indicating that $S$ is deprived of control) and seems linked to the presence of a locative PP: *tedu* ‘to collide’ (Nagaya 2010:6). In the Lamaléra dialect description, the remains of a former semantic alignment system are totally lexicalized (see for instance Keraf (1978:111), and no occurrences of Fluid-$S$ can be found. Only the example of *fai* ‘water’ > *faika* ‘become water’ (Keraf 1978:214), presented as a lexical derivation, recalls East Adonara perfect aspect marking.

Holton (2008: 275) underlined a cross-linguistic tendency for the ‘erosion’ of semantic alignment systems. In the East Adonara dialect, the semantic alignment system is especially lively and systematic, retaining for a large part a Fluid-$S$. It has, however, faded away and/or been largely lexicalised in the ‘peripheral’ areas of the Lamaholot dialect chain, as represented in the map below.

![Figure 2: Active and eroded Split Intransitivity systems within the Lamaholot dialect chain.](image)

### 6 A proposed scenario for the formation of the Lamaholot dialect chain

The phonological, lexical and grammatical data suggest that the Adonara cluster (including Tanjung Bunga and Eastern Solor) hosts the most conservative Lamaholot dialects. Geographically, the Adonara dialect cluster lies at the centre of the dialect chain, and as a middle link this cluster seems to be at the origin of the western and eastern peripheral clusters.

It is unlikely that only the East Adonara dialect evolved a more complex semantic alignment system than the neighbouring dialects, where semantic alignment is highly eroded and generally lexicalized. Moreover, the remnants of semantic alignment in each of the peripheral Lamaholot dialects are very different\(^\text{16}\), and so they must originate from a morphologically more complex protolanguage.

There are almost no historical sources on the Lamaholot language area predating the sixteenth century Portuguese archives. From lexical data, Fernandes (1996) has argued convincingly that Lamaholot and Kédang originated from a protolanguage he named

\(^{16}\) However, to date we can only compare a small subset of the dialects that constitute the Lamaholot dialect chain.
Proto-Flores-Lembata. Sikka also probably originated from this protolanguage. The emergence of Proto-Flores-Lembata is obviously related to the arrival of Austronesian settlers on the coasts of Flores, Adonara, Solor and Lembata. An archaeological survey by ARKENAS (Indonesian National Archaeological Board) is currently under way in some areas of the East Flores regency. The typical Austronesian urn burials that have been discovered can be dated to between 2500 B.P. and 2000 B.P. (500 to 0 BC), which is consistent with the consensus on the arrival of Austronesian settlers in the Lesser Sunda Islands, including Timor (J-C. Galipaud & H. T. Simanjuntak, 2012, pers. comm.).

Human historical genetic studies also provide some insights. Mona et al. (2009) analysed DNA samples from seven islands (Flores, Adonara, Solor, Lembata, Pantar, Alor, and eastern Timor). Their data ‘suggest a complex genetic history in eastern Indonesia, with components that reflect contributions from Austronesian-speaking migrants from East Asia and Non-Austronesian-speaking migrants from Melanesia. In addition, [they] find components that appear to reflect an ancient indigenous eastern Indonesian gene pool that is partly shared with Australia. Thus, genetically, [this area] is a melting pot.’ The dating indicated by the archaeological survey in east Flores is also consistent with recent historical genetic analyses by Pierron et al. (2013): ‘we have attempted to date the admixture between Papuan and Indonsian components. The admixture dates for the islands of east Indonesia are around 100 generations ago.’ The admixture between Papuan and Austronesian languages in the Eastern Nusantara islands had been noticed by linguists as well (e.g. Klamer 2010).

Hence the following tentative scenario for the formation of the Lamaholot dialect chain:

- Proto Flores-Lembata originated by the merger of one or several Austronesian language(s) and one or several unidentified non-Austronesian language(s), probably Papuan (Trans New Guinea phylum). Proto-Flores-Lembata is potentially the result of a creolisation and not a lexical borrowing from indigenous languages, because the lexicon is almost entirely Austronesian, while the syntax has a wealth of non-Austronesian features. In addition, later influences from Papuan or other Eastern Nusantara languages may have occurred.
- Proto Flores-Lembata underwent diversification, either through language differentiation under the influence of various non-Austronesian indigenous languages, or by separation, leading to the linkage observed today (from west to east: Sikka, Lamaholot dialect chain, Kédang).
- The island of Adonara (still a fertile and densely populated area today, especially the eastern region) is potentially the area where Lamaholot originated.
- Lamaholot extended westward (eastern tip of Flores) and eastward (Lembata), mixing with closely related vernacular languages, and subsequently splitting into many dialects. This is why no phonological or lexical innovations are shared by the western and eastern peripheral dialects.
- The Adonara dialects remained more conservative, retaining the fluid semantic alignment (Fluid-S) and the fixed semantic alignment (Split-S). In the peripheral

---

17 Mona et al. (2009) found that ‘The frequency of Melanesian NRY haplogroups ranged from 45.2% in Austronesian speakers from Lembata to 100% in Austronesian speakers from Alor (albeit only four individuals) or 84.4% in Austronesian speakers from Adonara’. The authors make no distinction between the terms Papuan and Melanesian. They also identified a probable ‘signature of the initial aboriginal hunter-gatherer population in the Pleistocene that occupied both [these islands] and New Guinea.’ They believe that the surprising genetic links with Australian Aboriginals ‘reflects the pre-Neolithic gene pool of the area, rather than a signature of more recent migration from/to Australia.’

18 Local traditions or colonial historical sources keep traces of immigrations episodes from Ceram, in the Moluccas (for instance a prominent family of Kota Kaya rulers, the Seran Goran); from Bima (the Koré Bima family, land owners in Kiwang Ona, Adonara), and others.
dialects semantic alignment underwent erosion and simplification, a consequence of languages/dialects being in contact.

- According to Klamer (2010), in approximately ad 1300-1500 some Lamaholot speakers settled to the east on the coast of Alor Island.

7 Conclusion

The Lamaholot dialect chain contains many dialects, potentially up to 33 according to Keraf (1978), extending over the eastern tip of Flores, Adonara, Solor, and Lembata (Lesser Sunda Islands, Indonesia). Dialect diversity is high; speakers of a peripheral dialect may find it challenging to understand Lamaholot speakers from the other side of the chain. These dialects can, however, be split in three clusters, from west to east: Southeast Flores, Tanjung Flores—Adonara—East part of Solor, and Lembata (except the eastern region of Lembata Island where Kédang is spoken). Within the Lamaholot dialect chain, phonological and lexical data consistently show that:

- Adonara is the most conservative dialect; innovations are not shared between the peripheral dialect clusters of Southeast Flores and Lembata.
- If the Adonara cluster presents an innovation, a more conservative phonology or lexicon will be retained and usually shared by Southeast Flores and Lembata.
- Adonara innovations are often common to one of the peripheral dialect clusters, suggesting that it has spread eastwards (to Lembata) or westwards (to Tanjung Flores, Eastern Solor).

This is coherent with the geographical central position of Adonara in this dialect chain. Moreover, morphosyntax proves more complex in Adonara, whose most striking feature is an extensive and complex split intransitivity system, a system which has undergone serious erosion in the peripheral dialects. It is unlikely that the East Adonara dialect alone evolved into a more sophisticated split intransitivity system compared to neighbouring dialects. Indeed, the remnants of split intransitivity systems in the peripheral Lamaholot dialects are very different from one dialect to another. Therefore, they must originate from a morphologically more complex protolanguage.

It can be hypothesised that within a large region whose inhabitants spoke a linkage of languages originating from Proto-Flores-Lembata (that can be classified as Central-Malayo-Polynesian), Lamaholot originated in Adonara, then spread westwards (eastern tip of Flores) and eastwards (Lembata), mixing with closely related vernacular languages and subsequently splitting into many dialects.

Keraf (1978) conducted a comprehensive study of his mother-language, the Lamaléra Lamaholot dialect (south coast of Lembata Island). According to him, there are many more dialects in Lembata, which are highly diversified. A priori, significant local differentiation may signal a very ancient settlement. Therefore, a survey of the Lamaholot dialects on Lembata Island is needed, to validate or refute Adonara as the homeland of the Lamaholot language.

References


4 The voice systems of Wotu, Barang-barang and Wolio: Synchronic and diachronic perspectives

DAVID MEAD AND JOANNA SMITH

This paper investigates three languages belonging to the small Wotu-Wolio subgroup of Sulawesi, Indonesia. These three languages are Wotu, Barang-barang, and Wolio. Following the pioneering work of Donohue (2004) and Van den Berg (2008), which focused on historical sound change in this subgroup, in this paper we take a different tack and look at pronouns and the morphosyntax of voice constructions. Our aim is to answer the following questions: What are the person markers and voice system that can be reconstructed for their common ancestor, Proto Wotu-Wolio? Are there developments which the Wotu-Wolio languages share, which would strengthen the hypothesis that they constitute a valid genetic grouping? What are the developments which led to the distinct, present-day systems?

Following a brief overview of the Wotu-Wolio subgroup, we present thumbnail sketches of the voice systems of Wotu, Barang-barang, and Wolio from a synchronic perspective. From this basis, we then reconstruct the person markers and verbal affixation that constituted the basic voice system of their common ancestor, Proto Wotu-Wolio. We also trace the developments which must have transpired to produce the systems seen today.

1 The Wotu-Wolio subgroup

Five languages belong to the Wotu-Wolio subgroup, or six, if one considers Barang-barang and Laiyolo to be separate languages. These languages are dispersed along the margins of Bone Bay, the body of water separating the southern and southeastern peninsulas of Sulawesi.

- Wotu (500 speakers), spoken at the head of the Gulf of Bone. The archeological record indicates that Wotu emerged in the thirteenth century as an early, important political center on the Gulf of Bone. As a polity under the Bugis kingdom of Luwu', the Wotu held sway over the Kalaena River basin, an important access route linking the bay area with the highlands of central Sulawesi (Schrauwers 1997:363; Bulbeck 2000:7).
- Wolio (>40,000 speakers), spoken in the town of Baubau and nearby villages in southwestern Buton Island. Wolio is well known as the court language of the former sultanate of Buton (mid-sixteenth century to 1960).
• Kamaru (4,000–5,000 speakers), spoken in six villages on the eastern side of Buton Island. Little is known about this language, although initial word-list data suggest that it is closely related to Wolio.
• Laiyolo (fewer than 250 speakers) and Barang-barang (550 speakers), spoken in separate enclaves on the southern part of Selayar Island, off the tip of southern Sulawesi.
• Kalao (500 speakers), also known as Lambego, spoken on the eastern half of Kalao Island to the southeast of Selayar Island.

As stated above, we are looking for shared innovations in the voice systems of the Wotu-Wolio languages. Rather than investigate all five (or six) languages, we concentrate on the three for which we have the best documentation, and which also give us good geographic spread: Wotu (in the north, head of Gulf of Bone), Barang-barang (in the southwest, on Selayar Island), and Wolio (in the southeast, on Buton Island).

2 Wotu

We begin with Wotu. As emerges below, Wotu is more conservative than either Barang-barang or Wolio, which makes it a logical starting point for our discussion. The following presentation of Wotu voice constructions has been summarized from Mead’s (2013) Wotu Grammar Notes. This grammar sketch in turn drew information from three published descriptions (Adriani 1898; Salombe, et al. 1987; Sande, et al. 1991), as well as from an unpublished Wotu sentence list, an unpublished lexicon, and two unpublished texts. The unpublished materials were collected and prepared in the 1990s by SIL researcher Wyn Laidig working with Stefanus Syuaib, a native speaker of Wotu.

2.1 Person markers

Wotu has four sets of person markers. The members of all four sets are listed in Table 1, even though only the last two sets are relevant to voice and grammatical relations. The alternate labels “Set F,” “Set P,” “Set A,” and “Set B” follow a convention established in the comparative studies of Van den Berg (1996) and Mead (2002) as a way of referring to pronoun sets that are etymologically related across languages, without necessarily making a statement about their function in any particular language. Across languages of this study, Set F are independent (“free”) forms. Set P person markers are genitive (“possessive”) clitics or suffixes. Set A are verb prefixes/proclitics which are historically related to the Set P markers, while Set B are verbal suffixes/enclitics (“bound” forms) which are historically related to the independent pronouns. In Wotu, first person plural inclusive forms ‘we’ (ita, -ta, etc.) can also be used in a polite/honorific way to refer to a respected ‘you.’

The third person Set A prefix has three allomorphs, la-, na- and a-, without any clear conditioning environment. Regarding the distribution of the -a and -ya allomorphs of the third person Set B enclitic, see the Appendix. For other allomorphs, see Mead (2013:12 ff).

---

1 Kaufman (2008), although aware of this previous research, uses an opposite convention, namely what are referred to here as Set A pronouns he refers to as Set B, and vice versa.
### Table 1: Wotu person markers.

<table>
<thead>
<tr>
<th>Independent</th>
<th>Genitive</th>
<th>Nominative</th>
<th>Absolutive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Set F</td>
<td>Set P</td>
<td>Set A</td>
<td>Set B</td>
</tr>
<tr>
<td>1SG</td>
<td>iyau, yau</td>
<td>-u/-yu</td>
<td>-u/-yu</td>
</tr>
<tr>
<td>1PL.EXCL</td>
<td>iyami, yami</td>
<td>-mami</td>
<td>i-</td>
</tr>
<tr>
<td>1PL.INCL</td>
<td>ita</td>
<td>-ta</td>
<td>ta-</td>
</tr>
<tr>
<td>2</td>
<td>iyo, yo</td>
<td>-mu</td>
<td>mu-</td>
</tr>
<tr>
<td>3SG</td>
<td>isia</td>
<td>-na/-nna</td>
<td>la-/na/-a-</td>
</tr>
<tr>
<td>3PL</td>
<td>sisia</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### 2.2 Intransitive verbs

In Wotu, both nominative (Set A) and absolutive (Set B) person markers are used to index subjects of intransitive verbs. The following examples illustrate the use of nominative forms.

**WOTU**

(1) *Ijiawi la-awa.*
    yesterday 3.NOM-come
    ‘He came yesterday.’

(2) *Mai=yo ta-ipa i banua-na anri-u!*
    come=2.ABS 1PL.INCL.NOM-go at house-3.GEN younger_sibling-1SG.GEN
    ‘Let’s go to my younger sibling’s house!’

The following examples illustrate the use of absolutive (Set B) forms to index intransitive subjects.

**WOTU**

(3) *Molanga=ba=u.*
    go_by_foot=only=1SG.ABS
    ‘I’m just going by foot.’

(4) *Maturu=ya.*
    sleep=3.ABS
    ‘He’s asleep.’

The factors which condition the use of a nominative versus absolutive marker are not well understood, except that consistently negation triggers the use of a nominative form.

---

² The form of the first person plural exclusive Set B clitic is currently unknown, and present data does not allow us to decide between three possibilities: (a) a form exists, but because of limited data does not occur in our corpus (based on comparative evidence, the expected form is *-*ami); (b) the enclitic -ta actually serves as a generalized first person plural form, unmarked for exclusive or inclusive; or (c) there is no enclitic form, and the independent form iyami/yami is used in contexts where a clitic form might be expected.
WOTU

(5) **Matabba sarro edo la-awa.**
    many very NEG 3.NOM-come
    ‘Very many did not come.’

(6) **Bulli=mo mu-maeka.**
    don’t=COMPL 2.NOM-afraid
    ‘Don’t be afraid.’

Wotu prefixes for deriving intransitive verbs include the reciprocal prefix *si*-(sijagurru ‘fistfight,’ *si*rewo ‘quarrel’) the stative prefix *ma*-(malampu ‘wild,’ *maputi ‘white,’ *masese ‘near’) and the non-agentive prefix *te*-ita-lo- (the first two usually with gemination of the initial stem consonant: *tessalle ‘poured out,’ tattutu ‘closed,’ torango ‘heard, overheard’). See also the discussion regarding the prefix *mo-* in § 2.4.

2.3 Voice constructions

In context, a transitive stem\(^3\) will occur with one of four different prefixes, each one corresponding to a different voice construction. These markers are *manga-* (active), *mu-* (also active), a Set A prefix (inverse),\(^4\) and *i-* (passive).

The two active constructions are similar in that the transitive object, if realized overtly, must immediately follow the verb. The primary difference is that with *manga-* the transitive object is never realized pronominally, but with *mu-* it may be realized pronominally, that is to say, by an independent pronoun or indexed by a Set B enclitic. In the following examples we underline the verb and the object which follows (the verb phrase constituent).

WOTU

(7) **Manga-paccingi bilassa pabilassa.**
    ACT-clean garden farmer
    ‘The farmer cleaned the garden.’

(8) **Anri-nna Laiya mu-ala doi=e.**
    younger_sibling-3SG.GEN Laiya ACT-take money=that
    ‘Laiya’s younger sibling took that money.’

(9) **Yo mu-ita vau.**
    2.INDEP ACT-see 1SG.INDEP
    ‘You see me.’

(10) **Yau mu-oti=o.**
    1SG.INDEP ACT-call=2.ABS
    ‘I called you.’

\(^3\) As with the other languages considered in this study, we limit our discussion to basic transitive stems, and omit the complications which would be introduced by a consideration of applicative suffixes.

\(^4\) Inverse voice is the only construction which allows both the subject and object to be indexed on a transitive verb. However, since the object is clearly pivot, Mead (2013) prefers the term ‘inverse.’ Whether it is an inverse in a pragmatic sense as defined e.g. by Givón (1994) has yet to be investigated.
Although it is possible for a verb with *manga* to be followed by a Set B enclitic, in this case the clitic does not index the transitive object (as above with *mu*, see example 10), but rather the transitive subject.

**WOTU**

(11)  *Mang-arra=ta anrale.*  
      ACT-boil=1PL.INCL.ABS corn  
      ‘We boil corn.’

In both actor voice constructions, the transitive subject is the pivot. This can be seen, for example, in that *manga-* and *mu-* forms are used when a transitive subject is relativized (examples 12 and 13) or interrogated (example 14).

**WOTU**

(12)  *puwa-u iyya manga-lua=e pongka pappua oge tomai*  
      grandparent-1SG.GEN REL ACT-fell=that trunk wood big hither  
      ‘my grandfather who felled that big tree’

(13)  *ito iyya mu-patei=yye sapi-mu*  
      person REL ACT-kill=that cow-2.GEN  
      ‘the person who killed your cow’

(14)  *Sema mu-antara ama-mu?*  
      who ACT-accompany father-2.GEN  
      ‘Who accompanied your father?’

In the inverse construction, the verb is prefixed with a nominative (Set A) prefix that indexes the transitive subject, and is optionally followed by an absolutive (Set B) enclitic that indexes the object. This is the only construction in which both the subject and the object can be indexed on the transitive verb stem.

**WOTU**

(15)  *Ajia punti-u. Bulli mu-anre=a!*  
      that banana-1S.GEN don’t 2.NOM-eat=3.ABS  
      ‘Those are my bananas. Don't eat them!’

(16)  *Tulu ta-soro=a lemba eni.*  
      help 1PL.INCL.NOM-push=3.ABS boat this  
      ‘Help us push this boat.’

Imperatives which omit indexing of an understood you are included as representatives of this construction.

**WOTU**

(17)  *Patei=ya!*  
      kill-3.ABS  
      ‘Kill him!’

In most cases, however, object indexing on the verb (Set B) lapses when the object is present elsewhere in the clause, whether before or after the verb.
In the inverse construction, the transitive object is the pivot. Note the use of inverse forms when the patient is relativized as in (20) or interrogated as in (21).

In the fourth construction, the transitive verb is preceded by the passive marker i-. Basic passive verbs have two core arguments, a derived subject (the underlying patient), and a demoted agent. In most corpus examples of passive clauses, the demoted agent is omitted.

In fact it is possible to consider i- verbs to be syntactically intransitive. When a nominative (Set A) or absolutive (Set B) form occurs with the verb it indexes the derived subject.

Based on comparative evidence, we might expect the demoted agent to be indexed on i- verbs using a genitive (Set P) person marker, but this pattern has not been observed (in fact we have no corpus examples of the demoted agent realized pronominally). In the rare cases when a demoted agent is expressed, two patterns have been observed. In one text example, it follows the verb, introduced by the compound preposition awa i ‘from’:
WOTU
(25)  ...i-carita awa i ina-na puwa-u tomai motae...
    PASS-tell from at mother-3.GEN grandparent-1SG.GEN hither say
    ‘...it was told by (lit. from) the mother of my grandfather that …’

Salombe et al. (1987:83) provide three examples in which the demoted agent is
expressed following the verb without oblique marking. For example:

WOTU
(26)  Marota=e i-papaccingi kaka-u.
    dirty=that PASS-clean older_sibling-1SG.GEN
    ‘That trash was cleaned up by my older brother.’

2.4 The prefix mo- (po-)

In addition to the affixes discussed above, Wotu also has a prefix mo-. With some stems
the mo- derivation is clearly intransitive, for example mobulu ‘hairy,’ mongorro ‘snore,’
omata ‘wake up from sleep,’ and molaa ‘fly.’ In other cases, however, the stem to which
mo- attaches appears to be transitive. In fact in our corpus we have encountered no less
than twenty stems which can take either mo- or manga-, with the expectation that further
research will reveal others. For example:

WOTU
(27)  mobuau mangabuau ‘make’
    mosaringi mangasaringi ‘sweep’
    motannu mangatannu ‘weave’ (cloth)

Partly on this basis, Salombe et al. (1987:82) regard both manga- and mo- to be markers
of the active voice. Our own preliminary analysis, however, is that mo- forms are
syntactically intransitive, including cases as in examples (29) and (30) where an object has
been incorporated.

WOTU
(28)  Ya=massa-nna pura mo-sipullu.
    REL=all-3.GEN finish INTR-gather
    ‘All of them are already gathered.’

(29)  Yau awa mo-manu tulu dia yo.
    1SG.INDEP come INTR-request help to 2.INDEP
    ‘I have come to request help of you.’

Nominative and absolutive markers that occur with a mo- verb index the (intransitive)
subject, never the object. When preceded by a nominative marker, as in example (31),
the prefix assumes the form po-.

WOTU
(30)  Mo-paru=ya banua.
    INTR-build=3.ABS house
    ‘He builds a house, he is house-building.’
(31)  
\[ U\text{-}po\text{-}tana\text{=}me \quad \text{dia} \quad \text{ranga}\text{-}u \quad \text{Tomai} \quad \text{mo\text{-}tae}, \quad \text{“Aga...} \]

\[ \text{I}5\text{G.}\text{NOM-INTR}\text{-}ask\text{=}\text{COMPL to friend-1}\text{G.}\text{GEN} \quad \text{Hither INTR}\text{-}say \quad \text{what} \]

\[ \text{‘I asked my friend, saying: What ...’} \]

2.5  Summary

Table 2 summarizes the voice constructions of Wotu. In general we find that Wotu morphosyntax supports the notion of transitivity as a continuum (Hopper and Thompson 1980). In fact in the presentation chosen for Table 2, transitivity is highest in the middle rows, and lowest in the rows at the top and bottom.

**Table 2: Verbal inflection potential in Wotu**

<table>
<thead>
<tr>
<th>Type</th>
<th>Inflection</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intransitive (including O-incorporation)</td>
<td>( mo\text{-}+ V )</td>
<td>S argument may be indexed by Set A or Set B; incorporated O (if present) must follow the verb</td>
</tr>
<tr>
<td>Active(_1)</td>
<td>( manga\text{-}+ V )</td>
<td>O argument must follow the verb, cannot be realized pronominally; A is pivot</td>
</tr>
<tr>
<td>Active(_2)</td>
<td>( mu\text{-}+ V + (Set B) )</td>
<td>O argument must follow the verb, may be indexed by Set B enclitic; A is pivot</td>
</tr>
<tr>
<td>Inverse</td>
<td>( Set A + V + (Set B) )</td>
<td>Set A indexes A argument; O argument may be realized anywhere in the clause, including Set B; O is pivot</td>
</tr>
<tr>
<td>Passive</td>
<td>( i\text{-}+ V )</td>
<td>derived S argument may be indexed by Set A or Set B; derived S is pivot</td>
</tr>
<tr>
<td>Non-agentive</td>
<td>( te\text{-}lita\text{-}ito\text{-}+ V )</td>
<td>S argument may be indexed by Set A or Set B; derived S is pivot</td>
</tr>
<tr>
<td>Reciprocal</td>
<td>( si\text{-}+ V )</td>
<td>Set A indexes A argument; O argument may be realized anywhere in the clause, including Set B; O is pivot</td>
</tr>
<tr>
<td>Stative</td>
<td>( ma\text{-}+ V )</td>
<td>S is pivot</td>
</tr>
</tbody>
</table>

For example, passive verbs are transitive-like in that the demoted agent may (in rare cases, as most Wotu passives are agentless) be expressed as a noun phrase following the verb, either with or without oblique marking (our Wotu corpus provides no examples of a demoted agent which is realized pronominally when the verb is marked with \( i\)-). On the other hand, they are intransitive-like in that the derived subject (\( S_d \)) can be indexed by a Set A or a Set B person marker, just as with other intransitives.

