Issues in Austronesian historical phonology

Lynch, J. editor. *Issues in Austronesian historical phonology*.
PL-550, vii + 227 pages. Pacific Linguistics, The Australian National University, 2003. DOI:*10.15144/PL-550.cover*

Also in Pacific Linguistics

Karen Davis, 2003, *A grammar of the Hoava language, Western Solomons.*

Catharina Williams-van Klinken, John Hajek and Rachel Nordlinger, 2002, *Tetun Dili: a grammar of an East Timorese language.*

John Bowden, 2001, *Taba: description of a South Halmahera Austronesian language.*

Joel Bradshaw and Kenneth Rehg, eds, 2001, *Issues in Austronesian morphology: a festschrift for Byron W. Bender.*

Fay Wouk and Malcolm Ross, eds, 2001, *The history and typology of western Austronesian voice systems.*

Malcolm Ross, Andrew Pawley and Meredith Osmond, 1998, *The lexicon of Proto Oceanic: the culture and environment of ancestral Oceanic socity,* vol. 1: *material culture.*

Pacific Linguistics is a publisher specialising in grammars and linguistic descriptions, dictionaries and other materials on languages of the Pacific, the Philippines, Indonesia, East Timor, southeast and south Asia, and Australia.

Pacific Linguistics, established in 1963 through an initial grant from the Hunter Douglas Fund, is associated with the Research School of Pacific and Asian Studies at The Australian National University. The authors and editors of Pacific Linguistics publications are drawn from a wide range of institutions around the world. Publications are refereed by scholars with relevant expertise, who are usually not members of the editorial board.

Pacific Linguistics 550

Issues in Austronesian historical phonology

Edited by
John Lynch

Pacific Linguistics
Research School of Pacific and Asian Studies
The Australian National University

Published by Pacific Linguistics
Research School of Pacific and Asian Studies
The Australian National University
Canberra ACT 0200
Australia

First published 2003

National Library of Australia Cataloguing-in-Publication entry:

John Lynch, ed.
Issues in Austronesian historical phonology

Bibliography
ISBN 0 85883 503 7

1. Austronesian languages – Phonology. I. Lynch, John.
II. The Australian National University. Research School
of Pacific and Asian Studies. Pacific Linguistics.
III. International Conference on Austronesian Linguistics
(9th : 2002 : Canberra, ACT). IV. International Conference
on Oceanic Linguistics (5th : 2002 : Canberra, ACT).

499.2

Copyedited by Basil Wilson
Typeset by Jeanette Coombes
Cover design by Emily Brissenden
Printed and bound by Union Offset Printers, Fyshwick, Canberra

Table of contents

Contributors to this volume

ROBERT BLUST
Dept. of Linguistics
University of Hawai'i at Mānoa
Honolulu, Hawai'i 96822
USA

JOHN LYNCH
Pacific Languages Unit
University of the South Pacific – Emalus Campus
Port Vila
VANUATU

RICHARD McGINN
Linguistics Department
Ohio University
Athens, Ohio 45701
USA

DAVID MEAD
SIL International
Davao Regional Office
PO Box 81439
8000 Davao City
PHILIPPINES

HANS SCHMIDT
Asien-Afrika-Institut
Abteilung für Indonesische und Südsee-Sprachen
Edmund-Siemers-Allee 1
D-20146 Hamburg
GERMANY

URI TADMOR
Jakarta Field Station, Department of Linguistics
Max Planck Institute for Evolutionary Anthropology
PKBB, Unika Atma Jaya, Jl. Sudirman 51
Jakarta 12930
INDONESIA

RENÉ VAN DEN BERG
SIL
PO Berrimah
Darwin ·NT 0828
AUSTRALIA

JOHN U. WOLFF
Department of Linguistics
217 Morrill Hall
Cornell University
Ithaca, New York 14853
USA

Preface

The Ninth International Conference on Austronesian Linguistics and the Fifth International Conference on Oceanic Linguistics were both held at The Australian National University in Canberra during the month of January, 2002.

Rather than attempt the mammoth task of compiling a single collection of conference papers — which would be quite diverse in scope — the conference organisers favoured a series of smaller compilations on specific topical areas.