Similarly it has been suggested to us that \( mo\)- derivations would be better labeled ‘semi-transitive.’ While some \( mo\)- derivations are transitive-like in allowing the expression of a second participant—namely an incorporated-O (e.g. mopatu surra ‘send letters, letter-send, correspond’)—this is possible only with a subset of \( mo\)- verbs, and in all other respects they are treated as intransitive. Therefore we feel the label ‘intransitive’ is justified.\(^5\)

---

\(^5\) Compare Hopper and Thompson who write “antipassive and O-incorporation constructions serve exactly the same purpose: they code clauses which are very low in Transitivity as ‘intransitive’—even ones which may, strictly speaking, have two participants. That is, morphosyntactic markings tend to be
3 Barang-barang

Our presentation of the Barang-barang voice system follows Belding, Laidig & Maingak (2001) and Smith (2002:45 ff.). Both of these studies were based on primary data collected in the field, consisting of several texts and a large electronic lexical database provided by a single native speaker. This material was supplemented by more limited data from other native speakers. We follow the same general order as we did with Wotu, beginning with person markers, followed by intransitive verbs, and finally voice constructions.

3.1 Person markers

Barang-barang has four sets of person markers, shown in Table 3. First person plural inclusive forms ‘we’ (kita, -ka, etc.) can also be used in a polite/honorific way to refer to a respected ‘you.’ Although paradigmatically sia, -na and -ea/-a occupy the third person singular row, they can also have plural readings and thus function more like unmarked forms. By contrast, sianai, -na'i and -'ia are explicitly plural forms.

Table 3: Barang-barang person markers.

<table>
<thead>
<tr>
<th>Independent</th>
<th>Possessive</th>
<th>Subject</th>
<th>Object</th>
</tr>
</thead>
<tbody>
<tr>
<td>Set F</td>
<td>Set P</td>
<td>Set A</td>
<td>Set B</td>
</tr>
<tr>
<td>1SG</td>
<td>aku</td>
<td>-ku</td>
<td>ku-</td>
</tr>
<tr>
<td>1PL.EXCL</td>
<td>kami</td>
<td>-mami</td>
<td>pi-</td>
</tr>
<tr>
<td>1PL.INCL</td>
<td>kita</td>
<td>-ka</td>
<td>la-</td>
</tr>
<tr>
<td>2SG</td>
<td>ko'o</td>
<td>-mu/-u</td>
<td>mu-</td>
</tr>
<tr>
<td>2PL</td>
<td>ko'omiu</td>
<td>-mu/-miu</td>
<td>-ko'miu/-komiu</td>
</tr>
<tr>
<td>3(SG)</td>
<td>sia</td>
<td>-na</td>
<td>la-</td>
</tr>
<tr>
<td>3PL</td>
<td>sianai</td>
<td>-na'i</td>
<td>-'ia</td>
</tr>
</tbody>
</table>

3.2 Intransitive verbs

Intransitive stems are optionally indexed for their subject (S) with a Set A prefix, as in the following examples.

BARANG-BARANG
(32) Ne'e mu-lafong.  
don’t 2SG.SBJ-reply  
‘Don’t reply’

(33) Ku-mæininging.  
1SG.SBJ-cold  
‘I’m cold.’

It is also possible for the subject to be realized by a full NP in the clause, in which case the prefix may occur, as in example (34), but it may also be omitted, as in example (35). When a full noun phrase is present, the normal order is verb-subject.
BARANG-BARANG

3. SBJ-INTR-bark dog that
‘The dog is barking.’

35. SBJ-INTR-open door that
‘The door opened.’

The existential verb *nyia* is exceptional in that it is indexed for its subject with a Set B suffix, that is to say, the same set used to index transitive objects (see below), rather than a Set A prefix.

BARANG-BARANG

36. SBJ-INTR-continue exist-3SG.OBJ name Loa that

ri-pau lara ri kampong.
GOALFOC-word inside at village
‘But the name Loa still exists, spoken inside the village.’

Barang-barang prefixes for deriving intransitive verbs include the accidental prefix *tə-* (*təlua* ‘vomit,’ *təiobo* ‘capsized,’ *tərungkua* ‘punished’); the stative prefix *mə-* (*mərea* ‘sick,’ *məkeddi* ‘small,’ *mədinging* ‘cold’) which is sometimes reduced to *m-* (as in *moge* ‘big,’ *mekə* ‘afraid’); the intransitive prefix *pə-* (*pəsəpatu* ‘wear shoes,’ *pəbisara* ‘speak,’ *pənasu* ‘cook’); and the reciprocal prefix *pəsi-* (*pəsilonga* ‘meet,’ *pəsigaga* ‘argue,’ *pəsítəllengaka* ‘sink at the same time’). See further Belding, Laidig & Maingak (2001:29–34).

3.3 Goal focus constructions

In context, a transitive verb base will make its appearance in one of three voice constructions, which Smith (2002) designates as actor focus (verbal maker: *ma*-), goal focus₁ (verbal marker: Set A prefix), and goal focus₂ (verbal marker *ri*). According to their affixation potential, these voice constructions are distinguished as follows.

In goal focus₁, the transitive object is indexed by an object suffix. The primary difference between the two constructions is that in goal focus₂ the verb is prefixed by *ri*-. In goal focus₂ there is no potential for the transitive subject to be indexed on the verb—although as illustrated in example (39) it may appear as an oblique. Although the transitive subject is demoted, the patient is not promoted to subject position—it is not indexed with a Set A prefix (as

BARANG-BARANG

37. SBJ-INDEP SBJ-eat-3SG.OBJ banana-2SG.POSS
1SG.IND 1SG. SBJ- eat-3SG.OBJ banana-2SG.POSS
‘I ate your bananas.’

Goal focus₂ is illustrated in examples (38) and (39). As in goal focus₁, the transitive object is indexed by an object suffix. The primary difference between the two constructions is that in goal focus₂ the verb is prefixed by *ri*-. In goal focus₂ there is no potential for the transitive subject to be indexed on the verb—although as illustrated in example (39) it may appear as an oblique. Although the transitive subject is demoted, the patient is not promoted to subject position—it is not indexed with a Set A prefix (as
happens with subjects in canonical intransitives), but remains indexed with a Set B pronoun.

BARANG-BARANG

(38) *Loka toria ri-kanre-a.*
   banana that GOALFOC-eat-3SG.OBJ
   ‘That banana was eaten.’

(39) *Ri-tdle’esti-ko ri ana-mu.*
   GOALFOC-urinate_on-2SG.OBJ at child-2SG.POSS
   ‘You were urinated on by your child.’

In both goal focus constructions, there are conditions when indexing of the transitive object lapses or becomes optional.

BARANG-BARANG

(40) *Ne’e sia muni mu-pətamo manga pitina.*
   don’t 3SG.INDEP again 2SG.SBJ-make_heavy with slander
   ‘Don’t weigh him down with slander.’

(41) *Ri-duru-mo dua ganru tria.*
   GF-harvest-PFV also corn that
   ‘That corn is also harvested.’

The transitive object is the pivot in both goal focus constructions, compare for example when the object is relativized. In Barang-barang, relative clauses are usually marked by the relativizer anu, by a deictic element such as *toria* or *itu* placed in second position, or—as in these examples—by both.

BARANG-BARANG

(42) *loka anu ku-lamung toria ri taung ri kana*
   banana REL 1SG.SBJ-plant that at year at before
   ‘the bananas that I planted last year’

(43) *berəng anu ri-pake=tu be=ma-bunu ri sapi-ku*
   machete REL GOALFOC-use=that FUT=ACTFOC-kill at cow-1SG.POSS
   ‘the machete that will be used to kill my cow’

3.4 Actor focus constructions and the prefix ma-

In actor focus the verb is prefixed with *ma*- . In simple or matrix clauses there is always a noun phrase or independent pronoun that instantiates the transitive subject, usually placed preceding the verb. A verb marked with *ma*- is never followed by a Set B person marker—although it can be immediately followed by an object noun or noun phrase.

BARANG-BARANG

(44) *Anri-mu ma-kanre duriang pongane.*
   young sibling-2SG.POSS ACTFOC-eat durian earlier
   ‘Your younger sibling was eating durian earlier.’
(45) Sia ma-balú’ kopi mai ri aku.
3SG.INDEP ACTFOC-sell coffee here at 1SG.INDEP
‘She sells coffee to me.’

As in (44) and (45), actor focus is appropriate when the patient is not specific or individuated. Compare also example (46), in which the transitive subject is relativized:

BARANG-BARANG
(46) muane məlitau ma-səppe tria ganru
man young ACTFOC-pick that corn
‘the young men who pick corn’

In other cases of subject relativization, however, sometimes two possibilities emerge. Example (47) is licensed by the fact that the subject is the pivot, and here it is the subject that is relativized, while (48)—exceptionally, a goal focus verb is used even though the subject is relativized—is apparently licensed by the fact that the object is specific.

BARANG-BARANG
(47) ito ma-bunu tria sapi-mu (actor focus)
person ACTFOC-kill that cow-2SG.POSS
‘the person who killed your cow’

(48) ito la-bunu tria sapi-mu (goal focus)
person 3.SBJ-kill that cow-2SG.POSS
‘the person who killed your cow’

Another interesting facet of relativization not specifically mentioned by Smith (2002) is that even intransitive verbs are sometimes marked with ma- when their subjects are relativized.

BARANG-BARANG
(49) poali ganru ma-pəsua ri kampong Loa’
buyer corn ACTFOC-enter at village Loa
‘corn buyers who come into Loa village’

(50) soridau-na Serəng ma-matte ri Loa’
soldier-3SG.POSS Seram ACTFOC-die at Loa
‘soldiers of Seram who died at Loa’

Despite the gloss given to ma- in (49) and (50), it is hard to maintain that here it is still an actor focus marker. Indeed, beyond relativization there are other contexts—always when the verb is subordinated and the subject is controlled—in which ma- occurs with both transitive and intransitive stems. In the following pairs, the (a) examples illustrate ma- prefixed to a transitive stem, while the (b) examples illustrate a parallel case where ma- is prefixed to an intransitive stem.
BARANG-BARANG

(51) a. Sianai muni ri-tofongi ma-duru ganru.
3PL.INDEP also GOALFOC-invite ACTFOC-harvest corn
‘They are also called to harvest corn.’

b. Lile mu-tofongi-m-ea ma-bolili inrong-u.
tomorrow 2.SBJ-invite-PFV-3SG.OBJ ACTFOC-return mother-2SG.POSS
‘Tomorrow you must ask your mother to come home.’

(52) a. ...sia be=la-lengka ma-pup-pulu’ kopi.
3SG.INDEP FUT=3.SBJ-walk ACTFOC-DUR-pick coffee
‘…she will go pick coffee.’

b. ...ka=ta-pəpəruru ma-lengka.
and=1PL.INCL.SBJ-depart ACTFOC-walk
‘…then let’s be on our way.’

(53) a. La-longa-m-ea Punro’ ma-sissili kutu-na
3.SUB-see-PFV-3SG.OBJ Monkey ACTFOC-scratch louse-3SG.POSS

Jangang Ko’o ri karo-na.
Chicken Forest at body-3SG.POSS
‘Monkey saw Jungle Chicken scratching lice on his body.’

b. Ku-long-ea /ku+longa+ea/ ma-olo mai ri si’i’.  
1SG.SUB-see-3SG.OBJ ACTFOC-approach hither at this
‘I see him coming this way.’

We believe that the use of ma- with transitive stems (as seen in examples 46, 47, 51a, 52a and 53a) was prior and—as emerges below—its spread to intransitive stems (examples 49, 50, 51b, 52b and 53b) ‘anticipates’ the participle function that is fully instantiated by its cognate form in Wolio.

3.5 Summary

The voice patterns of Barang-barang are summarized in Table 4.

<table>
<thead>
<tr>
<th>Table 4: Verbal inflection potential in Barang-barang</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transitive</td>
</tr>
<tr>
<td>Actor Focus</td>
</tr>
<tr>
<td>Goal Focus1</td>
</tr>
<tr>
<td>Set A + V + (Set B)</td>
</tr>
<tr>
<td>Goal Focus2</td>
</tr>
<tr>
<td>Intransitive</td>
</tr>
<tr>
<td>(Set A) + V</td>
</tr>
<tr>
<td>Existential</td>
</tr>
</tbody>
</table>

4 Wolio

Wolio is the only Wotu-Wolio language with a published grammar (Anceaux 1988) and dictionary (Anceaux 1987). The following description is based on Anceaux’s grammar, supplemented with material from Alberth (2000).
4.1 Person markers

Wolio has four sets of person markers, shown in Table 5 (Anceaux 1988:25, 27, 33–34, 36, 42; Alberth 2000:30). Independent, ‘Possessive,’ ‘Actor’ and ‘Object’ are the labels given to them by Anceaux.

<table>
<thead>
<tr>
<th>Set</th>
<th>Independent</th>
<th>Possessive</th>
<th>Actor</th>
<th>Object</th>
</tr>
</thead>
<tbody>
<tr>
<td>1SG</td>
<td>iaku</td>
<td>-ku</td>
<td>ku-</td>
<td>-aku</td>
</tr>
<tr>
<td>1PL.EXCL</td>
<td>(ing)kami</td>
<td>-mami</td>
<td>ta-</td>
<td>-kami</td>
</tr>
<tr>
<td>1PL.INCL</td>
<td>(ing)kita</td>
<td>-ta</td>
<td>-</td>
<td>-kita</td>
</tr>
<tr>
<td>2SG</td>
<td>(ing)koo</td>
<td>-mu</td>
<td>u-</td>
<td>-ko</td>
</tr>
<tr>
<td>2PL</td>
<td>(ing)komiu</td>
<td>-miu</td>
<td>-</td>
<td>-komiu</td>
</tr>
<tr>
<td>3</td>
<td>incia</td>
<td>-na</td>
<td>a-</td>
<td>-a/-ia/-ea</td>
</tr>
</tbody>
</table>

Third person forms may be made explicitly plural by using manga, i.e. manga incia ‘they,’ -na manga ‘their,’ or in the case of the Set A person marker, a- followed by manga at the end of the verb phrase (see example 55) (Anceaux 1988:36, 37, 43).

4.2 Finite verbs

According to Anceaux, a Wolio verb is either a participle (used as a noun or noun modifier) or it is finite. Finite verbs are obligatorily cross-referenced for subject (whether transitive or intransitive, or the derived subject of a passive verb) with a Set A person marker, regardless of whether the subject appears elsewhere in the clause. A main clause must contain a finite verb (Anceaux 1988:42), of which the following are examples.

**WOLIO**

(54) Inca-na humai uka a-sapo-mo o eo. (intransitive)
    inside-3.POSS that also 3.ACT-set-COMPL ART sun
    ‘Meanwhile the sun had also set.’

(55) A-peelo-mo ikane manga. (transitive)
    3.ACT-seek-COMPL fish PL
    ‘They looked for fish’

Finite passives are marked with to-, and are always agentless (Anceaux 1988:28).

**WOLIO**

(56) Inda a-to-taurako. (passive)
    NEG 3.ACT-PASS-put_down
    ‘They are not laid down.’

The object of a transitive verb may be realized as a full noun as in example (55) above, with a Set B person marker as in (57), or zero as in the second clause of (58). In a few cases indexing and noun phrase realization are found combined, compare example (59).

---

6 Anceaux does not include the first person plural inclusive form -kita in his list of Wolio object markers. However it is confirmed by both Alberth (2000:30) and McDowell (2012).
WOLIO

(57)  A-ala-m-ea. /a+ala+mo+ea/  
3.ACT-take-COMPL-3.OBJ
‘They took it.’

(58)  Maka sa-opa-po La Ndokendoke a-peelo ikane,  
but one-what-INCOMPL CLF.M Monkey 3.ACT-seek fish
inda a-pokawaaka.  
NEG 3.ACT-encounter
‘But however intently Monkey looked for fish, he could not find [any fish].’

(59)  O Wa Turungkoleo a-kemba-ia i saripi-na.  
ART CLF.M Turungkoleo 3.ACT-summon-3.OBJ at side-3.POSS
‘She called Wa Turungkoleo to her side.’

4.3 Participles

Participles on the other hand are exclusively used as substantives or as attributes of nouns. Active participles are marked with an invariant mo- form. It is usually\(^7\) paired with the third person possessive form -na following the verb (Anceaux 1988:25).

WOLIO

(60)  manga mo-pe-ɓuani-na  
PL PTCP-INTR-castnet-3.POSS
‘fishermen’ (ones using castnets)

(61)  O ndoke-mo duka mo-ɗobui-na  
ART monkey-COMPL also PTCP-extract-3.POSS
bulu-na pani-na.  
feather-3.POSS wing-3.POSS
‘The monkey was also the one who pulled out the feathers of his wings.’

However -na does not occur when a transitive participle is followed instead by a Set B (object) marker as in (62).

WOLIO

(62)  o manga mia mo-lawati-a i kabalingko humai  
ART PL person PTCP-receive-3.OBJ at grave that
‘the people who receive him in the grave’

Passive participles are of two types. Agentless passive participles are marked by to- and, as above, wrapped with mo- -na.

---

\(^7\) A possessive pronoun does not follow the participle in fixed phrases, be it active or passive, for example kapala mo-polaka (ship PTCP-fly) ‘airplane’ (lit. flying ship) and tabako i-tange (tobacco PTCP.PASS-wrap) ‘wrapped tobacco’ (Anceaux 1988:25, 28).
WOLIO
(63) **mo-to-paisilamu-na** (passive to-)
    PTCP-PASS-circumcise-3.POSS
    ‘[ones] being circumcised’

(64) **mo-to-kemba-na** (passive to-)
    PTCP-PASS-invite-3.POSS
    ‘the ones who are invited’

Passive *i-* participles on the other hand are never preceded by *mo-. Furthermore the demoted agent may be indexed by a possessive (Set P) marker.

WOLIO
(65) **saha** *i-tobe-mu** _iti_ (passive *i-*
    chili_pepper PTCP.PASS-pick-2SG.POSS that
    ‘those chili peppers that you picked’

(66) **i-aso-na** _manga_ **mia** (passive *i-*
    PTCP.PASS-sell-3.POSS PL person
    ‘what is sold by the people’

(67) **a-po-gau-aka** _opea-mo_ **i-karajaa** (passive *i-*
    3.ACT-INTR-talk-APPL what-COMPL PTCP.PASS-work
    ‘they talked about what was to be done’

In addition to affixes mentioned above, Wolio prefixes used to derive intransitive stems include, among others, the stative prefix *ma-* (**maluntu** ‘lazy,’ **maali** ‘expensive,’ **maeka** ‘afraid’); and the prefixes *po-* and *pe-* which “occur in complimentary distribution and in many cases have similar meaning” (Anceaux 1988:18) e.g. **potawa** ‘to laugh,’ **pomeke** ‘to cough,’ **pewilu** ‘to smile,’ **pojao** ‘to paddle,’ **peikane** ‘to fish.’ The prefix *po-* is the usual reciprocal marker (**pokeni** ‘hold each other,’ **poewangi** ‘to fight each other’). See further Anceaux (1988:14 ff.) and Alberth (2000:85).

### 4.4 Summary

Table 6 summarizes the voice patterns of Wolio.

<table>
<thead>
<tr>
<th>Voice</th>
<th>Finite</th>
<th>Participle</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Intransitive</strong></td>
<td>Set A + <em>ma-V</em></td>
<td><em>mo-ma-V</em> (-na)</td>
</tr>
<tr>
<td></td>
<td>Set A + <em>pe-V</em></td>
<td><em>mo-pe-V</em> (-na)</td>
</tr>
<tr>
<td></td>
<td>Set A + <em>po-V</em></td>
<td><em>mo-po-V</em> (-na)</td>
</tr>
<tr>
<td></td>
<td>etc.</td>
<td>etc.</td>
</tr>
<tr>
<td><strong>Transitive</strong></td>
<td>Set A + V + (Set B)</td>
<td><em>mo-V</em> + (Set B/-na)</td>
</tr>
<tr>
<td><strong>Passive</strong></td>
<td>Set A + <em>to-V</em></td>
<td><em>i-V</em> + (Set P)</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>mo-to-V</em> (-na)</td>
</tr>
</tbody>
</table>
Focus systems of Wotu, Barang-barang and Wolio

5 Proto Wotu-Wolio and diachronic developments

Based on the above descriptions of Wotu, Barang-barang and Wolio, we now reconstruct the person markers and the voice system of their common ancestor, Proto Wotu-Wolio (PWW), and trace the innovations which gave rise to the present-day languages. We begin by covering each person marker set in turn, then verbal affixes, and finally the role of Set A markers in the voice system. Where it sheds light on a matter, we also bring in data from other languages that were not specifically discussed above, such as Laiyolo, Kalao, Kamaru, and Selayar.

5.1 PWW person markers

Wotu, Barang-barang and Wolio all have four sets of person markers, and we also reconstruct four sets for their common ancestor. Most PWW forms are unremarkable in that they continue forms which have been reconstructed for Proto Malayo-Polynesian (PMP) or for their more immediate Proto Celebic (PCel) ancestor.

<table>
<thead>
<tr>
<th>Set F</th>
<th>Set P</th>
<th>Set A</th>
<th>Set B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1S</td>
<td>*i-aku</td>
<td>*-ku</td>
<td>*-aku</td>
</tr>
<tr>
<td>1PN</td>
<td>*[ŋ]-kita</td>
<td>*ta-</td>
<td>*-kita</td>
</tr>
<tr>
<td>1PX</td>
<td>*[ŋ]-kami</td>
<td>*-mami</td>
<td>*-kami</td>
</tr>
<tr>
<td>2S</td>
<td>*[ŋ]-ko</td>
<td>*-mu</td>
<td>*-ko</td>
</tr>
<tr>
<td>2P</td>
<td>*[ŋ]-komiu</td>
<td>*-miu</td>
<td>*-komiu</td>
</tr>
<tr>
<td>3</td>
<td>*[ŋ]-sia</td>
<td>*-na</td>
<td>*-a/ia</td>
</tr>
</tbody>
</table>

In the Set A series, all Wotu-Wolio languages exhibit loss of the original second person plural form (cf. PCel *mi-). Although Table 7 attributes this loss to the proto-language, it could reflect areal influence from South Sulawesi languages since Bugis, Makassar, Konjo and Selayar all exhibit a similar loss. Notably Wotu is the only daughter language to have extended this trend, losing second person plural forms across all sets.

Similarly, third person plural forms were lost across all person marker sets (but in a few cases were renewed through the innovation of new third person plural forms). This loss also shows all the hallmarks of being an areal feature. In addition to the Wotu-Wolio languages currently under discussion, loss of original third person plural forms is characteristic of all South Sulawesi languages with the exception of the Tamanic languages of Borneo (Mills 1975:210 ff.; Kaufman 2008:14 ff.) as well as a number of Muna-Buton languages.

---

8 Reconstructions are also based on a consideration of person markers in Laiyolo, Kalao and Kamaru (not shown).

9 For Proto-Austronesian (PAN) and PMP person markers, see among others Blust (1977, 2009) and Ross (2002:36). Reconstructed PCel person markers have not been published.

10 Consider for example the independent pronouns. It is clear that Barang-barang, Kalao, Laiyolo, Wotu and Kamaru renewed the singular/plural distinction via the innovation of new (and various) third person plural independent pronouns, respectively sianai, sina'i, salelea, sisia and ngaisia. The clearest in etymology is Laiyolo salelea, which straightforwardly contains the stem lele ‘all.’ Barang-barang sianai and Kalao sina'i appear to trace back to sia followed by the third person plural genitive suffix -na'i. The Wotu form sisia is formed by reduplication; furthermore it is similar to third person plural independent pronouns found dialectally in Pamona, including Lage sisi'a (plural) (next to s'i,a, singular or plural) and Onda'e sisira (plural, respectful) (next to sira, singular or plural, respectful) (Adriani 1931:340). Kamaru ngaisia arose via clipping of earlier manga isia, in which manga is the plural marker found in both Kamaru and Wolio (cf. Wolio which has retained manga incia ‘they’).
languages with the exception of Muna (Van den Berg 2003:99–100). Conversely, north of these areas third person singular and plural forms have been retained in all Bungku-Tolaki languages (Mead 1998:121 ff.) and Kaili-Pamona languages except for Rampi (loss of third person plural free form) (Kaufman 2008:15, footnote 7) and Pamona (in process of losing the distinction across all person marker sets) (Adriani 1931:340–343). Although Table 7 suggests that the loss of the singular/plural distinction in third person had already been effected in PWW, one could also entertain the notion that the loss occurred more gradually, e.g. only partially effected in PWW, but eventually reaching its completion in all daughter languages.