This volume represents the first of these compilations, and contains ten papers in the area of Austronesian historical phonology, ranging from John Wolff's survey of the Proto Austronesian phonological system, which appears first in this volume, to detailed studies on individual languages or subgroups, arranged roughly geographically from west to east.

It is hoped that this collection of essays will stimulate further study of a range of issues in Austronesian historical phonology.

John Lynch

Lynch, J. editor. *Issues in Austronesian historical phonology*.
PL-550, vii + 227 pages. Pacific Linguistics, The Australian National University, 2003. DOI:*10.15144/PL-550.cover*

The sounds of Proto Austronesian

JOHN U. WOLFF

1 Introduction

Reconstruction of the phonology of a protolanguage involves two steps: first, listing of correspondence sets in the data provided by the attested languages and, second, determining the phonetic nature of the reconstructed form. This second step is crucial to an evaluation of the validity of the reconstruction. The history of investigations into the articulatory nature of the reconstructed form of Proto Austronesian (PAn) goes back to the earliest attempts of reconstruction. Dempwolff, in his ground-breaking work of the 1930s, which became the point of departure of all historical studies, presents a phonological system which implies certain articulatory characteristics of the reconstructed sounds. Beginning in the 1950s, a view prevailed that the nature of the sounds of the protolanguage were unknowable and irrelevant to the endeavour of historical linguistics, and this view led to the positing of a plethora of protophonemes — consonants which, if they were real, would have made PAn typologically a language with a larger consonantal inventory than the most complex of the Caucasian or the Salish languages, even though the current An languages across the board have a phonological system small and simple by world standards. Beginning in 1988, studies on the articulatory nature of the PAn consonant phonemes appeared (Wolff 1988; Ross 1992). This paper differs from these earlier works in that here we hypothesise a substantially different inventory of protophonemes than has been proposed heretofore in scholarship on the history of the An languages.

I begin by hypothesising a certain phonemic inventory for the protolanguage. This hypothesis is based on the historical An literature heretofore, starting with Dempwolff and the considerable literature of the past sixty years which has led to substantial modification of what Dempwolff first proposed. Then I test this hypothesis by positing a vocabulary of PAn with the hypothesised phonemes on the basis of the data and determining whether the data from current languages manifest regular reflection of these phonemes. This testing process leads to revisions in the hypothesis. An evaluation of the validity of the hypothesis comes from its ability to explain the development of the attested reflexes of protoforms in terms of natural phonological processes, in terms of changes brought about by analogical processes, or in terms of language contact phenomena. This work of testing this hypothesis is not done, but what I have hypothesised is sufficient to explain the data which I have managed to examine up to this point — that is, to explain the development of the attested forms from the hypothesised protoforms in terms of natural linguistic processes.

John Lynch, ed. *Issues in Austronesian historical phonology*, 1–14.
Canberra: Pacific Linguistics, 2003.

Wolff, J.U. "The sounds of Proto Austronesian". In Lynch, J. editor, *Issues in Austronesian historical phonology*.
PL-550:1-14. Pacific Linguistics, The Australian National University, 2003. DOI:*10.15144/PL-550.1*

2 The list of the hypothesised phonemes

Chart I gives a list of the hypothesised phonemes. Note some characteristics of this particular inventory which certainly make it an acceptable inventory if indeed the data support it. First, it is a nicely balanced system. Although I am proposing five points of articulation, such as only a few of the currently attested languages manifest, yet it is still very much on the order of what most of the current languages show — that is, there are no multiple sibilants, and the series of voiced apicoalveolar and apicodental stops is not overly rich: I have assumed just **d* and **j*. Nor have I posited the existence of multiple liquids, the articulatory properties of which are unclear. Further, the voiced and voiceless series is completely congruent. The nasals are congruent as well, for they lack only the very back position, and there is very good reason in terms of how these sounds are articulated for that position to be empty.

Chart I: PAn Consonants

voiced consonants[1]	b	d	j	g	ɣ
voiceless stops	p	t	c	k	q
nasals	m	n	ñ	ŋ	–
liquids	w	l	y		
sibilant		s			

Chart IA shows the relation between my hypothesised system and the transcription found in the Austronesian literature.

Chart IA: Wolff's phonemes and traditionally assigned phonemes

Wolff's transcription	Traditional transcription	Wolff's transcription	Traditional transcription
p	p	ɣ	R
t	C, t	m	m
k	k	n	n
none	T,c	ñ	ñ, N, L
q	q	ŋ	ŋ
b	b	l	l
none	d	none	r
d	D	c	s
j	Z	s	S
none	z		
g	j	w	w
none	g	y	y

In addition PAn had four vowels **i, *e, *a, *u*. There was stress contrastive on the word level, which occurred either on the final syllable or on the penult.

---

1 This series consists of stop consonants except in the case of the **ɣ*, which is post-velar, a position in which voiced stops are rare. The reflexes of **ɣ* are spirants (or developments therefrom) in all languages except in the languages in which **g* merged with **j* or with **ɣ* (in which case the reflex has a sound [g], for there was room for **ɣ* to move to an articulation further forward in the mouth).

3 Assumptions underlying the research

This section makes explicit my basic assumptions about language change.

First, we make two default assumptions — that is, assumptions that so-and-so was the case unless there is evidence to the contrary. The first of these is that phonemic contrasts which cannot be documented did not exist. Accordingly, I published studies that aimed to refute phonemic contrasts which my academic progenitors or my colleagues had proposed (Wolff 1975, 1982, 1997). The grounds for eliminating these phonemes were that no forms evinced the supposed contrasts. The irregular correspondences which these forms reflected and which led to the positing of the supposed contrasts are due to the secondary nature of the forms as borrowings, or they are due to analogical changes, or these are not irregular at all but in fact subject to previously unrecognised environmental conditioning. For example, Dempwolff's dotted **ḷ* is evinced only in forms spread by borrowing, and further his two voiced apical phonemes actually fail to contrast. In addition to the dotted **ḷ* there were other phonemes of Dempwolff's which had to be eliminated on the same grounds. Later on I proposed that the accentual pattern for the root provided an environment for explaining the contrasts between **t* and **C* and **N* or **L* and **ñ*: the proposition here is that there were only two phonemes: **ñ* and **t* (Wolff 1991, 1993).

My second default assumption is as follows: although the protolanguage surely had variation (no language is without, for variation plays a crucial role in establishing social order), unless there is evidence to the contrary, only one of two or more variant forms in the protolanguage can be assumed to have come down to modern times and be manifested in the attested data. To be sure, there are times when the only conclusion to draw is that more than one variant came down from the protolanguage, but there should be evidence for the exceptions. This default assumption is akin to Bloomfield's (1984, Chapter 22) account of competing forms, in which he notes on the basis of numerous examples that one of two competing forms gets lost. Indeed there are times when two variants left reflexes in the daughter languages, but we cannot assume such to be the case without evidence to that effect. For example, the form which is reconstructed with the meaning 'sleep' must be reconstructed both as **tiduɣ* and **tuduɣ*. The distribution of the forms with /i/ as opposed to the forms with /u/ is clearly not phonologically conditioned and has absolutely nothing to do with language groupings, geographical location, or anything which we know about language contact that could explain the occurrence of a reflex of /i/ or of /u/ in the penult of this form. This randomness of the occurrence of the /u/ as opposed to /i/ constitutes evidence for the existence of two variants in the protolanguage, both of which survived.

The most important group of assumptions has to do with the kind of changes which occur and the way in which they occur. First, changes have to be phonologically motivated — that is, the changes develop in conformity with articulatory processes found throughout the languages of the world: for example, assimilation and weakening of an unstressed syllable in languages which have heavy stress, and also in accordance with specific characteristics of Austronesian languages — most importantly, the 'pull' of the canonical disyllabic shape of the root. For example, in the case of a language which manifests intervocalic /-p-/ in a morpheme which is cognate with one in which a sister language manifests intervocalic /-v-/, a hypothesis that an earlier **p* became /v/ makes sense, for this is assimilation to the environment, but the reverse would not be true except in a very special situation. Thus Gitua, a language from New Guinea (Ross 1988:50), reflects PAn **p* with /p-/ initially and /-v-/ medially: PAn **paŋudan* 'pandanus' > Gitua *pada* 'pandanus', cf. Malay *pandan* 'pandanus'; Proto Oceanic **nipi* 'dream' (a re-formation of PAn **sinupi*) > Gitua *vivi*