One change that is significant for subgrouping is the development of conditioned allomorphs of the third person Set B marker, namely *-ia following consonants and the vowel /a/, and *-a following all other vowels (< PCel *ia). As originally noted by Noorduyn (1991:144–145), a similar allomorphy is unknown in any of the surrounding languages, therefore this innovation supports the notion of Wotu-Wolio as a valid genetic group. For how this original allomorphy has developed in the daughter languages, see the Appendix.

For the third person Set A marker, we reconstruct only one form, PWW *a-. As this form is also unique to the Wotu-Wolio languages, it must be considered a second innovation which supports the notion of a Wotu-Wolio subgroup. However, this reconstruction requires an explanation because not only do daughter languages exhibit multiple forms (na-, la- and a-), but some of these appear to continue the PCel third person forms respectively *na- (singular) and *ra- (plural).

The starting place for our discussion is Table 8, which gives third person forms in all Wotu-Wolio languages and selected surrounding languages (other language data has been taken from Kaufman 2008:54–55, Adriani 1931:341–342, Martens 1988:169, and Mead 1998:131). Forms are unmarked for plurality except as noted otherwise.

<table>
<thead>
<tr>
<th>Language</th>
<th>Form of third person marker</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sa’dan Toraja</td>
<td>na-</td>
</tr>
<tr>
<td>Bugis</td>
<td>na-</td>
</tr>
<tr>
<td>Makasar</td>
<td>na-</td>
</tr>
<tr>
<td>Konjo</td>
<td>na-</td>
</tr>
<tr>
<td>Selayar</td>
<td>la-</td>
</tr>
<tr>
<td>Barang-barang</td>
<td>la-</td>
</tr>
<tr>
<td>Laiyolo</td>
<td>a- (usual), la- (rare, two corpus examples following negator)</td>
</tr>
<tr>
<td>Kalao</td>
<td>la-</td>
</tr>
<tr>
<td>Wotu</td>
<td>la- ~ na- ~ a-</td>
</tr>
<tr>
<td>Wolio</td>
<td>a-</td>
</tr>
<tr>
<td>Kamaru</td>
<td>a-</td>
</tr>
<tr>
<td>Tukang Besi</td>
<td>na-, a- (irrealis), no-, o- (realis)</td>
</tr>
<tr>
<td>Lasalimu</td>
<td>na- (irrealis), no- (realis)</td>
</tr>
<tr>
<td>Muna</td>
<td>na- (sg, irr), no- (sg, realis), da- (pl, irr), do- (pl, realis)</td>
</tr>
<tr>
<td>Tolaki</td>
<td>no- (singular), ro- (plural)</td>
</tr>
<tr>
<td>Pamona</td>
<td>na- (singular and plural), ra-, nda- (generic they, people)</td>
</tr>
<tr>
<td>Uma</td>
<td>na- (singular), ra- (plural)</td>
</tr>
</tbody>
</table>

From this table, note that most surrounding languages have na- (Muna, Tolaki, Pamona and Uma are included in Table 8 as examples of languages which have retained reflexes of both PCel *na- and *ra-). Although Wotu also has na-, which appears to continue PCel...
*na-, we ignore it for purposes of reconstructing PWW on the assumption that *na- in Wotu is the result of more recent influence from Bugis. Of the other modern-day forms *la- and *a-, we reconstruct *a- for the protolanguage, because it has the widest distribution in Wotu-Wolio; furthermore the loss of the initial consonant appears to be an innovation in Wotu-Wolio, since the form *a- is not found in any surrounding languages except Tukang Besi.

As we have highlighted in Table 8, a form *la- is found not only in certain Wotu-Wolio languages, but it is also found in Selayar, a feature which distinguishes Selayar from all other South Sulawesi languages. Given the distribution of languages which exhibit *la-, it seems clear that Selayar Island must have been the geographic locus of this innovation. We presume that the innovation occurred in Selayar and spread to Barang-barang, Laiyolo and Kalao. To summarize, we propose the following historical scenario as the simplest explanation for the distribution of present-day forms.

- Third person plural *ra- is lost in South Sulawesi languages, with *na- becoming a generalized third person form unmarked for plurality (Kaufman 2008).
- The PCel distinction between *ra- ‘3PL’ and *na- ‘3GS’ is lost in PWW, with the form *a- becoming a generalized third person form unmarked for plurality. Presumably the form *a- (with irregular loss of the initial consonant) reflects analogy with the similar Set B marker. Note also the irregular loss of the initial nasal in PWW passive marker *i- from prior *ni- (see § 5.2).
- PSS *na- irregularly shifts to *la- in Selayar.
- Selayar *la- spreads to Barang-barang, Kalao, and Laiyolo, as well as Wotu, replacing or nearly replacing inherited *a-.
- Third person *na- spreads from Bugis to Wotu, where it takes its place as a minor variant next to *la- and *a-.

Two objections can be raised against this proposal. First, while we know that Barang-barang, Laiyolo and Kalao have been influenced by Selayar, this was not true of Wotu. The spread of *la- to Wotu could not have been directly from Selayar, but (presumably) must have been via languages of the Kalao subgroup. Second, from a phonetic viewpoint it would be more reasonable to assume that *la- reflects the third person plural form *ra-.

Making this assumption suggests three alternative scenarios. For the benefit of our readers, we list these scenarios, but endorse none of them, since each is problematic in one way or another.

Scenario 1: a singular/plural distinction in third person proclitic forms was maintained into Proto South Sulawesi. Difficulties: one would have to assume that all languages lost the plural form (*ra-), except for Selayar which lost *na- (and reflects *ra- as *la-).

Scenario 2: a singular/plural distinction was maintained into PWW, with *ra- reflected in present-day languages as *la-, and *na- as *a- (or possibly as *na- in Wotu). Difficulties: Wotu and Laiyolo retained reflexes of both forms, but lost the singular/plural distinction anyway; direction of borrowing of *la- must have been from Wotu-Wolio into Selayar.

Scenario 3: when the singular/plural distinction collapsed prior to PWW, *na- was lost while *ra- was retained; by the stage of PWW the latter had developed doublets *la- and *a-. Difficulties: everywhere else (South Sulawesi languages, other Wotu-Wolio person

---

* The reverse possibility could also be entertained, namely that the innovation occurred in the Kalao subgroup and spread to Selayar. This would require us to assume that the status of Laiyolo etc. viz-à-viz Selayar was at one time reversed from what it is in the present-day. Furthermore, although *na- > *la- is irregular, nonetheless a shift of *n > l seems more likely that the creation of an l ex nihilo, as would be required by *a- > *la-.
marker sets) singular forms were retained while the plural forms were lost; direction of borrowing of *la- must have been from Wotu-Wolio into Selayar.

5.2 Passive prefix *i-

For the passive marker we reconstruct PWW *i-. Presumably this form developed via irregular consonant loss from *ni-, a form which itself traces back to the PMP infix *-in-. Present-day languages exhibit two forms of the passive marker. In Wotu and Wolio, the passive marker has the form i-. In Barang-barang, Laiyolo and Kalao, the passive marker has the form ri-, but we consider this form to be borrowed from neighboring South Sulawesi languages, since a passive marker ri- is also found in Selayar, Konjo and Bugis. Interestingly, all Wotu-Wolio languages (with the possible exception of Kamaru, for which we have no data) exhibit a homophony between the form of their passive marker and their general locative preposition. This curious state of affairs, which is found in a number of Sulawesi languages, was first noted by Van den Berg (2004:546). Expanding his chart, we can report the following languages (Konjo, Makasar and Tolaki are included to illustrate the typical case, namely languages which do not exhibit this homophony).

Table 9: Forms of the passive marker and locative preposition in selected languages.

<table>
<thead>
<tr>
<th>Language</th>
<th>Passive marker</th>
<th>Locative preposition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indonesian</td>
<td>di-</td>
<td>di</td>
</tr>
<tr>
<td>Bugis</td>
<td>ri-</td>
<td>ri</td>
</tr>
<tr>
<td>Selayar</td>
<td>ri-</td>
<td>ri</td>
</tr>
<tr>
<td>Barang-barang</td>
<td>ri-</td>
<td>ri</td>
</tr>
<tr>
<td>Kalao</td>
<td>ri-</td>
<td>ri</td>
</tr>
<tr>
<td>Laiyolo</td>
<td>ri-</td>
<td>ri</td>
</tr>
<tr>
<td>Wotu</td>
<td>i-</td>
<td>i</td>
</tr>
<tr>
<td>Wolio</td>
<td>i-</td>
<td>i</td>
</tr>
<tr>
<td>Muna</td>
<td>ne-</td>
<td>ne</td>
</tr>
<tr>
<td>Konjo</td>
<td>ni-</td>
<td>ri</td>
</tr>
<tr>
<td>Makasar</td>
<td>ni-</td>
<td>ri, i</td>
</tr>
<tr>
<td>Tolaki</td>
<td>-in-</td>
<td>i</td>
</tr>
</tbody>
</table>

Based on this chart, we suggest that not only was the form of the passive marker in the Kalao subgroup (ri-) influenced by South Sulawesi languages, but so too was the form of the preposition (ri).

In PWW, passive forms must have taken genitive indexing for the demoted agent, just as passive verbs still do in present-day Wolio, cf. i-kande-ta (PTCP.PASS-eat-1PL.INCL.POSS) ‘eaten by us,’ i-kamata-mu (PTCP.PASS-see-2SG.POSS) ‘seen by you’ (Alberth 2000:71, Anceaux 1988:28). The use of genitive markers to index the agent in non-agent voice constructions is an old pattern that traces back to Proto Austronesian. It is somewhat noteworthy, then, that between Wotu, Wolio and Barang-barang, only in Wolio has this function survived—and even then it is found only in so-called participle forms (for example relative clauses), and does not occur in main clauses.

12 An infixed form of the passive marker survives in certain substantive constructions. We cite the following examples, but since they do not relate to the voice system we have not investigated such forms systematically: Wotu minanga ‘river’ and inusa ‘spittle’ (both cited by Adriani 1898:137); Barang-barang konanre, Layolo kinanre, Kalao and Wolio kinande ‘food,’ literally ‘that which is eaten.’
In Wotu by contrast, passive forms have come to be treated as ordinary intransitive verbs, and for the most part are agentless. When the demoted agent is expressed, it is not indexed on the verb but rather follows it, with or without oblique marking. As with other intransitives, verbs marked with i- can be indexed for their intransitive subject—viz. the underlying patient—with either a Set A or a Set B person marker.

In Barang-barang, only Set B person markers are used. Given that apart from the existential verb nyia, intransitive subjects are indexed with Set A forms exclusively, this gives rise to an odd situation in which it appears that—even though the verb has passive morphology—the patient appears not to have been promoted to subject position. For this reason, Smith (2002:52) was unwilling to call stems marked with ri- ‘passive verbs’ and instead concluded that they were best considered a second goal focus construction. Interestingly, the same situation is found in Selayar: “The ri- passive is not a typical one, because it seems only to demote the agent and does not promote the patient, in comparison with a canonical passive” (Jackson 2008:18). Example (68) illustrates the non-canonical or ‘defective’ passive of Selayar (example from Ceria 1993, cited in Jackson 2008:18).

SELAYAR
(68) Ri-úru’-i bayken-na i-Sitti ( ri i-Dénji’ ).
PASS-massage-3SG.ABS leg-3SG.POSS CLF-Sitti by CLF-Denji
‘Sitti’s leg got massaged (by Denji).’

The Selayar construction is entirely parallel with what we find in Barang-barang; compare example (68) with example (39) above.

A third possible analysis would be to consider Barang-barang to have a split-S system: ri-verbs are neither goal focus or passive, but rather are intransitive, and they along with nyia belong to the class of intransitive verbs which index their subjects like an object (S_o), while all other intransitive verbs index their subjects like the subjects of transitive verbs (S_a). We do not pursue here which is the correct analysis for present-day Barang-barang, but we suggest the third possibility as at least a plausible historical pathway by which it developed.

5.3 Non-agentive prefix *to-

A non-agentive prefix *to- is reconstructed for PWW on the basis of Wotu to-, Barang-barang and Kalao ta-, and Wolio to-. Wotu has alternate forms of this prefix, te- and ta-, following which the initial consonant of the stem is doubled. We regard this pattern to reflect influence or borrowing from Bugis.

Verbs marked with *to- were syntactically intransitive. As regards the voice system, this prefix is notable in only one respect: in Wolio to- took the place of i- as the de facto marker of passive voice in main clauses, when the latter became restricted to subordinate clauses.

5.4 Stative prefix *ma-

A stative prefix *ma- is reconstructed for PWW on the basis of Barang-barang ma- and Laiyolo, Kalao, Wotu and Wolio ma-. Stems marked with *ma- were syntactically intransitive.
5.5 Prefixes *mu-, *maŋ- and *mo- and the PWW voice system

We consider these three prefixes together, and suggest that PWW must have had a voice system similar to that found in Wotu. Namely, in PWW there existed a contrast between *mu-, used with specific, definite objects to mark the stem as subject-pivot, and which contrasted with stems inflected with a Set A person marker, which were object-pivot; *maŋ- (possibly with an epenthetic vowel *a added before consonant-initial stems), used with indefinite or non-specific objects, at any rate it disallowed the object to be realized pronominally; and *mo-, which derived intransitive verbs from various bases, including (at least some) transitive stems. All three prefixes marked the verb as subject pivot, and required any realization of the patient to immediately follow the verb.

It is also to be noted that *mo-, and presumably also *maŋ-, had the alternate forms *po- and *paŋ- when they were immediately preceded by a Set A person marker. In all, PWW had six voice categories, as we identify in Table 5.

Table 10: The voice system of Proto Wotu-Wolio

<table>
<thead>
<tr>
<th>Intransitive / O-incorporation</th>
<th>*mo-V</th>
<th>*mo-V + Set B</th>
<th>Set A + *po-V</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antipassive</td>
<td>*maŋ-V</td>
<td>*maŋ-V + Set B</td>
<td>Set A + *paŋ-V</td>
</tr>
<tr>
<td>Active-direct</td>
<td>*mu-V + Set B [=O]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Passive</td>
<td>*i-V + Set P [=A]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-agentive</td>
<td>*to-V</td>
<td>*to-V + Set B</td>
<td>Set A + *to-V</td>
</tr>
</tbody>
</table>

For the active-direct, inverse, and passive, the function of the Set A, Set B and Set P person marker sets were fixed in regard to which grammatical role they indexed, whether the transitive subject (=A) or the transitive object (=O). The subjects of other verbs could, just as in present-day Wotu, be indexed fluidly with either a Set A or a Set B marker.

5.6 From Proto Wotu-Wolio to present-day languages

The system shown in Table 10 is, in its essentials, the same system found in Wotu. However, in Wotu passive verbs are no longer indexed for their demoted agent with a Set P (genitive/possessive) marker, but have come to be treated as other intransitive verbs that allow their (derived) subjects to be indexed fluidly. Note also that reflexes corresponding to the presumed cell “Set A + *paŋ-V” have not been attested in our Wotu corpus.

From PWW, two important changes took place leading to the development of the systems found in Barang-barang and Wolio. The first was the collapse of the contrast between *mo-, *maŋ- and *mu-, leaving only *mo-. Second was the shift, on intransitive verbs, from fluid-subject indexing to a nominative-accusative system, that is to say, the growing use of Set A markers at the expense of Set B markers. Today, the older, more widespread use of Set B person markers is vestigially retained in Barang-barang in their use with the existential verb nyia (§ 3.2). Curiously, even in the Wolio existential particle daangia, now regarded as monomorphemic, the final vowel sequence derives historically from PWW third person *-ia.

13 In this paper we skirt the issue of how this system developed, since this would require bringing in more comparative evidence from across Sulawesi than space allows. For readers interested in the broader comparative picture, see Van den Berg (1996), Mead (2002), and Ross (2002) among others.
These two changes in themselves result in the system shown in Table 11, which not only strongly resembles the Barang-barang system but is also a precursor to the Wolio system.

Table 11: Development of the PWW voice system, intermediate stage

<table>
<thead>
<tr>
<th>Active</th>
<th>*mo-V : Set A + *po-V</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inverse</td>
<td></td>
</tr>
<tr>
<td>Passive</td>
<td>*i-V + Set P</td>
</tr>
<tr>
<td>Non-agentive</td>
<td>*to-V : Set A + *to-V</td>
</tr>
</tbody>
</table>

Not shown in Table 11 is the functional divergence of *mo- and *po-, which must have also begun to occur. The prefix *po- (→ Barang-barang pə-) retained its function of deriving detransitive and other intransitive verbs, while *mo- expanded in its function as a marker of subordinate verbs. This expansion is seen in its initial stages in Barang-barang, where ma-\(^{14}\) serves not only as a marker of actor focus, but also in certain contexts it attaches to subordinate intransitive verbs; see § 3.4, especially examples (49) through (53). This expansion progressed even further in Wolio, where by analogy mo- became the de facto participle marker for nearly all intransitive verb forms, including intransitive verbs with po-, stative verbs with ma-, and non-agentive verbs with to-.

In the final stage in Wolio, verbs inflected for subject (i.e. with a Set A marker) came to predominate in main clauses, in effect ‘kicking out’ verbs marked with mo- and i- from this position. With loss of mo- forms from main clauses, the original inverse construction (Set A + V + Set B) became the sole ‘transitive’ form. With the loss of i- forms from main clauses, an original non-agentive constructions (Set A + to-V) became the de facto passive form—but to this day still do not allow the expression of an agent.

6 Summary

By tracing developments in the voice systems of three languages of Sulawesi we hope to have taken another step in the progressively deeper understanding of voice phenomena in Austronesia. As home to a number of languages which are transitional between those with Philippine-type ‘focus’ marking and those with Indonesian-type voice systems, Sulawesi once again proves an interesting laboratory for understanding changes which can occur in voice systems.

This study also strengthens the case for considering Wotu-Wolio to be a valid genetic subgroup. To the set of sound changes discussed by Donohue (2004) and Van den Berg (2008) and the lexical innovations proposed by Sirk (1988), we can now add the following innovations: (a) development of conditioned allomorphs *a and *ia in the third person enclitic (Set B) person marker (< PCel *ia), see § 5.1 and the Appendix; (b) loss of initial nasal in the third person proclitic (Set A) person marker *a- (< PCel *na-), see § 5.1; and (c) loss of initial the nasal in the passive marker *i- (< *ni- < PCel *in-), see § 5.2.

---

\(^{14}\) Presumably from earlier *mo-, with the change in vowel quality occurring after the stative prefix *ma- had shifted to /mə-/.
Appendix

In § 5.1, we claimed that the third person enclitic marker had two allomorphs in PWW, namely *-ia, following consonants and the vowel *a, and *-a, following all other vowels, that is, after *i, *e, *o and *u. This can be demonstrated from a consideration of the forms which this person marker has in Wotu, Barang-barang and Wolio. However, as this original allomorphy has developed in different directions in the daughter languages, we treat each context individually.

A.1 Following the vowel /a/

In Wotu the third person enclitic always has the (orthographic) form -ya when the final vowel of the stem is /a/. Said another way, the high vowel i becomes non-syllabic between two other vowels. For example:

**WOTU**

(69)  
<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>bawa ‘bring’</td>
<td>bawaya ‘bring it’ (&lt; bawa-ia)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td>tarima ‘receive’</td>
<td>tarimaya ‘receive it’</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.</td>
<td>pajia ‘store’</td>
<td>pajiaya ‘store it’</td>
<td></td>
<td></td>
</tr>
<tr>
<td>d.</td>
<td>lopa ‘slice’</td>
<td>lopaya ‘slice it’</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Barang-barang exhibits the same pattern, except that the historical sequence /a+i/ merged as /e/.

**BARANG-BARANG**

(70)  
<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>bafa ‘carry’</td>
<td>bafea ‘carry it’ (&lt; bafa-ia)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td>longa ‘see’</td>
<td>longea ‘see it’</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.</td>
<td>sakka ‘capture’</td>
<td>sakkea ‘capture it’</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In Wolio, two patterns emerge (Anceaux 1988:27). If the final vowel of the stem is /a/, and the penultimate vowel is not /a/, then the suffix has the form /-ia/. Also if the penultimate and final vowels are both /a/, and are not separated by a consonant, then the suffix also has the form /-ia/:

**WOLIO**

(71)  
<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>papesua ‘make enter’</td>
<td>papesuaia ‘make him enter’</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td>parahatika ‘notice’</td>
<td>parahatikaia ‘notice it’</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.</td>
<td>sikisa ‘punish’</td>
<td>sikisaiaia ‘punish her’</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Otherwise, if the penultimate and final vowels are both /a/, and are separated by a consonant, then the final /a/ is replaced by /ea/ (as in Barang-barang).

**WOLIO**

(72)  
<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>ala ‘take’</td>
<td>alea ‘take it’ (&lt; ala-ia)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td>kamata ‘see’</td>
<td>kamatea ‘see it’</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.</td>
<td>pada ‘finish’</td>
<td>padea ‘finish it’</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
A.2 Following consonants

The only final consonants in Barang-barang are glottal stop and the velar nasal. Following glottal stop, the third person clitic has the form /-kea/. Following the nasal, it has the form /-ea/ (Belding, Laidig & Maingak 2001:14):

**BARANG-BARANG**

(73)  a. beso’ ‘throw’ beso’kea ‘throw it’
     b. bale’ ‘return’ bale’kea ‘return it’
     c. lamung ‘plant’ lamungea ‘plant it’
     d. pəts’bunge ‘drop’ pəts’bungea ‘drop it’

Presumably the original form was /ia/, but became /ea/ by analogy with the form it has following /a/ (cf. example 69). Wotu and Wolio have lost final consonants, and therefore are not germane to this context. Serendipitously the original form of the person marker has been preserved in Wolio daangia, the existential particle ‘there is, there are,’ historically daang (or dang) plus third person -ia, cf. Laiyolo dangia, Kalao dangea.

A.3 Following vowels other than /a/

Following the vowels i, e, o, and u, the third person clitic has the form -a in both Barang-barang and Wolio (Anceaux 1988:27).

**BARANG-BARANG**

(74)  a. keni ‘hold’ kenia ‘hold it’
     b. kanre ‘eat’ kanrea ‘eat it’
     c. rango ‘hear’ rangoa ‘hear it’
     d. bunu ‘kill’ bunua ‘kill it’

**WOLIO**

(75)  a. gaangi ‘leave’ gaangia ‘leave it’
     b. kande ‘eat’ kandea ‘eat it’
     c. potumpo ‘cut’ potumpoa ‘cut it’
     d. matau ‘know’ mataua ‘know it’

In Wotu, it has the form -ya following stem-final -i preceded by another vowel. The same phonemic principle is at work here as in example (69), except in this case the high vowel belongs to the stem rather than the clitic:

**WOTU**

(76)  a. patei ‘kill’ pateiya ‘kill him’
     b. doi ‘give’ doiya ‘give him’
     c. sullei ‘exchange’ sulleiya ‘exchange it’

When i is preceded by a consonant, or when the clitic follows other vowels, both -a and -ya have been attested, sometimes even on the same stem.
WOTU

(77) a. san ‘know’ sania, saniya ‘know it’
    b. anre ‘eat’ anrea, anreya ‘eat it’
    c. paru ‘build’ paruya ‘build it’
    d. tulu ‘help’ tulua ‘help him’

In example (77), presumably sania, anreya and tulua represent the original situation, but the form -ya is in process of spreading to other contexts, leading to newer formations such as saniya, anreya and paruya.

Following aspectual clitics mo and po

The aspectual particles mo ‘perfective’ and po ‘imperfective’ present a special case. In both Wolio and Barang-barang, when these are followed the third person clitic, they appear on the surface as mea and pea (not moa and poa, as might be expected by application of rules given above).