'dream'(the initial consonant having been changed by an analogical process), cf. Malay *mimpi* 'dream'. Similarly, **s* may and often does become /h/, but the reverse is, as far as I know, not attested.

The reshaping of roots to fit a canonical disyllabic shape has enabled us to reconstruct monosyllabic roots and bring together as a single reconstructed morpheme two, three, and more disparate morphemes with a common meaning (Wolff 1999). For example, for the word 'eat' it is possible to bring together a large number of widely attested and very different forms which derive from one root **kan*: by stretching out the monosyllabic form to two morae, we can connect this root with Ceb *kaʔun* 'eat' and Mongondow *kaʔan* 'eat', whose vowels otherwise would not correspond. Further, we can connect forms which were made disyllabic by re-analysing an affix as a part of the root, even though the affix does not occur currently in the language which manifests this. Thus Malay *pakan* 'feed' is derived from **kan* even though the affix **pa-* is no longer found in Malay. Monosyllabic roots may be made disyllabic by adding a prothetic vowel, as for example in the word for 'coconut'. By hypothesising **ñuɣ* we can connect Ngaju *eñuh* which was disyllabised by adding a prothetic vowel and Malay *ñiur*, which was disyllabised by stretching out the vowel nucleus.

Another example of phonologically motivated change is this: specific phonological characteristics of the protolanguage predetermined changes which occurred. The stress patterns of the protolanguage led to vowel weakening and loss of syllables. We must assume that the accented syllables were pronounced with greater force than the unstressed — i.e. accent primarily involved stress rather than length (although the stressed vowels may indeed have been longer than the unstressed), for this explains the vowel weakening and loss of syllables. In addition certain forms were cliticised (stress-less), and this fact accounts for the various reflexes of these forms found in the data. For example, the word for 'one' is manifested variously as reflecting three protoforms **ica*, **ca* and **eca*.[2] If we assume that this word was typically a clitic (as its reflexes are in most of the currently attested An languages), it is possible to reconstruct only one morpheme **ica* and show that the forms which reflect **ca* and **eca* in fact developed from **ica* by the phonological processes just described. **ca* developed from **ica* by syncope of the initial syllable when in proclitic position, as is widespread in the antepenult of any number of forms in developments throughout the Austronesian area. After **ca* became generalised to stressed position, the form **eca* developed from **ca* by the process of disyllabisation of monosyllabic roots.

Second, we make assumptions about the way sound change takes place — namely, that sound change proceeds on a word-by-word basis, and it is not completed until all forms with the phoneme in a given environment have been replaced by the innovation. These assumptions are based on the discovery by Labov and others in studying on-going changes in English that sound change proceeds morpheme by morpheme within a community as part of the process of creating social structure, and that the change does not begin by replacement of the earlier phonological shape but by the creation of an alternative pronunciation of individual items which then compete with the original and often (but not always) replace it. This implies that the sound change may at times not be carried out to completion. In this way it provides a considerable refinement of the principle enunciated more than 100 years ago by Brugmann that 'sound-laws admit of no exceptions'. We can now say that they admit of no exceptions when they have been completed, but when they remain uncompleted, they show exceptions in the form of changes that never got made. Although it is possible in the

---

2 Examples are as follows: PAn **ica* > Paiwan *ita*, Sa'a *ite*; PAn **ca* > Samoan *sa*, Ratahan *sa*; and PAn **eca* > Cebuano *usa*, Tondano *esa*.

literature to find citations from Austronesianists who pay lip-service to the Neo-grammarian principles of the unexceptionality of sound changes, in fact it is tacitly agreed that indeed there are changes which show exceptions.