WOLIO

(78) a. apajeremea /a+pajere+mo+ea/ ‘he followed him’
    b. beapapakipea /be+a+papaki+po+ea/ ‘it will bite him soon’

BARANG-BARANG

(79) a. kulongamea /ku+longa+mo+ea/ ‘I saw it’
    b. lakellaipea /la+kellai+po+ea/ ‘they still wanted it’

In Wotu, the usual forms of the aspectual clitics are -me and -pe, with -mo and -po occurring as minor variants (Mead 2013:37-39). Although it is possible that in origin the difference between these variants was coalescence with a third person marker (-me, -pe) or not (-mo, -po), such a distinction is apparently not felt today, compare example (80).

WOTU

(80) Poso=me=ga=o?
    tired=COMPL=Q=2.ABS
    ‘Are you tired already?’

The distribution of variant forms of the aspectual clitics in the current language is not well understood, and requires further investigation.

References


Adriani, N. 1898. Iets over de talen der To Sada en der To Wadu. Mededeelingen van wege het Nederlandsch Zendelinggenootschap 41. 111–150.


5 The Etymology of the sociative-progressive circumfix in Suwawa (Gorontalo-Mongondowic)

YUKO KITADA

1 Introduction*

The present paper deals with the etymology of the sociative\(^1\)-progressive circumfix \textit{gi-a} in Suwawa. I argue that this circumfix is not a sporadic areal innovation, but consists of two reflexes of the Proto-Malayo-Polynesian affixes \textit{*maR-si-} ‘actor voice.irrealis-sociative\(^2\)’, and \textit{*-an} ‘event plural’, which I newly reconstruct based on evidence from Suwawa and other western Austronesian languages mainly in the Philippines and Sulawesi.

The Suwawa language is spoken in some inland districts in the Gorontalo Province in North Sulawesi, Indonesia. It is one of the Gorontalic languages of the Gorontalo-Mongondowic branch. The Gorontalic subgroup also includes Gorontalese, Buol etc. and the Mongondowic subgroup includes Mongondow and Ponosakan (Lewis et al. 2013). The Gorontalo-Mongondowic subgroup occupies an important place in Austronesian historical and comparative linguistics and in the history of Austronesian migration, because it is spoken on the borderline area between the two linguistic typological areas generally referred to as Philippine-type and Indonesian-type (Himmelmann 2005a). It is also the only subgroup in Sulawesi classified into the Greater Central Philippines subgroup placed in the Philippine group, which is one of the Malayo-Polynesian subfamily of the Austronesian family (Blust 1991). Figure 1 illustrates the classification:

\(*\) I would like to express my gratitude to Nikolaus Himmelmann, Sonja Gipper and Malcolm Ross for their valuable feedback and generous support. The usual disclaimer applies.
\(^1\) Following Lichtenberk’s definition, I use the term sociative to refer to a situation where two or more participants are involved together in the same actor (his ‘Initiator-type’) role with a (semantically) non-singular subject: typically, but not necessarily, the involvement is simultaneous (Lichtenberk 2000:35-36). Note that other terminologies for sociative include associative, collective, cooperative (Nedjalkov 2007:33).
\(^2\) This refers to the \textit{m}-initial series of the tripartite mood system with the initial \textit{m}, \textit{n} or \textit{p/Ø} in actor voice formatives, which is widely attested in western Austronesian languages especially in the Philippines. Several different labels have been used with reference to the \textit{m-} and \textit{n}-initial series in the system, however, I label the \textit{m}-initial series as irrealis, and the \textit{n}-initial series as realis (cf. Himmelmann 2005a:168).
There are a few publications on the Suwawa language, such as Kasim et al. (1981), Pateda et al. (1985) and Tome et al. (1988), however, no study discusses the sociative-progressive construction in Suwawa. The present study is based completely on fieldwork data collected by the author.

Recent discussions on western Austronesian morphosyntax primarily focus on voice and mood paradigms and person markers (see Himmelmann 2002, 2005a, Lémarchal 2010, Reid 2002, Ross 2002, 2006, 2009 et passim), however, relatively less attention has been paid to the sociative and related derivational categories involving event plurality such as reciprocals and iteratives. The reconstructions I propose for the etymology of the Suwawa sociative-progressive circumfix *gi-a would provide evidence that some western Austronesian affixes are homomorphemic but distinct in function and origin, from a different perspective than that found in the voice and mood paradigms.

The structure of the present study is as follows: In Section 2, the sociative-progressive construction in Suwawa is introduced with some examples. In Section 3, I discuss the etymology of the Suwawa sociative circumfix *gi-a, using data from related constructions in Suwawa and other western Austronesian languages. In Section 4, I conclude that the Suwawa sociative-progressive circumfix *gi-a diachronically consists of two parts; the prefixal *gi- indicating sociative and the suffixal -*a indicating progressive. I argue that the prefixal *gi- is a reflex of the Proto-Malayo-Polynesian sociative prefix *maR-si-, while the suffixal -*a reflects the Proto-Malayo-Polynesian suffix *-an indicating event plurality.

---

3 See Liao (2011a) for PMP *maki-/*paki- ‘comitative’ and Zeitoun (2002) for Formosan *pa-Ca- ~ *ma-Ca- ‘dynamic reciprocal’ and *paR- ~ *maR- ‘stative reciprocal’; note that Lichtenberk (2000) and Bril (2005) suggest Proto-Oceanic *paRi- ‘plurality of relations (Lichtenberk 2000)’, which also has reciprocal and sociative meanings. Liao (2011b) deals with PMP *si- which is of central concern here. See further fn. 9.
2 The sociative-progressive in Suwawa

In Suwawa, the sociative-progressive construction is formed with the circumfix *gi- -a*, as shown in the examples below:

(1) Tea do *gilaqoa* ado sikola.
    tea do gi-*laqo-a* ado sikola
    3.PL already SOC-go-EP to school
    ‘They are going to school together.’

(2) Monganaqo bitu *gipohogilanga* no bali.
    mongo-wanaqo bitu gi-*pohogilang-a* no bali
    PL-child DIST SOC-play-EP GEN ball
    ‘Those children are playing ball together.’

In Examples (1) and (2), tea ‘3.PL’ and monganaqo bitu ‘those children’ are plural actors encoded as a subject. The sociative-progressive circumfix *gi- -a* forms the verb *gilaqoa* ‘are going together’ with the base *laqo* ‘to go’ in (3) and the verb *gipohogilanga* ‘are playing together’ with the base *pohogilang* ‘to play’ in (4).

A remarkable point about the Suwawa sociative-progressive is, as the hyphenated term shows, that the semantics of the two categories, sociative and progressive, co-occur. The sociative-progressive construction lacks corresponding simple constructions, namely, a ‘simple sociative’ construction, or a ‘simple progressive’ construction marked with part of the sociative-progressive circumfix.

Another point to note is that this construction does not take voice and mood inflections, which are normally obligatory for verbal bases in Suwawa. The sociative-progressive construction has the actor voice, present tense and progressive aspect reading; this suggests that the circumfix had actor voice marking, such as *

In Section 3, I propose an etymological explanation for the Suwawa sociative-progressive marker with two new reconstructed proto-forms based on comparative evidence.

3 Etymological evidence

Etymologically, I propose that the Suwawa sociative-progressive marker can be divided into two parts: the prefixal *gi-* and the suffixal *-a*. Comparative evidence shows that the prefixal *gi-* is the sociative marker ‘proper’, while the suffixal *-a* indicates progressive, which I suggest is a realization of event plurality. The two parts of the circumfix are

---

4 It has a free-variant allomorph *di- -a*, which appears mostly when the base-initial consonant is alveolar.
5 The letter *q* represents a glottal stop.
6 Note that Suwawa has no obligatory number marking on nouns. Thus, nouns can be interpreted as singular or plural when unmarked.
7 The same is true in other Gorontalic languages. See 3.1.
8 The Suwawa language has four voices (the actor, patient, locative voice and conveyance voices) and three moods (the irrealis, realis and imperative moods).
discussed separately: Section 3.1. deals with the prefixal *gi-* and Section 3.2 with the suffixal -*a*.

### 3.1 The prefixal *gi-*

Comparative data on western Austronesian languages allows the reconstruction of Proto-Malayo-Polynesian *maR-si-* ‘actor voice.irrealis-sociative’, and suggests that the Suwawa prefix *gi-* reflects this proto-form. The proto-form *maR-si-* consists of two parts: Proto-Malayo-Polynesian *maR-* ‘actor voice.irrealis’, which has been reconstructed by many authors, such as Blust (2003), Lemarèchal (2010) and Ross (2002), and Proto-Malayo-Polynesian *si-* ‘sociative’, which has also been proposed by Liao (2011b), but with a different meaning. Its reflexes with the sociative semantics are, for example, *magsi-* in Tagalog, *agsi-* in Ilocano and *mosi-* in Pendau, as seen in the examples below:

(3) **Tagalog** magsi-
magsi-kanta ‘to sing (pl.)’ < kanta ‘to sing’

(Schachter and Otanes 1972:335)

(4) **Ilocano** agsi-
agsi-tulong ‘to help out with many people’ < tulong ‘to help’

(Rubino 1997:200)

(5) **Pendau** mosi- (realis nosi-)
nosi-inom ‘drank from the same glass [by plural actors, YK]’ < inom ‘to drink’

(Quick 2003:297)

In other Gorontalo-Mongondowic languages, affixes reflecting *maR-si-* with plural actors are also attested: In the Mongondowic subgroup, Mongondow has the prefix *mosi-* indicating that a plural entity is engaging in the activity or state indicated by the root (Dunnebier 1929:552):

---

9 Liao (2011b) reconstructs the proto-form *si-* based on its reflexes mainly in Philippine languages and proposes that it denotes ‘simultaneous and concurrent aspect’. As this proposal is only available in the form of a conference presentation handout, it is not possible to discuss the different meanings in detail. Note that the ‘sociative’ analysis proposed here predicts that forms thus marked do not have to refer to actions carried out by several actors simultaneously and concurrently. See fn. 23 for an example.

10 Note that Adelaar also suggests that the ‘(West) Malayo-Polynesian *si-* forming reciprocal (and medial?) verbs’ (Adelaar 1992:395), based on his Old Malay, Modern Malay, Salako, Minangkabau and Toba Batak data. However, he makes no arguments on its sociative function. I suppose that his reconstruction and my reconstruction refer to the identical marker, though the meanings are distinct; this is because the languages that he and I examine are different. To connect the two reconstructions with different semantics requires further discussion, and is beyond the scope of the present paper. For a more detailed discussion on the Proto-Malayo-Polynesian prefix *si-*, see Kitada (forthcoming).
The etymology of the sociative-progressive circumfix in Suwawa
(GorontaloMongondowic)

Mongondow: mosi- ‘plural entity’; mosilituq ‘to sit (by plural actors)’ < lituq ‘to sit’
(6) Kinabayaaqan-mai mosi-lituq-makow kong ganderia (dangkulon)
arrived-VEN SOC-sit-AND at gallery (front veranda)

imosia.11
3.PL
‘When (I) arrived, they were all sitting at the front veranda.’
(Dunnebier 1929:552, My glossing.)

In Mongondow, the prefix mosi-, which is a reflex of *maR-si-, indicates the sociative
situation by plural actors; this is the same construction as in Tagalog, Ilocano and Pendau.
However, the languages in the Gorontalic subgroup have a different construction; the
reflex of *maR-si- appears together with a reflex of *-an indicating progressive (see in 3.2)
and forms the sociative-progressive circumfix. For example, Gorontalese has the circumfix
hi- -a marking the durative aspect with a plural subject (Little n.d.:245) and Buol has a
circumfix gi- -an, which makes a ‘special plural form for progressives’ (Zobel 2005:645).12

Gorontalese: hi- -a ‘durative aspect with a plural subject [or actor, YK]’(Little n.d.:245):
hi-deqop-a13 ‘[some people, some animals] are/were catching’ < deqopo ‘catch’ (ibid.). No
sentential examples have been found with the word hi-deqop-a: here I list another
sentential example with the word hi-laqo-laqo-a ‘are going around’ taken from Joest
(1883:12):
(7) monga u hi-laqo-laqo-a.14
moN-ka u hi-RDP-laqo-a
AV.IRR-eat REL SOC-??.go-EP
‘to eat any animals going around.’
( Joest 1883:12, My glossing)

Buol: gi- -an ‘special plural form for progressives’: gi-dolo-an ‘to bring (by plural actors)’
< dolo ‘to bring’
(8) Agu oluo tilo diti malavung gidoloa(n) manuko...
if exist 3.PL.NOM small PV.ST:many PRG.PL:bring chicken
‘If there are many boys bringing roosters with them’
(Zobel 2005:645)

As exemplified here, a reflex of *maR-si- appears with a reflex of the suffix *-an in
Gorontalese, Buol and Suwawa, which are grouped as Gorontalic; in the Gorontalese and
Buol languages, the circumfix also indicates the sociative-progressive meaning as in the
Suwawa language. It is safe to posit that this is a common feature in the Gorontalic
subgroup.

11 The Dutch-style spelling adopted by Dunnebier is modified to a modern transcription in Indonesian
languages in this paper; thus, eo → u, j → y. A glottal stop was originally represented by the Arabic letter
hamzah ⟨⟩, but it is replaced with q in my orthography.
12 Note that the Proto-Austronesian phonemes *R is reflected as /g/ in Suwawa and Buol, and as /h/ in
Gorontalese: thus Gorontalese has hi- and Buol has gi- as the reflexes of *maR-si-, respectively:
13 The letter q represents a glottal stop /ʔ/.
14 Gorontalese hi- -a may appear with reduplication. Note that the original spelling by Joest is modified
based on the transcription in Little (n.d.).
With this morphosyntactic and semantic information, I reconstruct the Proto-Gorontalic form *gi- -an that marks the sociative-progressive meaning.\textsuperscript{15} The Mongondow prefix mosi- suggests that the prefixal part gi- of the Proto-Gorontalic circumfix *gi- -an is a reflex of Proto-Gorontalo-Mongondowic *mog-si-. This suggests that the present tense indicated by Suwawa gi- -a traces back to Proto-Gorontalo-Mongondowic *mog- (and Proto-Malayo-Polynesian *maR-) that indicates irrealis (or non-past).\textsuperscript{16}

I now turn to discussing the historical sound change from Proto-Malayo-Polynesian *maR-si- to Suwawa gi-, which is illustrated in (9). Note that I use the term Pre-Proto-Gorontal to refer to an intermediate stage between Proto-Gorontalic and Proto-Gorontalo-Mongondowic:\textsuperscript{17}

(9) The sound change of Proto-Malayo-Polynesian *maR-si- into Suwawa gi-

<table>
<thead>
<tr>
<th>Proto-Malayo-Polynesian</th>
<th>*maR-si-</th>
<th>'actor voice.irrealis-sociative'</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proto-Gorontalo-Mongondowic</td>
<td>*mog-si-</td>
<td>'actor voice.irrealis-sociative'</td>
</tr>
<tr>
<td>Pre-Proto-Gorontalic</td>
<td>*mogi-</td>
<td>'actor voice.irrealis-sociative'</td>
</tr>
<tr>
<td>Proto-Gorontalic</td>
<td>*gi-</td>
<td>'actor voice.irrealis-sociative'</td>
</tr>
<tr>
<td>Suwawa</td>
<td>gi-</td>
<td>'sociative'</td>
</tr>
</tbody>
</table>

First, Proto-Malayo-Polynesian *maR-si- becomes Proto-Gorontalo-Mongondowic *mog-si-, due to the phonological changes that in the prefix *maR- (1) Proto-Malayo-Polynesian /a/ is reflected as /o/, possibly through schwa at an earlier stage (cf. Sneddon and Usup 1986:410 and Blust 1991:128), and that (2) Proto-Austronesian *R is merged with Proto-Austronesian *g and reflected as *g in Proto-Greater Central Philippines, which is an earlier stage than Proto-Gorontalo-Mongondowic (Blust 1991).\textsuperscript{18}

Secondly, the phoneme /s/ in the Proto-Gorontalo-Mongondowic *mog-si- is lost. Admittedly, no parallel phonological change can be aduced to support the cluster simplification *gs > /g/. However, other alternatives for deriving gi- are even less plausible, since (a) there is no prefix **gi- with the sociative meaning attested anywhere in western Austronesian languages, and (b) the derivational prefix PMP *ma-ki- ‘comitative; requestive’ (cf. Liao 2011a), which has similar semantics with PMP *si- ‘sociative’ and whose reflexes are commonly observed as well, is also less likely because it is reflected as moqi-/moʔi/-l ‘requestive’ in Suwawa. Furthermore, it is implausible to think that the PMP *ma-ki- splits into two Suwawa reflexes, i.e. moqi- and gi- (< earlier **mogi-) because there is no evidence that the proto-phoneme *k is reflected /g/ in Suwawa. Comparatively, the split of the PMP *ma-ki- as having the distinct functions i.e. requestive and sociative with the phonological change of ki- is not observed in any languages as far as I know.

\textsuperscript{15} For other Gorontalo-Mongondowic languages, few data are available. There is no description of the cognate form of the Suwawa gi- -a in Attinggola (Gorontalic) (Kasim et al. 1983): it is not clear if the form is absent or if the description is missing.

\textsuperscript{16} Note that the fact that the Gorontalese data have both past and non-past translations in English by Little (n.d.), and that Mongondow mosi- is inflected for voice and mood (as in Tagalog, Ilocano and Pendau) suggests that the proto-form *gi- is actually the reflex of two different prefixes; Proto-Gorontalo-Mongondowic *mog-si- ‘actor voice.irrealis-sociative’ and *mog-si- ‘actor voice.realis-sociative’. However, for simplicity, the forms with the initial m- is employed in the present paper.

\textsuperscript{17} See Section 1 for the genealogical relations.

\textsuperscript{18} The Suwawa reflex of PMP *R is g as in the following examples (the PMP forms are from Blust (n.d.)): PMP *Ratus ‘hundred’ > Suwawa gatuto ‘ibid.’, PMP *daRat ‘littoral sea, sea near the shore’ > Suwawa dagato ‘ibid.’, PMP *hipaR ‘brother-in-law, sister-in-law’ > Suwawa yipago ‘ibid.’
Finally, the first syllable *mo of the Pre-Proto-Gorontalic mogi- is lost, thus it now appears as *gi-. Loss of the first phoneme or syllable of a (complex) prefix is sporadically observed in Austronesian languages such as Proto-Austronesian *ma-* ‘stative’ from Pre-Proto-Austronesian *kuma-* (< *<um>ka-) (Blust 2003:440 and Ross 1995:740) and Bontok ka- ‘ordinal numerals’ from Proto-Malayo-Polynesian *ika- coming from Proto-Austronesian Sika- (Blust 2003:444). The CV truncation of Pre-Proto-Gorontalic *mogi-* has taken place on the Suwawa prefixal *gi-, possibly in order to avoid the confusion with the different morpheme mogi- that means ‘wear, utilize’ attached to nouns of clothing and tools, such as mogi-abaya ‘wear a dress’ (< abaya ‘dress’) and mogi-sabonge ‘use soap’ (< sabonge ‘soap’).

Note, however, that as discussed in 3.1, in the Gorontalic languages, Proto-Gorontalic *gi-, which reflects Proto-Malayo-Polynesian *maR-si-, does not appear alone as in Tagalog etc. Rather, it co-occurs with Proto-Gorontalic *-a and forms the sociative-progressive circumfix.

In this section, I have argued that the prefixal *gi- in the Suwawa sociative-progressive circumfix historically reflects Proto-Malayo-Polynesian *maR-si- ‘actor voice.irrealis-sociative’ based on comparative evidence. In Section 3.2, I discuss the suffixal *-a of the Suwawa sociative-progressive circumfix *gi- -a, based on its uses and comparative evidence.

### 3.2 The suffixal *-a

Comparatively, it seems unusual to find a form like *-an indicating the progressive meaning in western Austronesian languages. Progressive is indicated inflectionally with a realis marker and monosyllabic reduplication in Tagalog (Schachter and Otanes 1972:346 and Himmelmann 2005b:365)\(^{19}\), or periphrastically with the progressive word sedang in Malay (Sneddon et al. 2010:205). However, there is some evidence that reflexes of the proto-form *-an are present in many western Austronesian languages in different constructions, as seen in the following sections.

#### 3.2.1 The use of the suffixal *-a with verbal bases in Suwawa

In Suwawa, there are at least two unrelated indicative verbal suffixes which have the form *-a, whose proto-form is Proto-Malayo-Polynesian *-an\(^{20}\); one is voice-marking and the other is non-voice-marking. The voice-marking suffix *-a marks the locative voice, which is a subtype of the undergoer voice; it also appears as a locative nominalizer. In contrast, the non-voice-marking suffix *-a appears as part of the sociative-progressive, reciprocal and habitual circumfixes. They are summarized in (10):

\[
(10) \text{Voice-marking and non-voice-marking } *-a.
\]

**Voice-Marking**

- Locative Voice/ Locative Nominalizer

---

\(^{19}\) In Tagalog, this form is also labeled as ‘imperfective’ (see Schachter and Otanes 1972 et passim).

\(^{20}\) Suwawa does not allow consonant-endings, thus the alveolar nasal /n/ in the proto-form *-an is dropped.
Non-Voice-Marking (The prefix -a as part of circumfixes)

- Sociative-Progressive: gi- -a
- Reciprocal\(^{21}\): (a) mo(N)- -a, (b) CV- -a, (c) mo.CV- -a
- Habitual: mo(N)- -a

Here are some examples of the three non-voice-marking constructions with the suffix -a:

**Sociative-Progressive:** gilaqoa ‘be going together’

(11) Tea do gilaqoa ado sikola.
    tea do gi-laqo-a ado sikola
    3.PL already SOC-go-EP to school
    ‘They are going to school together.’

**Reciprocal (a) (mo(N)- -a): moqotawa ‘to know each other’**

(12) Watea moqotawa onota
    watea mo-qa-taw-a onota
    1.SGHON AV.IRR-ST know-EP 3.SG.LOC
    ‘He and I know each other’

*notaalia ‘to buy from each other’*

(13) Ta dea do notaalia.
    ta dea do no-taali-a
    NR(H) two already AV.RLS-buy-EP
    ‘Two people have already bought from each other.’

**Reciprocal (b) (CV- -a): wuwumbada ‘to hit each other’**

(14) Tea wuwumbada.
    tea CV-wumbad-a
    3.PL REC-hit-EP
    ‘They hit each other.’

**Reciprocal (c) (mo-CV- -a): nogogoqoda ‘to hug each other’**

(15) Tea do nogogoqodamaqo.
    tea do no-CV-goqod-a-maqo
    3.PL already AV.RLS-REC-hug-EP-AND
    ‘They hug each other.’

**Habitual:** mongaana ‘eat habitually’

(16) Ti Hasan mongaana no tahu.
    ti Hasan moN-an-a no tahu
    PN Hasan AV.IRR-eat-EP GEN tofu, bean curd
    ‘Hasan habitually eats tofu.’

The question then arises as to what is the commonality that triggers the shared grammatical marker, i.e. the suffixal -a, in these constructions. This is accountable if one considers the following three semantic properties: the number of sub-events, the number of...

\(^{21}\) The affixes mo- -a and mo-CV- -a are inflected for mood: mo-(CV-) -a marks irrealis, while no-(CV-) -a marks realis.
participants, and the temporal organization of sub-events.\textsuperscript{22} I propose that (1) the sociative-progressive, habitual and reciprocal meanings have the common feature that the event indicated by the predicate consists of plural sub-events, (2) the participants are also plural in the sociative-progressive and reciprocal constructions, but not necessarily in the habitual construction and (3) the temporal organization of sub-events is typically simultaneous, but may also be sequential in the sociative-progressive situation (cf. Lichtenberk 2000:37 for his argument on sociative situations)\textsuperscript{23}, simultaneous or sequential in the reciprocal situation (Majid et al. 2011:32), and only sequential in the habitual situation. These features above are summarized in Table 1:

\textbf{Table 1: Properties of sociative-progressive, habitual and reciprocal in Suwawa}

<table>
<thead>
<tr>
<th></th>
<th>Sub-Events</th>
<th>Participants</th>
<th>Temporal Organization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sociative-Progressive</td>
<td>Plural</td>
<td>Plural</td>
<td>Simultaneous (but also Sequential is possible)</td>
</tr>
<tr>
<td>Reciprocal</td>
<td>Plural</td>
<td>Plural</td>
<td>Simultaneous or Sequential</td>
</tr>
<tr>
<td>Habitual</td>
<td>Plural</td>
<td>Singular or Plural</td>
<td>Sequential</td>
</tr>
</tbody>
</table>

The examination into the semantic properties of sociative-progressive, habitual and reciprocal constructions suggest that the common feature is event plurality. With this analysis, I posit that the suffixal \textit{-a} indicates event plurality. The Section 3.2.2 gives comparative evidence for the semantics of the suffixal \textit{-a}.

\textbf{3.2.2 Comparative evidence for Proto-Malayo-Polynesian \textit{*-an}}

As in Suwawa, in some western Austronesian languages, it is observed that there are also at least two distinct suffixes having the proto-form \textit{*-an} that appears with verbal bases; the locative voice marker and the non-voice marker.

I propose that the Proto-Malayo-Polynesian prefix \textit{*-an} ‘event plurality’ has the tendency to form a circumfix. In some Malayo-Polynesian languages, the suffix \textit{*-an} indicating event plurality appears in reciprocal, sociative or distributive constructions. For instance, in Tagalog, the circumfix \textit{mag-} \textit{-an} indicates reciprocal, in Muna, \textit{si-} \textit{-ha} indicates sociative and in Malay, \textit{ber-} \textit{-an} indicates distributive, as exemplified below:

(17) Tagalog: \textit{mag-} \textit{-an} ‘reciprocal’\textsuperscript{24}

\textit{mag-tulong-an} ‘to help each other’ < \textit{tulong} ‘to help’

(Shkarban and Rachkov 2007:900)

\textsuperscript{22} Two distinct temporal organizations, simultaneous and sequential (successive), are employed here. (cf. Lichtenberk 2000:37 and Majid et al. 2011:32).

\textsuperscript{23} As Lichtenberk (2000:37) states, sociativity is irrelevant for the temporal organization of the denoted event. For example, the sociative-progressive form \textit{gipokarajawa} ‘are working’ in the Suwawa sentence \textit{Ibu-ibu bitu gipokarajawa o pest}. ‘Those mothers are working at the party.’ (mother.PL dist SOC-work-EP LOC party) presupposes social cooperation with plural events that are in principle equivalent in terms of their quality and quantity, but with slightly distinct natures such as \textit{purchasing ingredients and party goods, cleaning, discussing the menus, cooking and serving}, which may occur simultaneously or sequentially. The sociative marker for these sub-events does not focus on their temporal organization but on the social cooperation of the overall event in Suwawa.

\textsuperscript{24} \textit{mag-} is a reflex of Proto-Malayo-Polynesian \textit{*maR-} ‘actor voice.irrealis’
Muna: *si- -ha ‘sociative’
   *si-suli- *ha ‘to return together’ < *suli ‘to return’

(Van den Berg 1989:315)

Malay: *ber- -an ‘distributive’
   *ber-lari- *an ‘to run in all directions’ < *lari ‘to run’

(Echols and Shadily 1993:331)

Semantically, as discussed in 3.2.1 for the constructions with the suffixal -a in Suwawa, I argue that the sociative, reciprocal and distributive meanings have the common feature of event plurality. With this comparative data, I conclude that Proto-Malayo-Polynesian *-an indicates event plurality occurring with other verbal prefixes. This suggests that the Suwawa suffixal -a in the circumfix gi- -a marks the progressive meaning, which is a realization of the event plurality.

Typologically, it is observed that event plural markers may also indicate duratives. For example, event plurals are formed by ablaut in Chechen, and this form also indicate duratives depending on lexemes, as in the following examples:

Chechen (North Caucasian)

<table>
<thead>
<tr>
<th>Base</th>
<th>Event plural form</th>
</tr>
</thead>
<tbody>
<tr>
<td>d.uttu ‘to pour’</td>
<td>d.yttu ‘to pour repeatedly’</td>
</tr>
<tr>
<td>molu ‘to drink’</td>
<td>myylu ‘to drink repeatedly’</td>
</tr>
<tr>
<td>loqu ‘to sing’</td>
<td>loequ ‘to sing (several subjects)’</td>
</tr>
<tr>
<td>huttu ‘to stand’</td>
<td>hyttu ‘to stand (several subjects)’</td>
</tr>
<tr>
<td>d.odu ‘to run’</td>
<td>ydu ‘to run for a while’</td>
</tr>
<tr>
<td>teqa ‘to crawl’</td>
<td>tieqa ‘to crawl for a while’</td>
</tr>
<tr>
<td>lakha ‘to look for’</td>
<td>liekha ‘to look for a while’</td>
</tr>
</tbody>
</table>

(Yu 2003:293, 295, 299)

The event plural forms indicate repeated events as in (20) and events performed by several subjects as in (21), but also duratives i.e. events that have prolonged periods as in (22); these functions are comparable to habituas, reciprocals and the sociative meaning of sociative-progressives, as well as their progressive meaning in Suwawa, respectively.

The Chechen examples that event plural forms may also signify duratives give justification for the analysis in which the Proto-Malayo-Polynesian event plural suffix *-an has changed to indicate the progressive meaning in Suwawa.

The semantic change from event-plural to progressive is, however, too involved to be treated here in detail. Thus, for now, it must suffice to point out that the plural sub-events marked with the suffixal -a became unindividuated and then conceptualized as a single progressive event. For a fuller account, see Kitada (forthcoming).

---

25 I suggest that the prefixal part *si- is a reflex of Proto-Malayo-Polynesian *si- ‘sociative’ In some languages, a reflex of the Proto-Malayo-Polynesian prefix *si- ‘sociative’ appears without that of the verbal prefix *maR-. (see Kitada (forthcoming) for further discussion).

26 *ber- is a reflex of Proto-Malayo-Polynesian *maR- ‘actor voice.irrealis’ (Adelaar 1984:417)
4 Conclusion

In the present study, comparative evidence has been presented that suggests that the Suwawa sociative-progressive circumfix gi- -a can historically be divided into two parts, namely the prefixal gi- and the suffixal -a. The prefixal gi- is the reflex of Proto-Malayo-Polynesian *maR-si- ‘actor voice.irrealis-sociative’, and indicates the sociative meaning, i.e. a situation where plural participants in the actor role encoded as a subject perform a cooperative event. The suffixal -a, on the other hand, indicates the progressive meaning, which is a realization of Proto-Malayo-Polynesian *-an indicating event plurality; the plural sub-events are unindividuated and conceptualized as a single progressive event.

Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>AV</td>
<td>actor voice</td>
</tr>
<tr>
<td>DIST</td>
<td>distal</td>
</tr>
<tr>
<td>GEN</td>
<td>genitive</td>
</tr>
<tr>
<td>HON</td>
<td>honorific</td>
</tr>
<tr>
<td>LOC</td>
<td>locative</td>
</tr>
<tr>
<td>NOM</td>
<td>nominative</td>
</tr>
<tr>
<td>PN</td>
<td>proper noun</td>
</tr>
<tr>
<td>PV</td>
<td>patient voice</td>
</tr>
<tr>
<td>SOC</td>
<td>sociative</td>
</tr>
<tr>
<td>REC</td>
<td>reciprocal</td>
</tr>
<tr>
<td>RLS</td>
<td>realis</td>
</tr>
<tr>
<td>AND</td>
<td>andative</td>
</tr>
<tr>
<td>EP</td>
<td>event plurality</td>
</tr>
<tr>
<td>HAB</td>
<td>habitual</td>
</tr>
<tr>
<td>IRR</td>
<td>irrealis</td>
</tr>
<tr>
<td>NM(H)</td>
<td>nominalizer (human)</td>
</tr>
<tr>
<td>PL</td>
<td>plural</td>
</tr>
<tr>
<td>PRG</td>
<td>progressive</td>
</tr>
<tr>
<td>SG</td>
<td>singular</td>
</tr>
<tr>
<td>ST</td>
<td>stative</td>
</tr>
<tr>
<td>REL</td>
<td>relativizer</td>
</tr>
<tr>
<td>VEN</td>
<td>venitive</td>
</tr>
</tbody>
</table>

References


Chung, Henry Yungli, Lillian M. Huang and Dah-an Ho, eds. 2006. *Streams converging into an ocean: Festschrift in honor of Professor Paul Jen-kuei Li on his 70th...*


Nedjalkov, Vladimir P. 2007. Overview of the research. Definition of terms, framework,
The etymology of the sociative-progressive circumfix in Suwawa (GorontaloMongondowic)


Contact-induced sub-dialects in Toda Seediq

AMY PEI-JUNG LEE

1 Introduction

Seediq is an Austronesian language spoken in the central and eastern parts of Taiwan. It comprises one of the two main branches of the Atayalic subgroup (Blust 1999, Li 1981). From both anthropological and linguistic perspectives, Seediq is composed of three main dialects: Tgdaya, Truku, and Toda. In terms of population, the figure is around 5,000, 25,000, and 4,000 for the three dialects, respectively.

The aim of this paper is two-fold: First, this paper offers a description of some grammatical aspects in Toda Seediq, which has been little studied in the literature. Secondly, based on the observation that Toda Seediq has developed into two sub-dialects, Central Toda and Eastern Toda (Lee 2012), this paper proposes that this development is due to long-term contact with its nearby Truku dialect, which is more dominant in terms of distribution and population.

The impact of contact is evidenced by intra-dialectal variations in terms of segments and stress assignment in Central Toda (but not exhibited in Eastern Toda), and phonemic as well as lexical similarities with Truku in Eastern Toda. However, despite being close to the other Seediq dialects, speakers of the two Toda sub-dialects still preserve some phonological features as a way of maintaining self-identity in order not to be totally assimilated with the other dialectal groups.

Therefore, this paper is structured as follows. Section 2 gives a discussion regarding the definition of ‘sub-dialect’ as opposed to ‘dialectal variation’. This is followed by a description of linguistic discrepancies between Central Toda and Eastern Toda, including their phonology, lexicon, and some syntactic structures, so as to support my observation that Toda Seediq has evolved into two sub-dialects, with Eastern Toda being greatly influenced by Truku spoken in Hualien in terms of its phonological processes and morpho-syntactic structures.
2 Defining ‘sub-dialect’

Setting the criteria to distinguish ‘language’ from ‘dialect’ has long been a challenge for linguists. Chambers and Trudgill (1998) point out that the distinction is far from just simply taking ‘mutual intelligibility’ into account. For example, Norwegian, Swedish, and Danish are considered as ‘languages’ not only for linguistic reasons. Political, geographical, historical, sociological, as well as cultural factors all come into play. The distinction of ‘language’ and ‘dialect’ depends on which criterion is considered as the highest priority.

A solution suggested by Chambers and Trudgill (1998: 5) is to employ the neutral term ‘variety’ to refer to any particular kind of language considered as an entity. In this way, ‘dialect’ refers to ‘varieties which are grammatically (and perhaps lexically) as well as phonologically different from other varieties’.

Therefore, in this paper a ‘sub-dialect’, following Chambers and Trudgill’s sense, can be defined as a variety within a variety, which are grammatically (and perhaps lexically) as well as phonologically different from other varieties, but the differences are not to the extent that the native speakers consider themselves as different from those who speak the other varieties. It follows that the definition of ‘sub-dialect’ is not purely linguistic, but also ethno-geographic.

This is the reason why in this paper I adopt the term ‘sub-dialect’ rather than ‘variety’ to apply to the dialectal situation in Toda Seediq. It is generally stated in the literature that there are three main dialects (or varieties) in Seediq: Tgdaya, Truku, and Toda (Holmer 1996, Li 1981, Yang 1976, to name a few). The native speakers of the Toda dialect I interviewed consider themselves as the Toda group from historical and anthropological rather than linguistic perspectives, given that a higher degree of mutual intelligibility is still maintained among the three groups. Intermarriage between different dialectal groups is also common, and some Toda speakers can easily switch to the other dialects.

Geographically speaking, the representative villages of Toda Seediq are Sakura (Snuin) and Teuda (or Toda) in Nantou County, and Tawsay in Hualien County. Based on their grammatical differences, the variety spoken in Sakura and Teuda are termed as ‘Central Toda’ (henceforth CT), and that spoken in Tawsay as ‘Eastern Toda’ (henceforth ET) (Lee 2012). These two varieties are considered as the developing sub-dialects in Toda Seediq.

3 Linguistic differences between Central Toda and Eastern Toda

Given the observation that Toda Seediq has developed into two sub-dialects, Central Toda and Eastern Toda, based on the phonemic inventory (Lee 2012), this section further explores the linguistic differences between the two sub-dialects in terms of phonological processes, lexicon, and grammatical paradigms.\(^1\)

An overview of the phonological as well as lexical similarities and differences found in Central Toda, Eastern Toda, and Truku is provided in Table 1, suggesting that more similarities are found between Eastern Toda and Truku.\(^2\)

---

1 The inventory of consonants in the dialects is as follows: /p, t, k, q, ?, b, d, y, s, x, h, m, n, η, r, ɲ, j, w/ in Truku (Tsukida 2009, Lee 2010), /p, t, k, q, ?, b, d, s, x, h, ts, m, n, η, r, l, j, w/ in Central Toda, and /p, t, k, q, ?, b, d, s, x, h, m, n, η, r, ɲ, j, w/ in Eastern Toda (Lee 2012). To facilitate presentation of the data, some letters are used to replace the IPA symbols: g for [ɣ]; c for [ts] or [tc] (an allophone of /t/ in Truku and Eastern Toda); h for [h]; f for voiced palatal stop [ʃ] (an allophone of /d/ in Tuku and Eastern Toda); ng for [ŋ]; r for [ɾ]; l for [ɭ]; y for [j]. There are four vowels /i, a, o, u/ and three diphthongs /ai, aw, uy/. The mid vowels [e] and [o] are allophones of /ai/ and /aw/, respectively.

2 The Seediq dialects are abbreviated as follows: Truku as Tk, Central Toda as CTd, and Eastern Toda as ETd.
Table 1: Phonological differences in Seediq dialects

<table>
<thead>
<tr>
<th>Difference</th>
<th>Central Toda</th>
<th>Eastern Toda</th>
<th>Truku</th>
</tr>
</thead>
<tbody>
<tr>
<td>/cl/-ls/ correspondence</td>
<td>qcúrux ‘fish’</td>
<td>qsúrux</td>
<td>qsúrux</td>
</tr>
<tr>
<td>[o]-aw/ correspondence</td>
<td>mósə ‘AV.go’</td>
<td>máwsa/mósə</td>
<td>mósə</td>
</tr>
<tr>
<td>[e]-ay/ correspondence</td>
<td>métáq ‘AV.stab’</td>
<td>máytaq</td>
<td>métaq</td>
</tr>
<tr>
<td>/g/-w/ correspondence</td>
<td>əgú/ów ‘many’</td>
<td>lála</td>
<td>lála</td>
</tr>
<tr>
<td></td>
<td>wai ‘that’</td>
<td>wawa</td>
<td>gaga</td>
</tr>
<tr>
<td>stress assignment</td>
<td>dóhá ‘two’</td>
<td>dóha</td>
<td>dóha</td>
</tr>
<tr>
<td>palatalisation</td>
<td>tímú ‘salt’</td>
<td>címu</td>
<td>címu</td>
</tr>
<tr>
<td>vowel harmony</td>
<td>dɔyáw-un ‘help-UV’</td>
<td>jiyáw-un</td>
<td>jyág-un</td>
</tr>
<tr>
<td>metathesis</td>
<td>p-c(hɔ)da-ani CAUS-cook-IMP.CV</td>
<td>p-sdóha-ani</td>
<td>N/A</td>
</tr>
</tbody>
</table>

3.1 Segments

In terms of phonemic inventory, Central Toda differs from Eastern Toda in several points. First, the affricate /c/ (ts) is retained in the former, as it has merged with /s/ in the latter, for example, ricah ‘plum’ in CTD corresponds to risah in ETd. See the following examples:

Table 2: /c/-/s/ sound correspondence in CTD and ETd

<table>
<thead>
<tr>
<th></th>
<th>Central Toda</th>
<th>Eastern Toda</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td>cakus</td>
<td>sakus</td>
<td>‘camphor tree’</td>
</tr>
<tr>
<td>(b)</td>
<td>chiyaw</td>
<td>sbiyaw</td>
<td>‘ago; before’</td>
</tr>
<tr>
<td>(c)</td>
<td>ricah</td>
<td>risah</td>
<td>‘plum (tree)’</td>
</tr>
<tr>
<td>(d)</td>
<td>qcurux</td>
<td>qsúrux</td>
<td>‘fish’</td>
</tr>
<tr>
<td>(e)</td>
<td>ucílun</td>
<td>usílun</td>
<td>‘lake; sea’</td>
</tr>
<tr>
<td>(f)</td>
<td>mi-icu</td>
<td>mi-isu</td>
<td>‘AV-afraid’</td>
</tr>
</tbody>
</table>

Secondly, mid vowels [o] and [e] which occur at word-medial position in Central Toda and Truku correspond to [aw] and [ay] in Eastern Toda. This sound correspondence shows that the proto-Atayalic *aw and *ay as reconstructed by Li (1981) are retained in Eastern Toda, while it has changed to [o] and [e] in other Seediq dialects.3

---

3 However, in ETd [o] also occurs as a variant form, as in the verb /m-usa/ ‘go’, presumably due to contact. The underlying form of the verb ‘go’ is considered as /m-usa/ based on the citation form in the imperative construction usā ‘Go!’. The vowel in this verb is undergoing a sound change in Seediq dialects.
Table 3: [o]-/aw/ correspondence in CTd and ETd

<table>
<thead>
<tr>
<th></th>
<th>Central Toda</th>
<th>Eastern Toda</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td>soki</td>
<td>sawki</td>
<td>‘sickle’</td>
</tr>
<tr>
<td>(b)</td>
<td>doriq</td>
<td>dawriq</td>
<td>‘eye’</td>
</tr>
<tr>
<td>(c)</td>
<td>tokan</td>
<td>tawkan</td>
<td>‘man’s basket’</td>
</tr>
<tr>
<td>(d)</td>
<td>rodux</td>
<td>rawdux</td>
<td>‘chicken’</td>
</tr>
<tr>
<td>(e)</td>
<td>mosa</td>
<td>mawsa/mosa</td>
<td>‘AV.IRR-go’</td>
</tr>
</tbody>
</table>

Thirdly, given the sound correspondence of /g/-/ɣ/-/w/ in Tgdaya, Truku, and Toda, respectively, the voiced velar stop/fricative in other dialects becomes /w/ in Toda (Lee 2012, Li 1981); thus presumably there is no voiced velar stop/fricative in Toda. However, the voiced velar stop or fricative is heard in Central Toda, yet with some variations. It appears that the voiced velar segments in the words borrowed from another dialect either remained the same or changed to the glide /w/. The voiced velar stop is assumed to result from intra-dialectal contact, as the segment is found in Sakura village and those who have been in contact with the Truku people, whereas in Toda village this segment is only found in Japanese loanwords.

Table 4: The velar segments heard in the two villages of Central Toda

<table>
<thead>
<tr>
<th></th>
<th>Sakura village</th>
<th>Toda village</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td>[ʔəŋ.ɡú] or [ʔə.ɡú]</td>
<td>[ʔów] or [ʔúu]</td>
<td>‘many + N’</td>
</tr>
<tr>
<td>(b)</td>
<td>[lāɣay]</td>
<td>not found</td>
<td>‘much + V’</td>
</tr>
<tr>
<td>(c)</td>
<td>[ɣϕ.ϕ-un]</td>
<td></td>
<td>‘steal-UV’</td>
</tr>
</tbody>
</table>

It is possible that the informant who pronounces ‘many’ as [ʔəgu] is influenced by the pronunciation in Tgdaya, though he insists that the word is of Toda origin. Another speculation is that the voiced velar stop is retained in his idiolect, while those speakers in Toda have completed the sound change of /ɡ/ to /w/. If this is the case, then it suggests that the sound change of /ɡ/ to /w/ in Toda dialects occurred in the last sixty years, as my consultant from Sakura village is in his seventies.

On the other hand, the voiced velar obstruent is absent in Eastern Toda, suggesting that the sound change has completed in this sub-dialect.

3.2 Stress assignment

In Seediq dialects stress generally falls on the penultimate syllable (Holmer 1996, Lee 2010, Tsukida 2005, Yang 1976). However, in Central Toda stress falls on the final syllable when the penultimate syllable is a schwa. The following examples show that the constraint *ā (no stressed schwa) is at work in Central Toda. Such a stress assignment seldom occurs in Eastern Toda. Taking the same lexical items in Truku into account, Table 5 shows that Eastern Toda exhibits the same stress pattern as that in Truku. The influence of Truku might have helped Eastern Toda retain the penultimate stressed schwa, while Central Toda is undergoing the stress shift due to avoidance of a stressed schwa.
### Table 5: Stressed syllables in Central Toda and Eastern Toda (cf. Truku)

<table>
<thead>
<tr>
<th></th>
<th>Central Toda</th>
<th>Eastern Toda</th>
<th>Truku</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td>sănáw</td>
<td>sănaw</td>
<td>sănaw</td>
<td>‘man’</td>
</tr>
<tr>
<td>(b)</td>
<td>hənnáng</td>
<td>hənnang</td>
<td>hənnang</td>
<td>‘sound’</td>
</tr>
<tr>
<td>(c)</td>
<td>dəhá</td>
<td>dáha</td>
<td>dáha</td>
<td>‘two’</td>
</tr>
</tbody>
</table>

However, it must be noted that such a stress assignment is only a tendency, as the stressed schwa at penultimate syllable is also attested. Central Toda speakers tend to shift the stress to the final syllable if the penult is a schwa, yet they also accept the variant with a stressed schwa.

A dialectical comparison suggests that such a stress assignment should be a later development based on two observations: First, the stress in most Seediq dialects falls on the penultimate syllable, whereas the stress falling on the final syllable if the penult is a schwa is only found in Central Toda. Secondly, the tendency of avoiding stressed schwa is attested in many languages. This process is also linked with vowel reduction.

### 3.3 Phonological process

In this section three phonological processes are presented: palatalisation, vowel harmony, and metathesis. The following table shows how the three processes occur in Seediq dialects. More details are explored in the sub-sections.

### Table 6: A comparison of phonological process in Seediq dialects

<table>
<thead>
<tr>
<th>Dialect</th>
<th>Process</th>
<th>palatalisation</th>
<th>vowel harmony</th>
<th>metathesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tgdaya</td>
<td>✗</td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Truku</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Central Toda</td>
<td>✗</td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Eastern Toda</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>

### 3.4 Palatalisation

In terms of phonological process, palatalisation is characteristic of the Truku dialect. As for other Seediq dialects, palatalisation of /t/ to [tɕ] and /d/ to [d] before high front segments is completed in Eastern Toda, while in Central Toda it appears to be an ongoing change. On the other hand, this process does not occur in Tgdaya.

Central Toda speakers sometimes alternate between alveolar stops and their palatalised counterparts. Words pronounced with the palatal stops are perceived as variants which do not contrast in meaning. In a way, such a palatalisation in Central Toda can be seen as a phonological feature with idiosyncrasy. The speakers who have had contact with the Truku tend to mix up the alveolar segments with their palatalised variants in their speech.

### 3.5 Vowel harmony

Vowel harmony is commonly seen in Seediq dialects (Lee 2010, Li 1991, Tsukida 2005, Yang 1976). It is a morpheme-internal as well as hetero-morphemic process, which is

---

4 So far the metathesis data in this dialect found in my fieldwork are limited and irrelevant to the discussion here.
triggered by affixation. Eastern Toda differs from Central Toda in the context where vowel harmony is applied.

Given that palatalisation occurs in Eastern Toda, the reduced vowel within a morpheme becomes a full vowel which shares the [+high] feature with its following glide/vowel, as shown in (1), while in Central Toda the schwa remains the same (i.e. not harmonised), as shown in (2).

(1) /dayaw/ ‘help’ (Eastern Toda)  [jiyaw-un=mu ka dəhiya.]
    help-UV=1SG GEN NOM 3PL NOM
    ‘I am helping them.’

(2) /dayaw/ ‘help’ (Central Toda)  [dəyaw-un=mu ka dəhiya.
    help-UV=1SG GEN NOM 3PL NOM
    ‘I am helping them.’

Toda in general tolerates the sequence of a schwa followed by a glide, which is usually harmonised in Truku. Example (3) shows that the antepenult schwa is not harmonised, while in Truku the same form is pronounced as [pi.ya.mi], and in Central Toda it is either [pə.in.ya.mi], in which metathesis also occurs, as shown in (4a), or harmonised, as shown in (4b).

(3) /hapuy/ ‘cook’ (Eastern Toda)  [iya pay-ani ka laqi.
    NEG IMP cook-IMP CV NOM child
    ‘Don’t cook for the children.’