An example is found in the history of Javanese. An important change in the development of Javanese is the merger of PAn **d* and **g* as /r/. Only it is not a complete merger. Some forms with **d* become retroflexed /dh/ (i.e. remain unchanged), and a good proportion of these are in fact doublets of forms with /r/ (e.g. **deŋeŋ* 'hear' > Old Javanese *dhengö* and *rengö*.) Now in the case of this particular Javanese development, people have wiggled out of admitting that sound changes may indeed have exceptions by resorting to language contact phenomena to explain the double reflexes, for Javanese was in contact with Malay and strongly influenced by Malay throughout its known history. Malay reflects PAn **d* as /d/, and Javanese borrows with a retroflexed /dh/ forms which in Malay had /d/. In this way it is possible to maintain that Javanese 'borrowed' the forms with /d/ and rescue the Neo-grammarian hypothesis in its pristine form.

However, there is much to indicate that a hypothesis of borrowing does not hold water. First, there are about as many forms which reflect **d* with retroflexed /dh/ as those which reflect this phoneme with /r/, and there is nothing about their semantic characteristics which would induce borrowing of them. Second, a number of these forms do not even occur currently in Malay, so that we would have to assume the occurrence of forms which now no longer exist at an older stage of Malay with which Javanese was in contact. In short, borrowing cannot explain the forms with retroflexed /dh/, and a hypothesis that the change of **d* to /r/ remained incomplete, very much like the incomplete changes which Labov (1994, Parts C and D) adduces for English, makes much more sense. We find similar phenomena of double reflexes that can only be laid to sound changes which did not spread in the vocabulary of languages over the entire Austronesian area from west to east.

4 The articulatory features of the protophonemes

Now that we have the basic assumptions that underlie the methodology, we can finally get to the topic of this paper: what the PAn phonemes sounded like. Chart II lists six reconstructed forms and their reflexes in a language from each of Taiwan, the Philippines, Sumatra, Java and Borneo, along with Malay. This does not cover the entire range of data which shed light on the articulatory characteristics of these phonemes, and we will adduce additional data to elucidate the nature of other phonemes in Pan.

Chart II: Selected forms in six An languages and the PAn reconstructions for them

Meaning	rice	heavy	rot	be flat	road, way	weaver's sword
Reconstruction	**beɣas*	**beɣeqat*	**buyuk*	**dayat*	**jalan*	**baliga*
Paiwan	*vat* 'grain'	*v'qatj*	*vuk*	*kazatjan* 'level land'	*djalan*	*valida*
Cebuano	*bugas*	*bug?at*	*buguk*	*dágat* 'sea'	*dálan*	*balíla*
Toba Batak	*boras*	*borat*	*buruk*	*darat* 'land'	*dalan*	*baliga*
Old Java	*w'as*	–	*wuuk*	*raat* 'world'	*dalan*	*walira* 'part of loom'
Ngaju Dayak	*behas*	*behat*	–	–	*jalan*	–
Malay	*b'ɣas*	*b'ɣat*	*buɣu?*	*dayat* 'land'	*jalan*	*b'liya*

Chart IIA: Reconstructed reflexes for the forms listed in Chart II

PAn	*b	*d	*j	*g	*ɣ	*qeɣ[2]	*c	*t	*k	*ə	*a	*u	*i	*ŋ
Paiwan	v	z, j[1]	z, j[1]	d	Ø	q	t	tj	k	ə	a	u	i	ʔ
Cebuano	b	d	d	l	g	g[ʔ][2]	s	t	k	u	a	u	i	ŋ
Toba Batak	b	d	d	g	r	r	s	t	k	o	a	u	i	n
Old Javanese	w	r	d	r	Ø	Ø	s	t	k	ə	a	u	i	ŋ
Ngaju Dayak[3]	b	r	j	r	h	h	s	t	k	e	a	a	i	ŋ
Malay	b	d	j	ɣ	ɣ	ɣ	s	t	ʔ	ə	a	u	i	ŋ

1 This **d* and **j* fell together in Paiwan and are reflected by two phonemes /z/ and /j/ (written *dj*). Probably the intervocalic reflex was /-z-/ and the initial and preconsonantal reflex was /j/, but analogical developments have obscured this distribution.