(4) /hapuy/ ‘cook’ (Central Toda)  
    a. [iya pəhay-ani ka laqi.
        NEG IMP cook-IMP CV NOM child
        ‘Don’t cook for the children.’
    b. [usa hapiy-ani laqi han.
        go cook-IMP CV child PART
        ‘Go to cook for the children.’

Another difference pertaining to vowel harmony occurs in prefixation. Like Truku, Central Toda harmonises the vowel in the prefix with the following vowel, while in Eastern Toda no vowel harmony occurs. Compare the following two sentences:

(5) /uyas/ ‘sing’ (Central Toda)  [s-m-kuxun mu-uyas mi r<m>ri ka ubus
    VBLZ-AV likeness AV sing and <AV> dance NOM Ubus
    ‘Ubus likes singing and dancing.’

---

5 The Leipzig Glossing Rules (http://www.eva.mpg.de/lingua/resources/glossing-rules.php) is adopted in this paper, in addition to the following: AV, actor voice; COS, change of state; CV, conveyance voice; IDEO, ideophone; LIG, ligature; LV, locative voice; MOD, modal; PART, particle; STAT, stative; UV, undergoer voice; VBLZ, verbaliser.

6 Diphthongization in Eastern Toda can be triggered by suffixation. See the following examples:
   /paux/ ‘to turn over’
   a. sa-paux [sa.pəux] ‘somersault’
   b. paux-i [pəw.xi] ‘Turn it over (imperative!’

---
Contact-induced sub-dialects in Toda Seediq

As mentioned above, Eastern Toda retains the Proto-Atayalic *aw in both word-medial and word-final position, while it has changed to [ə] at word-medial position in Truku and Central Toda, one is tempted to consider that the diphthong in example (6) might be a result of analogy. However, example (7) suggests that the root of ‘song’ is /əwəyas/.

> (6) /əwəyas/ ‘sing’ (Eastern Toda)
> m-əwəyas kjiyax ka laqi niyi.
> AV-sing often NOM child this
> ‘This child often sings.’

> (7) /əwəyas/ ‘song’ (Eastern Toda)
> po-əyas-i awəyas ka ubus wawa
> CAUS-sing-IMP.UV song NOM Ubus that
> ‘Make that Ubus sing a song!’

Li (1981: 270) states that among the Atayalic dialects he investigates, only Maspaziʔ and Mayrinax retain *aw in non-final position, while in other Seediq dialects it has changed to [ə]. For example, the proto-form of ‘die’ is reconstructed as *mana-hawqil, which is m-huqil in Seediq dialects such as Tongan, Toda, and Inago. My finding in Eastern Toda shows that not all Seediq dialects have undergone this sound change. While some Toda speakers may still retain the proto-sound, the others have been affected by the nearby Truku speech so that [ə] is also heard in word-medial position.

This observation leads to an implication that the vowel shift of aw>ə>u is in progress in Seediq dialects. Whereas in Eastern Toda the proto-sound *aw is retained, in Tgdaya and in Central Toda it has changed to [ə], and in some varieties of Truku it has even changed to [u]. Take ‘chicken’, for example, the following three variations are heard: rawdux (Eastern Toda) ~ rodux (Central Toda, Truku, Tgdaya) ~ rudux (Truku).

In Eastern Toda, parallel to the retention of *aw is the retention of *ay, which occurs in both non-final and final positions. Word-medial *ay has changed to [e] in most Atayalic dialects except Maspaziʔ (Li 1981: 272). This sound change does not occur in Eastern Toda.

However, a word-medial [e] can be derived from vowel harmony in Eastern Toda. When the schwa in a prefix is followed by a high front vowel, it is raised to mid vowel [ə], though this is an optional process.

> (8) /wa inu/ ‘be where’ (Eastern Toda)
> a. we inu ka hiya
> PROG where NOM 3SG.NOM
> ‘Where is he?’
> b. we inu ka kuxun=su
> PROG where NOM likeness=2SG.GEN
> ‘Which one do you like?’

3.6 Metathesis

Metathesis is the process of reversing the order of two segments in a required phonological or morphophonemic context. In Eastern Toda, occurrence of metathesis appears to be triggered by suffixation, especially when the segments /h/, /ŋ/, and /d/ are involved. Two examples are found so far: /hada/ ‘cook (meals)’ and /hangu/ ‘cook (soup)’, which are also semantically related.

Metathesis occurs in imperative construction when the stem is suffixed by -ani or -anaw. For the example /hada/, metathesis is also triggered by the suffix –i, whereas for the other example, metathesis does not occur when suffixed by the same marker. The motivation of this metathesis remains unclear and is pending further research.
(9) /shada/ ‘cook (meals)’ (Eastern Toda)
a. iya p-shada uqun laqi.
   NEG.IMP CAUS-COOK food child
   ‘Don’t cook the food for the children.’
b. iya p-shada-i ka uqun laqi.
   NEG.IMP CAUS-COOK-IMP.UV NOM food child
   ‘Don’t cook the food for the children.’
c. p-dasha-ani uqun ka laqi.
   CAUS-COOK-IMP.CV food NOM child
   ‘Cook meals for the children!’
d. p-dasha-i laqi ka uqun.
   CAUS-COOK-IMP.UV child NOM food
   ‘Tell the child to cook (meals)’

(10) /hangud/ ‘cook (soup)’ (Eastern Toda)
a. s-hangud=ku=na qsurux ka bubu.
   CV-cook.soup=1SG.NOM=3.GEN fish NOM mother
   ‘Mother cooked the fish soup for me.’
b. hangoj-i=saku kingal rawdux
   cook.soup-IMP.UV=2SG.GEN+1SG.NOM one chicken
   ‘(You) Cook a (bowl of) chicken soup for me, please.’
c. hodang-ani=misu han.
   cook=IMP.CV=1SG.GEN+2SG.NOM PART
   ‘Let me cook soup for you.’
d. hodang-anaw=misu damac han ha
   cook-IMP.CV=1SG.GEN+2SG.NOM meals PART IDEO
   ‘Let me cook soup for you first, ok? (While you eat, I have to leave first).’

Similarly, metathesis occurs with the two stems in Central Toda. There is a variation of
the stem /chada/—the syllable [ha] can be dropped.
(11) /chada/ ‘cook’ (Central Toda)
a. usa p-chada uqun laqi
   IMP.AV CAUS-cook food child
   ‘Go to cook (some) food for the children.’
b. p-c(ha)da-ani uqun laqi
   CAUS-cook-IMP.CV food child
   ‘Cook (some) food for the children.’
c. p-dasha-ani uqun laqi
   CAUS-cook-IMP.CV food child
   ‘Cook (some) food for the children.’

(12) /hangud/ ‘cook (soup)’ (Central Toda)
a. h<anguc=ku uqun bubu
   <AV.PROG>cook.soup=1SG.NOM food child
   ‘I am cooking (some) food for the children.’
b. p-hodang-i uqun laqi
   CAUS-cook.soup-IMP.UV food child
   ‘Make (him) cook soups for the children!’

7 The devoicing rule applies to the root /hangud/, which changes the final /d/ to [t] (in ETd) or [ts] (in CTd).
Although the data above show that both Central Toda and Eastern Toda share the metathesis in similar contexts, variations still exist between the two sub-dialects. With the imperative suffix –i, metathesis occurs in Central Toda as in (12b), while in Eastern Toda no metathesis but palatalisation occurs at suffixation, as in (10b).

### 3.7 Morpho-syntactic constructions

The section presents some differences in Central Toda and Eastern Toda regarding morpho-syntactic constructions, including personal pronouns, demonstratives, topicalisation, coordination, negation, and modal expressions.

The following table lists the lexical differences concerning syntactic structures in the three varieties: Central Toda, Eastern Toda, and Truku. It is obvious that Eastern Toda shares the same lexical items with those in Truku, especially the topic marker, the coordinator, the possessive negator, and the volition negator.

<table>
<thead>
<tr>
<th>Difference</th>
<th>Dialect</th>
<th>Central Toda</th>
<th>Eastern Toda</th>
<th>Truku (Eastern)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dem: far/visible</td>
<td>kiya</td>
<td>wa/wawa</td>
<td>gaga</td>
<td></td>
</tr>
<tr>
<td>Dem: far/invisible</td>
<td>wai</td>
<td>wai</td>
<td>gai</td>
<td></td>
</tr>
<tr>
<td>Deictic: far/visible</td>
<td>hiya</td>
<td>hiya</td>
<td>hiya</td>
<td></td>
</tr>
<tr>
<td>Deictic: far/invisible</td>
<td>wai</td>
<td>wowi</td>
<td>gaga hiya/gai</td>
<td></td>
</tr>
<tr>
<td>Topic marker</td>
<td>wa</td>
<td>?u</td>
<td>?u</td>
<td></td>
</tr>
<tr>
<td>Coordinator</td>
<td>mi</td>
<td>mi-ni</td>
<td>ni</td>
<td></td>
</tr>
<tr>
<td>Possessive negation</td>
<td>uka</td>
<td>ungot</td>
<td>ungot</td>
<td></td>
</tr>
<tr>
<td>Volition negation</td>
<td>uxay</td>
<td>aji</td>
<td>aji</td>
<td></td>
</tr>
</tbody>
</table>

### 3.8 Personal pronouns

Central Toda differs slightly from Eastern Toda in personal pronouns. However, before we compare the differences of the pronominal system between Central Toda and Eastern Toda, it might be useful to give a brief comparison of Truku and Toda due to contact issue.

Compared with Truku, Toda does not have oblique personal pronouns, such as knan, sunan, or hiyaan. Instead, nominative personal pronouns also function as oblique pronouns, so that in this paradigm they may be termed as ‘neutral’ (cf. Huang 1995). The following three tables list the pronominal system of personal pronouns in the two dialects.
Table 8: Personal pronouns in Truku (cf. Tsukida 2005)\(^8\)

<table>
<thead>
<tr>
<th></th>
<th>Bound</th>
<th></th>
<th>Free</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Nominative</td>
<td>Genitive</td>
<td>Nominative</td>
<td>Oblique</td>
<td>Possessive</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1SG</td>
<td>=ku</td>
<td>=mu</td>
<td>yaku</td>
<td>knan</td>
<td>nə-naku</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2SG</td>
<td>=su</td>
<td>=su</td>
<td>isu</td>
<td>sunan</td>
<td>nə-nisu</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3SG</td>
<td>=na</td>
<td>hiya</td>
<td>hiyaan</td>
<td>nə-hiya</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1PL.INCL</td>
<td>=ta</td>
<td>=ta</td>
<td>ita</td>
<td>tənan</td>
<td>nə-nita</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1PL.EXCL</td>
<td>=nami</td>
<td>=nami</td>
<td>yami</td>
<td>mənan</td>
<td>nə-nami</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2PL</td>
<td>=namu</td>
<td>=namu</td>
<td>yamu</td>
<td>mənan</td>
<td>nə-namu</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3PL</td>
<td></td>
<td>=dəha</td>
<td>dəhiya</td>
<td>dəhiyaan</td>
<td>nən-dəhiya</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 9: Personal pronouns in Eastern Toda

<table>
<thead>
<tr>
<th></th>
<th>Bound</th>
<th></th>
<th>Free</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Nominative</td>
<td>Genitive</td>
<td>Neutral</td>
<td>Possessive</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1SG</td>
<td>=ku</td>
<td>=mu</td>
<td>yaku</td>
<td>naku</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2SG</td>
<td>=su</td>
<td>=su</td>
<td>isu</td>
<td>nisu</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3SG</td>
<td>=na/=nia</td>
<td>hiya</td>
<td>nə-hiya</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1PL.INCL</td>
<td>=ta</td>
<td>=ta</td>
<td>ita</td>
<td>nita</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1PL.EXCL</td>
<td>=nami</td>
<td>=nami</td>
<td>yami</td>
<td>nami</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2PL</td>
<td>=namu</td>
<td>=namu</td>
<td>yamu</td>
<td>namu</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3PL</td>
<td></td>
<td>=nəha</td>
<td>dəhiya</td>
<td>nən-dəhiya</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 10: Personal pronouns in Central Toda

<table>
<thead>
<tr>
<th></th>
<th>Bound</th>
<th></th>
<th>Free</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Nominative</td>
<td>Genitive</td>
<td>Neutral</td>
<td>Possessive</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1SG</td>
<td>=ku</td>
<td>=mu</td>
<td>yaku</td>
<td>naku</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2SG</td>
<td>=su</td>
<td>=su</td>
<td>isu</td>
<td>nisu</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3SG</td>
<td>=na/=nia</td>
<td>hiya</td>
<td>nə-hiya</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1PL.INCL</td>
<td>=ta</td>
<td>=ta</td>
<td>ita</td>
<td>nita</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1PL.EXCL</td>
<td>=nami</td>
<td>=nami</td>
<td>yami</td>
<td>nami</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2PL</td>
<td>=namu</td>
<td>=namu</td>
<td>yamu</td>
<td>namu</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3PL</td>
<td></td>
<td>=dəha</td>
<td>dəhiya</td>
<td>nən-dəhiya</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Apart from the difference in the oblique case, Toda is different from Truku in the form of the genitive case regarding the third person and the first person plural inclusive. First, in Toda there are two variants of the bound genitive third person singular pronouns: =na and =nia, whose distributions appear to be overlapping. Secondly, the third person plural pronoun in Eastern Toda is =nəha, but in Truku and Central Toda it is =dəha. This shows that the differences of personal pronouns between the two dialects are mainly in the bound form of the genitive pronouns.

---

\(^8\) Tsukida (2005: 302) lists the oblique case of the 1st personal plural exclusive as mənani, which appears to be a typing error.
This comparison shows that Eastern Toda is distinctive for having \( n\text{\textligature{a}}ha \) instead of \( d\text{\textligature{a}}ha \), which is found in the other dialects. The personal pronouns appear to be the most resistant part of this dialect from being influenced by Truku.\(^9\)

### 3.9 Demonstratives

In terms of deictic expressions, Seediq makes a distinction of a three-degree distance from speaker with visibility as a parameter. Based on my fieldwork, Seediq deictics are summarised in the tables below.

**Table 11: Seediq deictics in Truku (cf. Tsukida 2005)**

<table>
<thead>
<tr>
<th>Deictic</th>
<th>Deictic Adverbial</th>
<th>Gloss</th>
<th>Deictic</th>
<th>Deictic Adverbial</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>Near speaker</td>
<td>niyi</td>
<td>this; this one</td>
<td>hini</td>
<td>hini</td>
<td>here</td>
</tr>
<tr>
<td>Far and visible</td>
<td>ga/gaga</td>
<td>that; that one</td>
<td>hiya</td>
<td>hiya</td>
<td>there</td>
</tr>
<tr>
<td>Far and invisible</td>
<td>gai</td>
<td>that; that one</td>
<td>gaga hiya /gai</td>
<td>gaga hiya /gai</td>
<td>over there</td>
</tr>
</tbody>
</table>

In the paradigms the segment /g/ in Truku corresponds to /w/ in Eastern Toda, where the demonstrative pronoun denoting ‘that’ is wawa, which is not found in Central Toda. The equivalent form in Central Toda is kiya.

**Table 12: Seediq deictics in Central Toda**

<table>
<thead>
<tr>
<th>Deictic</th>
<th>Deictic Adverbial</th>
<th>Gloss</th>
<th>Deictic</th>
<th>Deictic Adverbial</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>Near speaker</td>
<td>niyi</td>
<td>this; this one</td>
<td>hini</td>
<td>hini</td>
<td>here</td>
</tr>
<tr>
<td>Far and visible</td>
<td>kiya</td>
<td>that; that one</td>
<td>hiya</td>
<td>hiya</td>
<td>there</td>
</tr>
<tr>
<td>Far and invisible</td>
<td>wai</td>
<td>that; that one</td>
<td>wai</td>
<td>wai</td>
<td>over there</td>
</tr>
</tbody>
</table>

**Table 13: Seediq deictics in Eastern Toda**

<table>
<thead>
<tr>
<th>Deictic</th>
<th>Deictic Adverbial</th>
<th>Gloss</th>
<th>Deictic</th>
<th>Deictic Adverbial</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>Near speaker</td>
<td>niyi</td>
<td>this; this one</td>
<td>hini</td>
<td>hini</td>
<td>here</td>
</tr>
<tr>
<td>Far and visible</td>
<td>wa/wawa</td>
<td>that; that one</td>
<td>hiya</td>
<td>hiya</td>
<td>there</td>
</tr>
<tr>
<td>Far and invisible</td>
<td>wai</td>
<td>that; that one</td>
<td>wowi</td>
<td>wowi</td>
<td>over there</td>
</tr>
</tbody>
</table>

(13) /kiya/ ‘that’ (Central Toda)

\[ \text{uka} \ sapah \ ka \ rudan \ kiya \]
\[ \text{NEG} \, \text{house} \, \text{NOM} \, \text{old.person} \, \text{that} \]

‘That old person has no house.’

(14) /wawa/ ‘that’ (Eastern Toda)

\[ \text{sa-talax-i} \ ka \ damac \ wawa. \]
\[ \text{CV=heat.up=IMP.CV} \, \text{NOM} \, \text{dishes} \, \text{that} \]

‘Heat up/warm up those dishes!’

---

\(^9\) The reviewer points out that if the form \( d\text{\textligature{a}}ha \) found in Central Toda is the result of influence, it would clearly be influenced from Tgdaya, which has \( daha \), thus indicating that the differentiation between Central Toda and Eastern Toda is not just a case of Eastern Toda being subjected to Truku influence, but rather that both are affected, each by its closest neighbour (Truku for Eastern and Tgdaya for Central). However, given that Truku also has the form \( d\text{\textligature{a}}ha \), it is not clear whether this is an influence or rather a reflex of the shared cognate.
104 Amy Pei-jung Lee

(15) /wai/ ‘over there (invisible)’ (Central Toda)
ima wa m-uyas wai
who PROG AV-sing over there
‘Who is singing over there?’

(16) /wowi/ ‘over there (invisible)’ (Eastern Toda)
wam niq wowi ka hulin=mu.
PROG AV-exist over there NOM dog=1SG GEN
‘My dog is over there.’

3.10 Topic marker
In Truku Seediq spoken in Hualien there is a topic marker [ʔu], which is treated as a conjunction by Tsukida (2005). It usually occurs in a topicalised construction. This topic marker is found in Eastern Toda, but absent in Central Toda, suggesting that it is influenced by Truku. The correspondent form in Central Toda is wa, which is usually omitted, whereas /ʔu/ is obligatory in Eastern Toda.

(17) /wa/ ‘topic marker’ (Central Toda)
a. yaku (wa) watan basaw.
1SG.NOM (TOP) Watan Basaw
‘As for me, I am Watan Basaw.’
b. bubu=mu wa, s-hanguc=dąha qsurux ka dąhiya.
mother=1SG GEN TOP, CV-cook=3PL GEN fish NOM 3PL NOM
‘As for my mother, she cooks fish for them.’

(18) /ʔu/ ‘topic marker’ (Eastern Toda)
a. yaku ʔu watan basaw.
1SG GEN TOP Watan Basaw
‘As for me, I am Watan Basaw.’
b. bubu ʔu, s-hangut=nąha qsurux ka dąhiya.
mother TOP, CV-cook=3PL GEN fish NOM 3PL NOM
‘As for Mother, (she) cooks fish for them.’

3.11 Coordinator
In Toda the coordinator is mi, as opposed to ma in Tgdaya and ni in Truku. However, in Eastern Toda it can be either mi or ni. The latter is obviously borrowed from Truku (cf. Lee 2010).

(19) /mi/ ‘COORDINATOR, and’ (Central Toda)
laqi ubus mi yabung do paru kana da.
child Ubus and Yabung CONJ big all COS
‘Ubus and Yabung’s children have all grown up.’

(20) /mi/ or /ni/ ‘COORDINATOR, and’ (Eastern Toda)
yaku ni/mi lupung=mu ka ma-usa matas.
1SG NOM and friend=1SG GEN NOM AV-go AV study
‘My friend and I went to school.’
3.12 Negative constructions

Central Toda and Eastern Toda differ in the use of negators. The following table lists the negators occurring in the three varieties: Central Toda, Eastern Toda, and Truku, showing that some negators in a certain dialect can have multiple semantic functions.

Table 14: Negators in Seediq dialects

<table>
<thead>
<tr>
<th>C. Toda</th>
<th>E. Toda</th>
<th>Truku</th>
<th>Syntactic distribution and function</th>
<th>English Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>uxay</td>
<td>haray</td>
<td>uxay/aji</td>
<td>Before nominal and verbal predicates</td>
<td>‘be not’</td>
</tr>
<tr>
<td>uxay</td>
<td>aji</td>
<td>aji</td>
<td>Before action verbs to negate volition and express prohibition</td>
<td>‘not want to’</td>
</tr>
<tr>
<td>—</td>
<td>aji</td>
<td>ini tduwa</td>
<td>Before action verbs to express prohibition</td>
<td>‘not allowed to’</td>
</tr>
<tr>
<td>uka</td>
<td>uka/ungat</td>
<td>ungat</td>
<td>Before NP arguments in possessive/ existential constructions</td>
<td>‘not have’</td>
</tr>
<tr>
<td>ini</td>
<td>ini</td>
<td>ini</td>
<td>Before action verbs and stative verbs</td>
<td>‘do not’</td>
</tr>
<tr>
<td>iya</td>
<td>iya</td>
<td>iya</td>
<td>Before action verbs in imperative constructions</td>
<td>‘Do not…’</td>
</tr>
</tbody>
</table>

To negate a nominal phrase, Central Toda uses uxay, while Eastern Toda uses haray, or sometimes following Truku, aji. In Eastern Toda, haray is used to negate a phrase, be it nominal or verbal, while aji is used as the same as it is in Truku. Neither aji nor haray is not found in Central Toda.

(21) /uxay/ ‘be not; do not (want to)’ (Central Toda)
   a. uxay=ku sədiq toda
      NEG=1SG.NOM person Toda
      ‘I am not a Toda person.’
   b. uxay=ku mə-kan binlo.
      NEG=1SG.NOM AV-eat betel.nut
      ‘I do not eat betel nuts. (It’s not my habit to eat betel nuts)’
   c. saw bey uxay/uka m-iyah.
      like very NEG AV-come
      ‘(He’s) probably not coming.’

(22) /haray/ ‘be not’ (Eastern Toda)
   a. haray hiyi kacing ka uqun=mu
      NEG meat ox NOM food=1SG GEN
      ‘What I eat is not beef.’
   b. haray=ku ungat pila
      NEG=1SG.NOM NEG money
      ‘It’s not that I have no money.’

To negate an existential or a possessive construction, the dialects in Nantou use uka, while Truku in Hualien uses ungat. According to my consultant, uka used to occur in Eastern Toda, but nowadays it has been replaced by ungat, following Truku.

---

The reviewer suggests that Central Toda might have adopted uxay found in its neighbouring dialect Tgdaya, which has uxe. Although the negator haray does not support the general idea of Truku influence on Eastern Toda, as pointed out by the reviewer, it suffices here to provide an example to demonstrate the differences between the two sub-dialects.
Like most Formosan languages, the negators in Toda Seediq occur in sentence-initial position, which can be followed by another negator to form double negation. The examples above show that all the negators function as verbs, which attract pronominal enclitics. If a negator follows another negator, it forms a serial verb construction.

3.13 Modal expressions

These two Toda sub-dialects also differ in modal expressions, where the combination of lexical items is slightly different. Epistemic modality is expressed by the lexical sequence of *ya be nii* in Central Toda and that of *ida ba ya* in Eastern Toda, both of which function as modal predicate.

(25) *ya be nii* ‘be likely’  (Central Toda)
    niyi saw m-sulung da karac niyi, *ya bey niyi* q<m>uyux da.
    this like AV-cloudyCOS sky thiys,YAvery this <AV>rain COS
    ‘The sky looks cloudy. It is likely to rain.’

(26) *ida ba ya* ‘be likely’  (Eastern Toda)
    *ida ba ya* m-eyah buwihur behing da.
    MOD very YA AV-come wind big COS
    ‘The typhoon is probably coming.’

Directive deontic modality is expressed in Central Toda by the modal verb *aka*, which is unheard of for my Eastern Toda consultant. Eastern Toda, similar to Truku, uses the modal verb *asi* for such an expression.

(27) *aka=su*  cicuh hari nə-mah-an=su.  (CT)
    MOD=2SG.NOM little a.bit PFV-drink-LV=2SG.GEN
    ‘You should drink less.’

(28) *asi=su*  ka m-usa p-sapuh da.  (ET)
    MOD=2SG.NOM NOM AV-go CAUS-medicine COS
    ‘You should go for a medical treatment (for your illness).’

---

11 The reviewer suggests that the form *aka* in Central Toda might be related to Tgdaya *maka*.
12 In Truku the modal verb *asi* expresses the deontic notion of weak obligation, followed by a clause introduced by the marker *ka*. 

(a) *asi=su*  ka m-usa q<m>pah tayhoku.
    MOD=2SG.NOM NOM AV-go <AV>work Taipei
    ‘You should go to work in Taipei. (It is necessary that you go to work in Taipei.)’
4 Concluding remarks

The Toda dialect spoken in Hualien has been influenced by the Truku dialect to some extent. However, the dialect maintains its distinctiveness in phonology, lexical items, and function words, which are distributed in different grammatical constructions.

Internally, the Toda dialect is developing into two sub-dialects, evidenced by the grammatical differences as pointed out in this paper, suggesting that this development is contact-induced. In general, Eastern Toda, being influenced by its nearby Truku dialect, appears to be more homogeneous than Central Toda.

Externally, while confronting nearby dominant dialect, speakers of the dialect still maintain their ethnic identity, which is reflected linguistically in personal pronominal system, the most stable grammatical structure generally immune to language contact.

This paper also suggests that in language or dialect contact, some grammatical features are more prone to borrowing than the others. For example, phonological processes such as palatalisation tend to spread across dialects, as it is the most common sound change in languages. Therefore, palatalisation occurring in Truku spreads to Eastern Toda and becomes an areal feature of the Seediq dialects spoken in the eastern part of Taiwan. Lexical items are generally prone to borrowing, yet this tends to be restricted to the open class. In the case of Eastern Toda, however, such borrowings extend to some function words, including some negators and the coordinator ni from Truku, which are not found in Central Toda.

References


7 Revising the reconstruction of early Austronesian personal pronouns

MALCOLM ROSS

1 Introduction

The purpose of this squib is to revise the reconstruction of Proto Austronesian (PAn) personal pronouns provided by Ross (2006) in the light of the Nuclear Austronesian subgrouping hypothesis (Ross 2009, 2012; Figure 1) and of Zeitoun & Teng’s (2014) modification thereof.

Reconstruction and subgrouping are intimately connected, as the languages one reconstructs are determined by one’s subgrouping, yet the innovations that emerge from reconstruction themselves define one’s subgroups. A subgrouping hypothesis is thus simultaneously a hypothesis about reconstruction. Ross (2006) refers to Sagart’s (2004) subgrouping hypothesis, but does not reconstruct forms for any protolanguage at a node lower than Proto Austronesian (except for some decidedly tentative Proto Malayo-Polynesian forms). The Nuclear Austronesian hypothesis is founded on the observation by Starosta, Pawley and Reid (1981), published as Starosta, Pawley and Reid (2009), that early Austronesian displays verb forms that are derived from nominalising morphology. Starosta, Pawley and Reid attributed this innovation to PAn. Ross (2009) observes that the innovation is not reflected in Tsou, Rukai or Puyuma, and proposes that all Austronesian languages other than these three reflect a shared innovation whereby forms containing certain PAn nominalising morphemes (but not all; see Ross 2012) became finite verbs. The languages which reflect this innovation were dubbed ‘Nuclear Austronesian’ languages. Proto Nuclear Austronesian (PNucAn) was thus the putative ancestor of all Austronesian...
languages other than Tsou, Rukai and Puyuma. Zeitoun & Teng (2014) modify the Nuclear Austronesian hypothesis by showing that Saaroa and Kanakanavu reflect interstages between PAn and PNucAn, as they do not reflect the full set of innovations found in Nuclear Austronesian languages.

When we reexamine the pronominal forms of Formosan languages in the light of the Nuclear Austronesian subgrouping, we find that the pronominal system attributed by Ross (2006) to PAn resembles the one that is now reconstructed for PNucAn, whilst the PAn system is significantly different from it, insofar as two out of its three pronoun sets were not casemarked.

Apart from the reanalysis of the data based on the Nuclear Austronesian subgrouping, the analysis here differs from Ross (2006) in two ways. First, the pronominal data which provide the input to the reconstructions have been updated in certain respects, as indicated in §2. Second, the approach to reconstruction is a little different from that embodied in Ross (2006), insofar as a little more attention is paid here to form and a little less to function and to syntactic status—on the assumption that forms can extend or reduce their functional range quite readily within a pronoun paradigm (e.g. free nominatives or obliques can become free neutral pronouns, or free forms can become clitics).

PAn and PNucAn were both ergative languages, as a majority Formosan languages are. There were two voices, an intransitive actor voice and a transitive undergoer voice. With non-pronominal arguments, the sole argument of an intransitive clause and the undergoer of a transitive clause appeared in nominative case, while the actor of a transitive clause occurred in the genitive (i.e. the same case as a possessor). The actor (intransitive) voice of a semantically transitive verb could take an undergoer argument, marked as oblique and typically generic or indefinite, but sometimes definite and low in discourse salience (Huang & Tanangkingsing 2011). The morphosyntactic systems of Formosan languages are rather different from most ergatively aligned systems, as the actor voice is simply marked as intransitive, i.e. there is no specially marked antipassive.

2 The data

The pronominal data given in the Appendix to Ross (2006) are not repeated here. The data for certain languages are shown both there and here as reconstructed forms based on a number of dialects. These languages are Rukai, Puyuma, Bunun, Atayalic and Amis. There are minor differences between the Proto Rukai and Proto Bunun reconstructions in Ross (2006) and those used here.

Additionally, Pan (2012) provides data for Saaroa and Teng & Zeitoun (2014) for Kanakanavu, and Adelaar (2011) revises the orthography of the extinct language Siraya.

None of these changes/additions has significant bearing on the revised reconstructions offered below, but additional data for Paiwan and Puyuma certainly do.

---

1 It is tempting to use the typologically more transparent labels ‘absolutive’ and ‘ergative’ in place of ‘nominative’ and ‘genitive’, but the ‘genitive’ pronominals are used for both actor and possessor, in Formosan languages, so this is an appropriate label. Descriptions of Austronesian languages of Taiwan and the Philippines and the theoretically oriented literature on Austronesian clitics (Billings & Kaufman 2004) retain ‘nominative’ and ‘genitive’.

2 Proto Atayalic was labelled ‘Proto Atayal’ in Ross (2006). ‘Proto Atayalic’ is a more appropriate label as the reconstruction is based on the dialects of two languages, Atayal and Seediq.

3 The history of Rukai pronominals is discussed by Ross (2013b).

4 De Busser (2009) provides information on Takivatan Bunun that was not previously available.
2.1 Paiwan

The Northern Paiwan pronominal forms presented in the appendix to Ross (2006) are shown in Table 1. These forms are drawn from Huang et al. (1999) and match those in Chang (2006). They also match the cognate forms from other dialects presented by Ferrell (1982), Egli (1990), Early & Whitehorn (2003) and Huang (2012). Curiously, however, there is an additional set of proclitics, shown in (1) that has gone largely unnoticed in the literature. I label them ‘old nominatives’ (NOMo), as they are evidently older than the innovatory nominative forms in Table 1.

(1)  
1SG  2SG  1INCL.PL  1EXCL.PL  2PL
NOMo kəә =, ku=  sa=, su=  tjə=  nə=  nu=

There are no third-person forms, and the vowels of the 1SG and 2SG forms vary across dialects.

Table 1: Northern Paiwan pronominal forms (from appendix to Ross 2006)

<table>
<thead>
<tr>
<th></th>
<th>1SG</th>
<th>2SG</th>
<th>3SG</th>
<th>1INCL.PL</th>
<th>1EXCL.PL</th>
<th>2PL</th>
<th>3PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>NEUT</td>
<td>ti-a-kəә</td>
<td>ti-sun</td>
<td>ti-maju</td>
<td>ti-tjə</td>
<td>ti-a-mən</td>
<td>ti-mun</td>
<td>ti-a-majə</td>
</tr>
<tr>
<td>NOM</td>
<td>=a</td>
<td>kəә</td>
<td>=a</td>
<td>sun</td>
<td>=a</td>
<td>tjə</td>
<td>=a</td>
</tr>
<tr>
<td>GEN</td>
<td>ku=</td>
<td>su=</td>
<td>—</td>
<td>tja=</td>
<td>nia=</td>
<td>nu=</td>
<td>—</td>
</tr>
<tr>
<td>PSR</td>
<td>ni-a-kəә</td>
<td>ni-sun</td>
<td>ni-maju</td>
<td>ni-tjə</td>
<td>ni-a-mən</td>
<td>ni-mun</td>
<td>ni-a-majə</td>
</tr>
<tr>
<td>OBL</td>
<td>tjanu-a-kəә</td>
<td>tjanu-sun</td>
<td>tjanu-maju</td>
<td>tjanu-tjə</td>
<td>tjanu-a-mən</td>
<td>tjanu-mun</td>
<td>tjanu-a-majə</td>
</tr>
</tbody>
</table>

The Paiwan NOMo proclitics have a curious scholarly history. Egli (1990:156) notes the series, but says simply that it occurs ‘after certain expressions’, of which he lists the negators ini ‘not’ and iru ‘certainly not’ and the interclausal conjunction sa ‘and’, all preverbs that are followed by a dependent verb form. Ferrell (1982:114) says that kəә occurs in certain set expressions, and gives i kəә ‘laŋ’ ‘I don’t understand’. The negator i is related to Egli’s ini. Only Early & Whitehorn (2003:574) recognise that these forms are nominative: they say that they follow sa ‘and’ and the genitive proclitic ku= 1SG. Because they are formally very similar to the genitive proclitics (and in some dialects perhaps identical), Chang (2006) and Huang (2012) mistake them for genitives, but context shows that they are not genitive but nominative. Old nominative clitics occur in the data of various authors and various dialects, so their existence is certain. They occur regularly after the preverbs listed by Egli, as the intransitive examples in (2) show.5

(2) Paiwan NOMo in AV clauses

a.  
<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>ru</th>
<th>nə=</th>
<th>kəә</th>
<th>kan</th>
</tr>
</thead>
<tbody>
<tr>
<td>NEG.EMPH</td>
<td>NOMo.1EXCL.PL=</td>
<td>eat.AV.DEP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>‘Surely we won’t eat.’</td>
<td>(Egli 1990:156)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

b.  
<table>
<thead>
<tr>
<th></th>
<th>aku</th>
<th>ini</th>
<th>su=</th>
<th>puvalavalaw</th>
</tr>
</thead>
<tbody>
<tr>
<td>why</td>
<td>NEG</td>
<td>NOMo.2SG=</td>
<td>take.spouse.AV.DEP</td>
<td></td>
</tr>
<tr>
<td>‘Why don’t you get married?’</td>
<td>(Early &amp; Whitehorn 2003, 011:077)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

c.  
<table>
<thead>
<tr>
<th></th>
<th>valjualjut</th>
<th>a</th>
<th>kəә=</th>
<th>pacay</th>
</tr>
</thead>
<tbody>
<tr>
<td>be.living=</td>
<td>NOM.1SG</td>
<td>NEG</td>
<td>LIG</td>
<td>NOMo.1SG=</td>
</tr>
<tr>
<td>‘I am still alive; (I am) not dead yet.’</td>
<td>(Huang 2012:69)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

d.  
|   | ʔ | a|mju~  | kəә | sa  | ku=  | kəә~kəә |
|---|---|----------|---|---|---|---|
| ⟨AV⟩IPFV~cry= | NOM.1SG | and | NOMo.1SG= | IPFV~cook.AV.DEP |

5 Abbreviations follow the Leipzig Glossing Rules. Additions to the latter are AV (Actor voice), DEP (dependent), EMPH (emphatic), NOMo (‘old’ nominative), UV (Undergoer voice).
‘I am/was crying and cooking.’ (Chang 2006:309)

The transitive examples in (3) confirm Early & Whitehorn’s observation that NOMo clitics follow ku= GEN.1SG. (3a) is a dictionary example devoid of context, so it is not clear whether a proclitic sequence occurs other than after a preverb (nu ‘when, if’ in 3b and 3c).

(3) Paiwan NOMo in UV clauses

a. \[ ku= \] nu= \[ səəkəu̯-an \]
   GEN.1SG= NOMo.2PL= send-UV
   ‘I am sending you.’ (Egli 2002:122)

b. \[ nu \] ku= \[ su= \] \[ tju¬mal-an \]
   if GEN.1SG= NOMo.2SG= tell-UV
   ‘If I tell you …’ (Egli 2002:122)

c. \[ nu \] ku= \[ səə= \] \[ aya-in … \]
   when GEN.1SG= NOMo.2SG= say-UV
   ‘When I said to you …’ (Early & Whitehorn 2003, 050:017)

Unlike the intransitive examples in (2) where the verbs are dependent, the verbs in the transitive examples in (3) are independent. The status of these verbs in Paiwan grammar and consequently the history of the construction in (3) are matters for investigation.

These constructions occur alongside the constructions with the innovatory nominative pronouns illustrated in (4).

(4) Paiwan NOM pronouns

a. \[ dʒə̯om`=ə̯a=davac \] =akə̯n
   ⟨AV⟩IPFV=walk =NOM.1SG
   ‘I am walking.’ (Chang 2006:67)

b. \[ ku=ko̯iρ̂om \] =ə̯sun katiaw
   GEN.1SG=PRF=beat.UV =NOM.2SG yesterday
   ‘I beat you yesterday.’ (Chang 2006:65)

All the Paiwan pronoun sets in Table 1, with the exception of the genitive proclitics, are the results of a process noted by Ross (2006) whereby the PNucAn freestanding accusative forms in *ə̯m (see §3.3) extended their function to become neutral forms, which were then casemarked or, in the case of the new nominatives, cliticised to the verb, as shown in (4). These new nominatives have evidently been gradually replacing the old nominatives in (1), which are important because of the light they cast on the functions of the PAn clitics (§3.4). The implications of the Paiwan old nominatives for PAn reconstruction are discussed further in Ross (in press).

2.2 Puyuma

The history of Puyuma pronouns is apparently rather similar to that of Paiwan pronouns. A transitive clause in Nanwang Puyuma, described by Teng (2008), has the structure shown in (5). It differs from the similar Paiwan construction in (4b) in that the Puyuma nominative enclitic reflects a PAn form, unlike the Paiwan enclitic.

(5) Nanwang Puyuma (Taiwan)

\[ ə̯di \] tu= \[ pa-kała̱qam-i \] =ku
NEG GEN.3= CAUS-know-UV =NOM.1SG
‘She didn’t let me know.’ (Teng 2008:207)
Recent work reported in Teng (2013) shows that Puyuma dialects other than Nanwang have the alternative transitive construction illustrated in (6a) whereby a sequence of two clitics, the first genitive, the second nominative, may precede the verb.

(6) Katripul Puyuma (Teng 2013)
   a. \( ku=nu=k\alpha\eta_{\gamma}\-aw \)
      GEN.1SG = NOM.2SG = pull.out-UV
      ‘I pulled you out.’
   b. \( ku=k\alpha\eta_{\gamma}-aw=u \)
      GEN.1SG=pull.out-UV=NOM.2SG
      ‘I pulled you out.’

The construction differs from the Paiwan clitic-sequence construction in that a preceding preverb is not required in Puyuma. However, the clitic sequences that occur in Puyuma are restricted to \( ku=nu= \) as in (6a) and sequences with a third-person genitive \( taw \) (which is apparently not of PAn antiquity as it has no known cognates).\(^6\) The Puyuma ‘old nominatives’, shown in (7), are identical in form to genitives.

(7) NOMo 1SG 2SG 1INCL.PL 1EXCL.PL 2PL
    \( ku=nu=ta=mi=mu= \)

The fact that both Puyuma and Paiwan examples regularly show old nominative clitics occurring only after \( ku= \) GEN.1SG is intriguing. It perhaps implies that the transitive construction with a sequence of clitics was already in the course of replacement when PAn broke up, and that sequences of other genitive clitics plus an old nominative had been lost in PAn, the most immediate shared ancestor of Puyuma and Paiwan. Puyuma is contiguous with Paiwan, and there is lexical evidence of contact between the two languages (Blust 1999:47–51). However, it is rather unlikely that the constructions of the two languages would match each other so exactly as a result of contact.

3 Revising early Austronesian pronominal reconstructions

Table 2 shows the personal pronouns reconstructed by Ross (2006). Table 3 revises that reconstruction. The 3SG forms are omitted, as these entail reconstructive issues that lie beyond the scope of this short paper.

Almost all the forms reconstructed in Table 2 are also reconstructed for PAn or PNucAn in Table 3, but Table 3 differs from Table 2 in two major respects. First, Table 3 contains reconstructions for both PAn and PNucAn, and two PAn sets in Table 2, the neutral and accusative sets, are now reconstructed only for PNucAn. Second, two sets are reconstructed with broader functions in Table 3. The NOM1 set (*aku etc) in Table 2 is a neutral set in Table 3, and the two enclitic sets NOM2 and GEN1 in Table 2 are replaced by a single NOM/GEN1 enclitic set in Table 3. This implies that there were fewer casemarked pronouns in PAn than in most modern Formosan languages. It also responds to an issue discussed by Ross (2006), namely that there seemed to be too many PAn pronoun sets in the 2006 reconstruction.

In the following subsections I present the supporting data for each of the PAn and PNucAn pronoun sets reconstructed in Table 3.

---

\(^6\) The Puyuma third person pronoun \( taw \) looks at first sight like a reflex of PAn *\( Cau \) ‘person’, but this is improbable, as PAn *\( C \) is regularly reflected in Puyuma as \( f \), not as \( t \).
Table 2: Preliminary reconstruction of Proto Austronesian personal pronouns according to Ross (2006)

<table>
<thead>
<tr>
<th></th>
<th>1SG</th>
<th>2SG</th>
<th>1INCL.PL</th>
<th>1EXCL.PL</th>
<th>2PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>NEUT</td>
<td>*-i-aku</td>
<td>*Su[qu]/iSu</td>
<td>*-i-amu</td>
<td>*-i-mu[qu]</td>
<td>*-i-amu</td>
</tr>
<tr>
<td>NOM1</td>
<td>*aku</td>
<td>*Su[qu]</td>
<td>*-i-ami</td>
<td>*mu[qu], *(am)umu</td>
<td>*-i-ami</td>
</tr>
<tr>
<td>ACC</td>
<td>*-i-ak-an</td>
<td>*Su[qu]-an</td>
<td>*ita-an</td>
<td>*i-ami-n</td>
<td>*-i-mu[qu]-n</td>
</tr>
<tr>
<td>NOM2</td>
<td>*=ku, *=Saku</td>
<td>*=Su</td>
<td>*=i-ta</td>
<td>*=mi[a], *=Sami</td>
<td>*=mu</td>
</tr>
<tr>
<td>GEN1</td>
<td>*=a</td>
<td>aku</td>
<td>*=Su</td>
<td>*=i-ta</td>
<td>*=mi[a]</td>
</tr>
<tr>
<td>GEN2</td>
<td>*=m-aku</td>
<td>*=m-iSu</td>
<td>*=m-ita</td>
<td>*((=)m-ami)</td>
<td>*(=)m-amu</td>
</tr>
<tr>
<td>GEN3</td>
<td>*=n-aku</td>
<td>*=n-iSu</td>
<td>*=ni-ta</td>
<td>*=ni-am, *=ami</td>
<td>*=ni-mu, *=n-amu</td>
</tr>
</tbody>
</table>

Table 3: Revised reconstruction of early Austronesian personal pronouns

<table>
<thead>
<tr>
<th></th>
<th>1SG</th>
<th>2SG</th>
<th>1INCL.PL</th>
<th>1EXCL.PL</th>
<th>2PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAn</td>
<td>NEUT</td>
<td>*aku</td>
<td>*Su, *Su[qu]/iSu</td>
<td>*(i)ta</td>
<td>*ami</td>
</tr>
<tr>
<td>PAn</td>
<td>NOM1</td>
<td>*aku</td>
<td>*Su[qu]</td>
<td>*(i)ta</td>
<td>*ami</td>
</tr>
<tr>
<td>PAn</td>
<td>NOM2/NOM1</td>
<td>*(i)uka</td>
<td>*Su[qu]</td>
<td>*(i)ta</td>
<td>*ami</td>
</tr>
<tr>
<td>PAn</td>
<td>ACC</td>
<td>*(i)ak-an</td>
<td>*Su[qu]-an</td>
<td>*(i)ta-an</td>
<td>*(i)am-an</td>
</tr>
<tr>
<td>PAn</td>
<td>NOM3</td>
<td>*=ku</td>
<td>*=Su</td>
<td>*=ta</td>
<td>*=mi[a]</td>
</tr>
<tr>
<td>PAn</td>
<td>GEN3</td>
<td>*=m-aku</td>
<td>*=m-iSu</td>
<td>*=m-ita</td>
<td>*=mi[a]</td>
</tr>
<tr>
<td>PAn</td>
<td>GEN2</td>
<td>*=n-aku</td>
<td>*=n-iSu</td>
<td>*=n-ita</td>
<td>*=ni-am, *=ami</td>
</tr>
</tbody>
</table>

3.1 The free neutral pronouns of PAn and PNucAn

Although a majority of forms in Table 4 are enclitics, it is noteworthy that the Tsou, Pazih, Sai siyat and Amis reflexes are freestanding. As free forms are more likely to become clitics than vice versa, I reconstruct this set as free forms in both PAn and PNucAn. In Proto Rukai and Tsou, reflexes of this set serve as neutral pronouns, reflecting *amis and *aku. As free pronouns are more likely to occur as both nominatives and genitives, pointing to the reconstruction of a PAn neutral function. In NucAn languages reflexes are sometimes nominative—and both in Thao and Amis—suggesting that the PNucAn forms were also neutral. In 2006 genetic reflexes of *aku 1SG were assigned to a reconstructed PAn genitive *=[a]ku. I now think that these reflect later clinisations of the free pronoun *aku.

Turning to the forms themselves, a pair of forms is reconstructed for PAn/PNucAn 2SG, namely *iSu and *Su[qu]. Nanwang Puyuma =yu, Siraya and Thao =uhu and Proto Amis *isu all attest to the form *Su. Forms reflecting *Su[qu] are either monosyllables or lengthened forms reflecting *Su (Proto Rukai *=su, Saarao -u, Proto Bunun *=su) or lengthened forms reflecting *Su[qu] (Tsou suu, Proto Bunun *=suʔu, Pazih siw, Sai siyat foʔo). There are two reconstructive issues here. The first is the status of the initial *i- of *Su. The second is the apparent existence of *Su and *Su[qu] alongside each other.

With regard to the first issue, in the 2006 reconstructions I recognised that *iSu 2SG and *ita 1INCL.PL each occurred with initial *i- as part of the base, i.e. *i- was not the personal article that occurred on forms like *i-aku 1SG (see §3.2). This is evident from the synthetic GEN2 and GEN3 forms in Table 3, where *m- and *n- respectively are prefixed to the pronominal bases of the neutral set reconstructed in Table 4. However, in 2006 I did...
not separate out the reflexes as I have done here. There I reconstructed *iSu[qu] and did not note that *i- and *-qu never cooccur in a single form. The confusion arose because forms with and without *i- are both reconstructable (PNucAn *ita and PAn/PNucAn *ta; PAn/PNucAn *iSu and PAn/PNucAn *Su).  

<table>
<thead>
<tr>
<th>Table 4: Revised reconstruction of early Austronesian free neutral pronouns [1]</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1SG</strong></td>
</tr>
<tr>
<td><strong>PAn</strong></td>
</tr>
<tr>
<td>P-Rukai</td>
</tr>
<tr>
<td>P-Puyuma</td>
</tr>
<tr>
<td>Tsou</td>
</tr>
<tr>
<td>Saaroa</td>
</tr>
<tr>
<td>Kanakanavu</td>
</tr>
<tr>
<td><strong>PNucAn</strong></td>
</tr>
<tr>
<td>Siraya</td>
</tr>
<tr>
<td>P-Bunun</td>
</tr>
<tr>
<td>Thao</td>
</tr>
<tr>
<td>Gen2</td>
</tr>
<tr>
<td>Pazih</td>
</tr>
<tr>
<td>Saisiyat</td>
</tr>
<tr>
<td>P-Atayal</td>
</tr>
<tr>
<td>P-Amis</td>
</tr>
</tbody>
</table>
| Gen | *=aku* | *=isu* | *=ita* | ... | *=amu* | ...
| PMP | NOM2 | *=aku* | ... | *=ta* | ... | ... |

This brings us to the second issue. The ‘lengthener’ *-qu is reconstructable on both *Suqu 2SG and *muqu 2PL. It perhaps reflects lengthening to avoid a monomoraic free pronoun. This cannot be the whole story, however, since *ta 1INCL.PL never acquires a lengthener. A more refined hypothesis is that it reflects lengthening to avoid a monomoraic free pronoun used vocatively.