2 This sequence is reconstructed on the basis of forms not listed here. Cebuano shows metathesis and syncope.

3 Not all of the Ngaju Dayak reflexes are illustrated in Chart II.

4.1 Voiced stops

First, let us look at the voiced stops. Note that the changes which Charts II and IIA imply are totally natural from the articulatory point of view. The voiced stops continue to be pronounced as voiced stops in the languages illustrated here except for Javanese, where lenition has replaced voicing as the distinctive feature which marks this series off from other stop consonants. The labials remain labials and the consonants which are articulated in the front of the mouth continue to be articulated in the front. The velar consonants which tend to be unstable do indeed change their points of articulation in several of the languages here illustrated. Further, in Javanese and Paiwan **b*, **d*, and **ɣ* develop in entirely analogous and natural ways, where **b* and **d* lose the occlusive feature entirely and become spirants or taps, and **ɣ* becomes lost altogether. Although not all voiced stops are weakened in all positions, there is a palpable consistency in this development which allows us to be secure in the conclusion that these stops were in fact voiced stops. Further, it is unequivocal that the point of articulation implied by the symbols of the chart are in fact the distinctive features of the PAn consonant inventory.

However, the chart does not show this entirely. First is the nature of the **d*/**j* distinction, for further data indicate it to be problematic. Although the chart does not show this, in fact Javanese reflects **d* with a retroflexed /dh/ as well as with /r/. Rukai, an Austronesian language from Taiwan, also reflects these two phonemes with an apicodental and a slightly retroflexed apical stop. Ross (1992) argued that the nature of the distinction in PAn was one of apicodental versus apicoalveolar or retroflex because this is the distinction found in two widely separated languages. Other languages provide little evidence one way or the other as to the nature of the difference, whereas the languages which clearly show a palatalised versus a non-palatalised reflex are contiguous. However, it should be noted that there is no great difference from an articulatory point of view in a contrast consisting of (1) apicodental articulation versus apical stop with slight retroflexion or (2) a contrast consisting of palatalised versus an unpalatalised apical stop. In short, **j* may actually have not been a palatal at all, or at least it developed a non-palatal pronunciation in Rukai, Javanese, and other languages as well.

As to the articulatory features of **g*: we note that in languages in which **g* has not merged with other phonemes it has a voiced alveolar stop reflex medially and finally, but initially it very often has a voiceless velar stop reflex [k]. On the other hand, if **g* was lost, it was by merger with other phonemes, and the nature of the merger can only lead to the conclusions that **g* was a voiced velar stop. In most languages in which **g* has been lost, it has merged with **j*. This fact has led previous scholars to assume a palatalised articulation for **g*, and most certainly **g* was fairly forward — if only to keep it apart from **ɣ*. In Javanese and some other languages **g* merged with **d* rather than **j*. This would make sense if **j* was an apical stop of some sort, and it did develop into that in Javanese. On the other hand **ɣ* merged with **g* in some languages. It makes sense to reconstruct [g] as the value of **-g-* for one more reason: namely, **g-* [g] has to be reconstructed for initial position. This means that a *g*-like sound in medial position must have been the same or nearly the same.

Finally **ɣ*: a voiced back spirant or possibly a back stop (which is acoustically almost indistinguishable from a back spirant) is the only sound which can explain the reflexes of **ɣ*. Chart IIA illustrates the following reflexes of **ɣ*: [ɣ], [g], [r], [h], and ∅. Other languages manifest [l] and [x]. The development of **ɣ* as [x], [h], ∅ involves a devoicing of the **ɣ* followed by a weakening of the voiceless velar spirant to [h] and subsequent complete loss. The change of **ɣ* to [r] and [l] involves a change widespread in the world's languages where a voiced velar spirant becomes a uvular trill and subsequently a tongue-tip trill which may then merge with /l/.

4.2 Voiceless stops

Now for the voiceless stops. Chart III presents data which clarifies the nature of these stops in eight forms from seven languages.