PNucAn *mu[qu]/ appears alongside *amu. Perhaps the latter arose by analogy with PAn/PNucAn *ami 1EXCL.PL.

A final note: in 2006 I reconstructed the alternant nominative enclitic forms *=[S]aku 1SG and *=[S]ami 1EXCL.PL to account for Proto Atayal *caku and *cami respectively, as well as for Proto Bunun forms whose reconstruction I am now uncertain about. The Atayalic reflexes occur only in Atayal, not in Seedieq. By the criterion that form has precedence, these forms belong to the free set reconstructed in Table 4, as they consist of an unexplained Atayal linker *=c- plus *aku and *ami. Furthermore, Proto Atayalic *c reflects PAn *C, not *S.

3.2 The second set of PNucAn free neutral pronouns

In the set of pronouns in Table 5, *i- is added to those members of the set in Table 4 that do not already begin with *i-. This set was reconstructed for PAn by Ross (2006), but, differentiating diachronically between PAn and PNucAn, it is clear that the set arose only in PNucAn. Its reflexes differ functionally from those in Table 4 in that no set is solely genitive, but the Thao and Saisiyat sets appear to be solely nominative.

---

7 We have no evidence for *ita in PAn: hence the reconstruction *(i)*ita, which says that we do not know whether the PAn variant *ita occurred.
Table 5: Revised reconstruction of early Austronesian free neutral pronouns [2]

<table>
<thead>
<tr>
<th>PNucAn</th>
<th>NEUT2 or NOM</th>
<th>1SG</th>
<th>2SG</th>
<th>1INCL.PL</th>
<th>1EXCL.PL</th>
<th>2PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Siraya</td>
<td>NEUT</td>
<td>yau</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>P-Bunun</td>
<td>NEUT</td>
<td>*iaku</td>
<td>*iSu[qu]</td>
<td>*ita</td>
<td>*i-ami</td>
<td>*i-mu[qu], (*i-amu)</td>
</tr>
<tr>
<td>Thao</td>
<td>NOM1</td>
<td>*iaku</td>
<td>*ita</td>
<td>*dami</td>
<td>*i-ami</td>
<td>*i-mu</td>
</tr>
<tr>
<td>Pazih</td>
<td>NEUT</td>
<td>yaku</td>
<td>ita</td>
<td>yami-n</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Saisiyat</td>
<td>NOM</td>
<td>*iaku</td>
<td>*ita</td>
<td>*i-ami</td>
<td>*i-mu</td>
<td>*i-mu</td>
</tr>
<tr>
<td>P-Atayal</td>
<td>NEUT</td>
<td>(*i-aku)</td>
<td>*isu</td>
<td>(*i-ami)</td>
<td>(*i-amu)</td>
<td>...</td>
</tr>
</tbody>
</table>

PAn *i was a PAn personal determiner. It is reflected as the personal nominative determiner in Puyuma and some Atayalic dialects, and in PAn it contrasted with the genitive determiner *ni. However, the PNucAn personal nominative determiner was evidently *si, and it is thus possible that *i was the neutral (uncasemarked) form. Whether or not it had a casemarking function, however, the occurrence before a pronominal base of the determiner that introduced a personal NP suggests that the pronoun was being treated as if it were a lexical NP, perhaps in preverbal topic position.

3.3 The Proto Nuclear Austronesian ‘accusative’ pronouns

‘Accusative’ in the present context labels a set of pronouns that were used for the undergoer of a PNucAn actor voice verb. As described in the introduction, this undergoer is typically indefinite or lacking discourse salience in Formosan languages, but personal pronouns are exceptions to this tendency and appear with some frequency as undergoers, and PNucAn apparently innovated a special set of pronouns for them. Ross (2006) attributed them to PAn but it is clear from their distribution that they were a PNucAn innovation, as the forms are not reflected in Tsou, Rukai or Puyuma. The accusative pronouns were formed by adding the suffix *-əәn to the forms in Table 5, but the morphophonemics of this addition are not entirely clear, as there is some variation from language to language.

Table 6: Reconstruction of PNucAn accusative pronouns

<table>
<thead>
<tr>
<th>PNucAn</th>
<th>ACC</th>
<th>1SG</th>
<th>2SG</th>
<th>1INCL.PL</th>
<th>1EXCL.PL</th>
<th>2PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paiwan</td>
<td>NOM</td>
<td>*=i[ak-əәn]</td>
<td>*=i[Su-n]</td>
<td>*=i-[ta-əәn]</td>
<td>*=i-[am-əәn]</td>
<td>*=i-mu-n *=i-mu2n</td>
</tr>
<tr>
<td>Takitu’du’ Bunun</td>
<td>ACC</td>
<td>*iak-əәn</td>
<td>*iSu-n</td>
<td>*i-[ta-əәn]</td>
<td>*i-[am-əәn]</td>
<td>*=i-mu-n *=i-mu2n</td>
</tr>
<tr>
<td>Thao</td>
<td>ACC?</td>
<td>yaku</td>
<td>ita</td>
<td>*=i-[am-in]</td>
<td>*=i-[am-in]</td>
<td>*=i-mu</td>
</tr>
<tr>
<td>Saisiyat</td>
<td>ACC</td>
<td>yaku</td>
<td>ita</td>
<td>*=i-[am-in]</td>
<td>*=i-[am-in]</td>
<td>*=i-mu</td>
</tr>
<tr>
<td>PMP</td>
<td>PSR</td>
<td>*=i[ak-əәn]</td>
<td>*=i-[at-əәn]</td>
<td>*=i-[am-əәn]</td>
<td>*=i-mu</td>
<td></td>
</tr>
</tbody>
</table>

As noted in §2.1, Paiwan reflexes of the accusative forms are no longer accusative, having become the base for other pronoun sets. The case of the Thao forms is shown with a question mark, as their case status is not clear (Wang 2004:127–128).

PAn also appears to have had a process whereby the enclitic *=an marked as oblique a personal noun phrase, including a personal pronoun. This included but was not restricted to actor-voice undergoers. In Isbukun Bunun and Pazeh *=an-marked personal pronouns

---

8 The Saisiyat form ʔi-ʃoʔ-o-n 2SG appears to be counter to the generalisation that forms either reflect *iSu or *Suqu, but not *Suqu. I have no explanation of this that is better than a speculation.

9 Seediq and Amis have *=an-marked personal noun phrases other than personal pronouns.
are only used locatively (‘at my place’ etc), but in Rukai, Kanakanavu, Siraya, Seediq, Atayal and Amis they are used in the same way as the accusative pronouns in Table 6. In fact, it seems reasonable to infer that the accusative forms arose as variants of *\(=\)an forms. PAn *\(=\)an forms are not reconstructable, however, since—as noted by Ross (2006)—the enclitic is attached to different bases in different languages.

3.4 The clitic pronouns of PAn and PNucAn

Table 7: Revised reconstruction of early Austronesian clitic pronouns

<table>
<thead>
<tr>
<th></th>
<th>1SG</th>
<th>2SG</th>
<th>1INCL.PL</th>
<th>1EXCL.PL</th>
<th>2PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAn</td>
<td>(*=)ku</td>
<td>(*=)su</td>
<td>(*=)ta</td>
<td>(*=)mi[a]</td>
<td>(*=)mu</td>
</tr>
<tr>
<td>P-Rukai NOM</td>
<td>(*=)ku</td>
<td>(*=)su</td>
<td>(*=)ta</td>
<td>...</td>
<td>(*=)mu</td>
</tr>
<tr>
<td>PSR</td>
<td>...</td>
<td>(*=)su</td>
<td>(*=)ta</td>
<td>...</td>
<td>(*=)mu</td>
</tr>
<tr>
<td>P-Puyuma NOM</td>
<td>(*=)ku</td>
<td>(=)u</td>
<td>(*=)ta</td>
<td>...</td>
<td>(*=)mu</td>
</tr>
<tr>
<td>PSR1</td>
<td>...</td>
<td>(*=)u</td>
<td>(*=)ta</td>
<td>...</td>
<td>(*=)mu</td>
</tr>
<tr>
<td>NOMo/ERG</td>
<td>(*=)ku</td>
<td>...</td>
<td>(*=)ta</td>
<td>(*=)mi</td>
<td>(*=)mu</td>
</tr>
<tr>
<td>Tsou NOM/GEN</td>
<td>(=)ku</td>
<td>(=)u</td>
<td>(=)ta</td>
<td>(=)mi</td>
<td>(=)mu</td>
</tr>
<tr>
<td>Saaroa GEN</td>
<td>(=)ku</td>
<td>(=)u</td>
<td>(=)ta</td>
<td>...</td>
<td>(=)mu</td>
</tr>
<tr>
<td>Kanakanavu NOM</td>
<td>(=)ku</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>GEN1</td>
<td>...</td>
<td>(=)su</td>
<td>(=)ta</td>
<td>(=)mi</td>
<td>(=)mu</td>
</tr>
<tr>
<td>PNucAn NOM2/GEN1</td>
<td>(*=)ku</td>
<td>(*=)su</td>
<td>(*=)ta</td>
<td>(*=)mi[a]</td>
<td>(*=)mu</td>
</tr>
<tr>
<td>Paiwan NOMo</td>
<td>(k=), (s=)</td>
<td>(t=)</td>
<td>(n=)</td>
<td>(n=)</td>
<td>(n=)</td>
</tr>
<tr>
<td>GEN</td>
<td>(k=)</td>
<td>(s=)</td>
<td>(t=)</td>
<td>(n=)</td>
<td>(n=)</td>
</tr>
<tr>
<td>Siraya NOM</td>
<td>(=)ko</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>P-Bunun NOM2/GEN</td>
<td>(*=)ku</td>
<td>(*=)su</td>
<td>(*=)ta</td>
<td>...</td>
<td>(*=)mu</td>
</tr>
<tr>
<td>Pazih GEN1</td>
<td>...</td>
<td>(=)ta</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>P-Atayal NOM</td>
<td>(*=)ku</td>
<td>(*=)su</td>
<td>(*=)ta</td>
<td>(*=)mi</td>
<td>(*=)mu</td>
</tr>
<tr>
<td>GEN</td>
<td>(*=)ku</td>
<td>(*=)su</td>
<td>(*=)ta</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Kavalan NOM</td>
<td>(=)i-ku</td>
<td>(=)su</td>
<td>(=)i-ta</td>
<td>(=)i-mi</td>
<td>(=)i-mu</td>
</tr>
<tr>
<td>GEN</td>
<td>(=)ku</td>
<td>(=)su</td>
<td>(=)ta</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>PMP GEN1</td>
<td>(=)ku</td>
<td>(=)mu</td>
<td>(=)ta</td>
<td>(=)mi</td>
<td>...</td>
</tr>
</tbody>
</table>

\(=\)u is reflected in the Katripul and Tamalakaw dialects of Puyuma (Teng 2013)

The revised reconstruction of early Austronesian clitic pronouns, shown in Table 7, is the area in which the current reconstruction differs most sharply from the 2006 reconstruction. There, as Table 2 shows, I reconstructed two clitic sets, one labelled NOM2, the other GEN1, on the assumption that since modern Formosan languages have two clitic sets, one nominative, the other genitive, this distinction could be inferred for PAn. A close look at form, however, suggests that this is untrue. First, among the non-Nuclear languages, Puyuma and Tsou each use a single set of clitics for actor (Puyuma ERG, Tsou GEN) and undergoer (Puyuma NOMo, Tsou NOM) arguments of transitive verbs. Second, among Nuclear languages, Paiwan, Proto Bunun and Proto Atayalic genitive and nominative reflexes are identical or near-identical. Finally, if the questionable forms in *S- (see §3.1) are omitted, we are left with just one formal difference between the two sets. In Table 2 NOM2.1SG is reconstructed as *\(=\)ku, GEN1.1SG as *\([a]\)ku, i.e. as *\(=\)ku or *\(=\)aku. If forms reflecting *\(=\)aku are taken to be later encliticisations of the free

10 Paiwan n\(\sigma\)=/nia= 1EXCL.PL and nu= 2P have n- where m- is expected. I don’t understand the reason for this, but the fact that both clitic sets were affected implies that they were at the time a single set.
form *aku in Table 4, then there is no formal difference between the two sets of enclitics. Instead, the case of a PAn clitic was indicated by its position. A sequence of two clitics after a preverb or verb were in the order genitive, then nominative, and a single clitic coreferenced the nominative argument of an intransitive verb (Ross 2013a). Included in Table 7 are forms that are enclitic to the preverb or verb, and forms that are usually taken to be proclitic to the verb after preverb deletion (Starosta, Pawley & Reid 1981 [2009]; Ross 2013a).

3.5 The synthetic genitive pronouns of PAn and PNucAn

Non-Nuclear and Nuclear languages both reflect two sets of synthetic genitive pronouns. Both have as their bases the neutral forms in Table 4. The first set is formed by prefixing *m-, the second by prefixing *n[i]-. The identity of *m- is unknown, but *n[i]- reflects *ni, the genitive singular personal determiner reflected as ni in Proto Puyuma, Paiwan, Saisiyat, Proto Atayalic, Proto Amis, Kavalan and Proto Malayo-Polynesian. Because these are synthetic forms, it is possible that in both PAn and PNucAn they were still bimorphic, as suggested by Ross (2006). The data in Table 9 suggest that this was almost certainly true of the forms in *n[i]-, which display variation between apparent *ni- and *n-am in PNucA in both PAn and PNucAn, and between *ni-mu and *n-amu in PNucAn. This variation is explicable if the transition from two morphemes to one took place independently in various daughter-languages. For this reason, ‘PAn’ and ‘PNucAn’ are marked with question marks in Table 9.

Since their synthesis, the synthetic forms have in a number of languages become enclitics. As an accompaniment of voice system changes in pre-Proto Rukai synthetic forms in *m- were reanalysed as neutral bases (Ross 2013b).

Table 8: Revised reconstruction of early Austronesian synthetic genitive pronouns in *m-

<table>
<thead>
<tr>
<th></th>
<th>1SG</th>
<th>2SG</th>
<th>1INCL.PL</th>
<th>1EXCL.PL</th>
<th>2PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>?? PAn</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P-Rukai</td>
<td>NEUT</td>
<td></td>
<td>*maku</td>
<td>*m-iSu</td>
<td></td>
</tr>
<tr>
<td>Kanakanavu</td>
<td>GEN2</td>
<td></td>
<td>=maku</td>
<td>=musu</td>
<td>*miamu</td>
</tr>
<tr>
<td>PNucAn</td>
<td>GEN2</td>
<td></td>
<td>*maku</td>
<td>=musu</td>
<td>*miamu</td>
</tr>
<tr>
<td>Siraya</td>
<td>GEN</td>
<td></td>
<td>=mau</td>
<td>=musu</td>
<td>*miamu</td>
</tr>
<tr>
<td>P-Bunun</td>
<td>GEN</td>
<td></td>
<td>*m[i]mita</td>
<td>m-</td>
<td>*miamu</td>
</tr>
<tr>
<td>Thao</td>
<td>GEN1</td>
<td></td>
<td>m-ihu</td>
<td>m-ita</td>
<td></td>
</tr>
<tr>
<td>Saisiyat</td>
<td>PSR1</td>
<td></td>
<td>m-</td>
<td>m-ita?</td>
<td></td>
</tr>
<tr>
<td>P-Atayalic</td>
<td>GEN</td>
<td></td>
<td>*=maku</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P-Amis</td>
<td>GEN2</td>
<td></td>
<td>*=maku</td>
<td>*=misu</td>
<td></td>
</tr>
<tr>
<td>PMP</td>
<td>GEN</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Comment on the 1EXCL.PL and 2PL forms in Table 8 is in order. It appears that the prefix was simply *m-. Prefixed to *mu 2PL, this gave *m-mu, with an epenthetic vowel in Proto Rukai *mumu and possible reduction of *mm- to *m- in Kanakanavu and Proto Bunun (I cannot say how Siraya =mumi arose; Atayal =m-amu contains the alternative base amu). Prefixed to *mia 1EXCL.PL, reduction of *mm- to *m- seems to have happened early, as reflexes point to PAn/PNucAn *mia. Note, however, that *mia is apparently the clitic from Table 7, not the neutral form *ami from Table 4. This suggests

---

11 For example, Kanakanavu =aku GEN1.1SG is included in Table 4, not in Table 7.
that a PAn *mia perhaps competed with PAn *ami as a free neutral pronoun, but there is no other evidence of this. The form *ami, on the other hand, appears in Proto Malayo-Polynesian *m-ami.

**Table 9:** Revised reconstruction of early Austronesian synthetic genitive pronouns in *n[i]-*

<table>
<thead>
<tr>
<th>?? PAn</th>
<th>GEN3</th>
<th>1SG</th>
<th>2SG</th>
<th>1INCL.PL</th>
<th>1EXCL.PL</th>
<th>2PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>P-Rukai BLT NOM/PSR</td>
<td>*n-aku</td>
<td>...</td>
<td>...</td>
<td>*n-a[m]i</td>
<td>*ni-mu</td>
<td></td>
</tr>
<tr>
<td>MBLT NOM</td>
<td>*n-au=</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P-Puyuma PSR2</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>niam=</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ERG</td>
<td>... nu=</td>
<td>...</td>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>?? PNucAn</th>
<th>GEN3</th>
<th>1SG</th>
<th>2SG</th>
<th>1INCL.PL</th>
<th>1EXCL.PL</th>
<th>2PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>P-Bunun</td>
<td>GEN</td>
<td>*[i]nak</td>
<td>...</td>
<td>...</td>
<td>*[i]n-am</td>
<td>...</td>
</tr>
<tr>
<td>Thao</td>
<td>GEN1</td>
<td>nak</td>
<td>...</td>
<td>n-am</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pazez</td>
<td>GEN2</td>
<td>n-aki</td>
<td>nisiw</td>
<td>nita</td>
<td>ni-am</td>
<td>ni-mu</td>
</tr>
<tr>
<td>Saisiyat</td>
<td>PSR1</td>
<td>...</td>
<td>n-ifo?</td>
<td>...</td>
<td>ni-mo-n</td>
<td></td>
</tr>
<tr>
<td>P-Atayalic GEN</td>
<td>...</td>
<td>...</td>
<td>*=ni-am</td>
<td>(*n-amu)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P-Amis</td>
<td>GEN2</td>
<td>(*n-aku)</td>
<td>(*n-isu)</td>
<td>...</td>
<td>*ni-am</td>
<td>*n-amu</td>
</tr>
<tr>
<td>Kavalan</td>
<td>GEN</td>
<td>...</td>
<td>...</td>
<td>-ni-aq</td>
<td>-numi</td>
<td></td>
</tr>
<tr>
<td>PMP</td>
<td>GEN2</td>
<td>*=n-(a)ku</td>
<td>*=n-ihu</td>
<td>...</td>
<td>*=n-ami</td>
<td>...</td>
</tr>
</tbody>
</table>

The one oddity in Table 9 is the Kavalan forms -niaq 1EXCL.PL and -numi 2PL, for expected niam and nami. However, as (8) shows, Kavalan pronouns have undergone considerable formal reorganisation, such that genitive enclitics (now suffixes) have provided the bases for a new nominative enclitic set with prefixed *i-, and these have in turn provided the bases for new free neutral forms in *a-. The originally synthetic genitive forms -niaq and -numi seem to have replaced earlier =mi and =mu, to judge from the nominative and neutral forms in (8), but the rest of the synthetic genitive forms have vanished and the reformation of the rest of the paradigm left -niaq and -numi with no formal analogues and thus ripe for formal change.

(8) Kavalan

<table>
<thead>
<tr>
<th>1SG</th>
<th>2SG</th>
<th>1INCL.PL</th>
<th>1INCL.PL</th>
<th>2PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>NEUT</td>
<td>a-iku</td>
<td>a-isa</td>
<td>a-ita</td>
<td>a-imu</td>
</tr>
<tr>
<td>NOM</td>
<td>=iku</td>
<td>=su</td>
<td>=ita</td>
<td>=imi</td>
</tr>
<tr>
<td>GEN</td>
<td>-ku</td>
<td>-su</td>
<td>-ta</td>
<td>-niaq</td>
</tr>
</tbody>
</table>

### 4 Conclusions

The initial goal of this short paper was to revise the reconstruction of Proto Austronesian (PAn) personal pronouns provided by Ross (2006) in the light of the Nuclear Austronesian hypothesis and in the process also to reconstruct PNucAn pronouns.

**Table 10:** Revised reconstruction of PAn personal pronouns

<table>
<thead>
<tr>
<th>1SG</th>
<th>2SG</th>
<th>1INCL.PL</th>
<th>1EXCL.PL</th>
<th>2PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>NEUT</td>
<td>*aku</td>
<td>*Su, *Su[qu]</td>
<td>*(i)i</td>
<td>ta</td>
</tr>
<tr>
<td>NOM/GEN1</td>
<td>*=aku</td>
<td>*=Su</td>
<td>*=ta</td>
<td>*=mi[a]</td>
</tr>
<tr>
<td>GEN2</td>
<td>*=m-aku</td>
<td>*=m-su</td>
<td>*=m-ita</td>
<td>*=mia</td>
</tr>
<tr>
<td>?? GEN3</td>
<td>*=n-aku</td>
<td>*=n-su</td>
<td>*=n-ita</td>
<td>*=ni-am, *n-am</td>
</tr>
</tbody>
</table>
In the case of the PAn paradigm, shown in Table 10, the results are more striking than one might have expected. From the seven sets reconstructed by Ross (2006; see Table 4), maximally four—and perhaps just three—date back to PAn. This reduction has two sources: [1] the reconstruction of just one clitic set rather than two, and [2] the attribution of the neutral/nominative set in *i- and the accusative set in *əәn to PNucAn rather than to PAn (Table 11). Adding to this the fact that at least the synthetic genitive set in *ni was possibly still bimorphemic in PAn (i.e. *ni aku etc), we see that the basic PAn system was actually rather simple, with a free neutral set and a nominative/genitive clitic set, neither of which was case-marked, and a probable free genitive set in *m-

PNucAn (Table 11) differed from PAn by the innovation of the neutral/nominative set in *i- and the ‘accusative’ set in *əәn.

**Table 11**: Revised reconstruction of PNucAn personal pronouns

<table>
<thead>
<tr>
<th></th>
<th>1SG</th>
<th>2SG</th>
<th>1INCL.PL</th>
<th>1EXCL.PL</th>
<th>2PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>NEUT1</td>
<td>*aku</td>
<td>*Su, *Su[qu]</td>
<td>*[i]ta</td>
<td>*ami</td>
<td>*mu[qu], *amu</td>
</tr>
<tr>
<td>NEUT2/NOM1</td>
<td>*i-aku</td>
<td>*Su[qu]</td>
<td>*ita</td>
<td>*i-ami</td>
<td>*i-mu[qu], (*i-amu)</td>
</tr>
<tr>
<td>NOM2/GEN1</td>
<td>*=ku</td>
<td>*=Su</td>
<td>*=ta</td>
<td>*=m[i]a</td>
<td>*=mu</td>
</tr>
<tr>
<td>GEN2</td>
<td>*=m-aku</td>
<td>*=m-Su</td>
<td>*=m-ita</td>
<td>*=mia</td>
<td>*=m-amu</td>
</tr>
<tr>
<td>?? GEN3</td>
<td>*n-aku</td>
<td>*n-Su</td>
<td>*n-ita</td>
<td>*n-am, *n-ami</td>
<td>*n-muʔ(*u), *n-amu</td>
</tr>
</tbody>
</table>

Finally, note that reducing the complexity of the PAn paradigm is not a reconstructional achievement in itself. If reconstructed protolanguages have any claim to replicating reality, then they will be as complex and as irregular as languages of the present day. It simply happens that PAn seems to have had a simpler system than most modern Formosan languages—with the exception of Tsou in (9), which retains the simplicity of PAn.

(9) Tsou

<table>
<thead>
<tr>
<th></th>
<th>1SG</th>
<th>2SG</th>
<th>1INCL.PL</th>
<th>1EXCL.PL</th>
<th>2PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>NEUT</td>
<td>aʔo</td>
<td>suu</td>
<td>aʔati</td>
<td>aʔami</td>
<td>muu</td>
</tr>
<tr>
<td>NOM-GEN</td>
<td>-ʔu/-ʔu</td>
<td>-su/-ko</td>
<td>-to</td>
<td>-mza</td>
<td>-mu</td>
</tr>
</tbody>
</table>

**References**