Chart III: Voiceless stops and spirants in seven languages

reconstructed meaning	four	liver	tree	ray of light	one	dog	weep	pull out (sword)
PAn form	**pat*	**qatay*	**kasiw*	**cinaɣ*	**ica*	**acu*	**taɲic*	**sunuc*
Paiwan	*se-patj*	*qatsay*	*kasiw*	*telyar*[1] 'lightning'	*ita*	*vatu*	*tsmangit*	
Amis	*s-pat*	*ʔataɣ*		*cida*	*c-cay*	*waco*	*tomangic*	*hodoc*
Cebuano	*upat*	*atay*	*káhuy*		*isa*		*tangis*	*húnus* 'drawer'
Muna	*fato-*	*ɣate*	*sau*		*ise*	*d-ahu*	–	–
Malay	*empat*	*hati*	*kayu*	*sinar*	*esa*	*gigi-asu* 'canine tooth'	*tangis*	*hunus*
Moken	*pat*	*katay*	*kae*		*sa*		*nangoy*	–
Tongan	*faa*	*ʔate*	*kau*	*maa-hina* 'moon'	*ta-ha*		*tangi*	*unuh-i*

[1] This form is not directly inherited from PAn. The final /r/ does not correspond to Malay /r/. See Chart II.

Chart IIIA: Correspondences of the forms in Chart III

reconstructed phoneme	*p	*t	*c	*k	*q	*s
Paiwan	p	ts, tj[1]	t	k	q	s
Amis	p	t	c	k	qʔ	h[1]
Cebuano	p	t	s	k	Ø, ʔ[1]	h
Muna[1]	p, f	t	s, h	k, s	ɣ	Ø
Malay	p	t	s	k, ʔ[1]	h	Ø, h[1]
Moken	p	t	s	k	k	Ø
Tongan	f	t	h	k	ʔ	Ø

[1] Where there are two reflexes, there is conditioning according to the phonemic environment with some cases of analogical spread. In the case of the Amis reflex of *s: Amis manifests /s/ in some forms in the *s correspondence set, and I have not yet done the research to develop an explanation. The situation is more complex in Muna. Some of the double reflexes arise from phonetic conditioning and analogical spread, but there are other double reflexes which developed when a sound change failed to spread through the entire vocabulary.

The reconstruction of **p* is pretty much self-evident. This is obviously the only sound from which the reflexes in the data could have originated. The /f/ in Tongan is simply a spirantisation of a stop, and this is also reflected in the voiced series (not illustrated here) and it could be argued that /h/ reflecting **c* is another instance of the phenomenon of weakening. Similarly **t* is obviously the only sound from which the reflexes in the data could have originated. Paiwan palatalised or affricated this phoneme, and this process may have been motivated by the depalatalisation of **c* in Paiwan. In any case it is clear that the protophoneme had a sound [t]. It does not make sense to hypothesise a palatal stop for this phoneme in PAn — [c] — for as we shall see immediately following, [c] is the sound which must be reconstructed for forms represented in the third column of the chart. I say the protosound which gave rise to the reflexes exemplified in column three must be reconstructed as **c*, for what other sound can give rise variously to [t], [c], [s], and [h]? We should add to this list the Saisiat reflex palatalised /ß/. In Paiwan (and a number of other languages outside of Taiwan) /t/ is simply a depalatalisation, a natural articulatory development from [c]. The most widespread reflex, /s/, is simply a loss of the stop articulation after the **c* developed a pronunciation [ts], a change similar to that documented for many of the western Romance languages, for the Satem languages of Indo-European, and many others. Tongan /h/ is simply a weakening of an earlier [s], which is manifested in many of the other Polynesian languages.[3]

[3] Dempwolff (1934–38) reconstructed this sound as **t'* (adducing approximately the same sound as I do with **c*). This is remarkably prescient, for none of Dempwolff's languages manifest anything by [s] or what is clearly a development from [s], [h] and Ø. Dempwolff's argument is morphophonemic. That is, in the western An languages (e.g. Malay) in roots beginning with the other stop consonants these initial stops are replaced by homorganic nasals in the morphonemic process which involves nasal replacement. However, for roots beginning with /s/ the /s/ is replaced by /ñ/. This was evidence for Dempwolff (and further evidence for us) that /s/ in Malay and the other Western An languages developed from an earlier palatalised sound.

The reconstruction of **s* fits in perfectly with the reconstruction of **c*. By a natural phonetic change, **s* is weakened to /h/ and then ∅ in some languages, but this happens only in languages in which **c* became /s/ (or became /s/ en route to a further development). In other words, the movement of **s* from its original articulatory features to /h/ is motivated by the development of **c* to /s/, or else the loss of **s* left open the phonetic space for a change of **c* to /s/ without having phonemic merger.

The reconstruction of **k* is uncontroversial. This phoneme is quite stable over the An area, although as we see in Moken and independently in some other languages in the Philippines, on Taiwan, and in Oceania, **k* merges with **q*, but this involved a change in the articulation of **q* rather than **k*. Further, in a few languages **k* became glottalised and the closure with the back of the tongue was lost, resulting in [ʔ].

**q* again must be reconstructed as a postvelar stop, for this is the only sound which could variously have produced [q], [k] [ʔ], [x] (not illustrated in our sample), [ɣ] and [h]. The Amis reflex, a glottal stop with the root of the tongue almost closing the air passage which I symbolise [ʔˤ], is an intermediate stage between the back velar closure [q] and the [ʔ]. The spirantisation of [q] as [x] is a natural development. In fact in the languages which manifest [q] as a reflex of this phoneme, **q* is often articulated with a non-distinctive affricate release — that is, the reflexes of **q* may have affricate allophones. This [x] became voiced in Muna and perhaps other languages, but most frequently it lost its fricative feature and became [h], a natural development paralleled in the history of languages all over the world.

4.3 Liquids and nasals

The liquids and nasals are for the most part stable over the entire An area, and the attested reflexes are with one exception quite similar. We may assume that the articulation of the protophonemes was rather similar to what is generally found as reflexes of them in the documented languages. **w* falls together with the voiced labial stop in a few languages, e.g. Javanese, but this is a matter of weakening of **b* rather than a change of **w*. The phoneme /l/ develops non-lateral articulations in some languages (i.e. becomes [r]), and in some of the languages of Taiwan the reflexes of **l* have a lateral articulation rather different from that found in most of the An languages. /l/ is also velarised and often lost after velarisation. We do not exemplify this phenomenon. It is a natural phonetic development and a widespread process in the the history of languages throughout the world. It took place independently in the An languages which exemplify the phenomenon.

Similarly all the nasals except **ñ* are stable, and except for the fact that some languages lose distinctions in word-final position, the reflexes of these phonemes are pronounced very similarly in the various languages across the area with few exceptions. The phoneme **ñ* has a wide range of diverse reflexes, and we need to look at the reflexes carefully to ascertain the articulatory nature of the protophoneme. Chart IV exemplifies **n*, **l*, and **ñ* in seven languages:

Chart IV: Reflexes of **n*, **l*, and **ñ* in seven languages

reconstructed meaning	six	five	swim	rain	roast	child	float in current
reconstruction	**nem*	**limá*	**ñaŋúy*	**qujáñ*	**tuñú*	**añák*	**qáñuj*
Tsou	*nomə*	*rimo*	*ruu-hnguzu*	*m-əchə*	*chu-a*	*me-ah?o* 'give birth'	*ng-ohcu*
Rukai (Budai)	*ənəmə*	*ríma*	*langoy*	*ódale*	*wa-culo*	*vlakə, la-valakə*	*mw-álodo* 'flow'
Bunun	*nuum*	*hima?*	–	*qudan*	*ma-tunu*	–	*mung-qanu?*
Paiwan	*enem*	*lima*	*lymanguy*	*qudjaly*	*culyu*	*alyak*	*si-qalyudj*
Amis	*q?nem*	*lima*	*dangoy*	*q?orad*	*todoh*	–	*q?alol*[1]
Tagalog	*ánim*	*limá*	*langúy*	*ulán*	–	*anák*	*ánud*
Malay	*enam*	*lima*	–	*hujan*	*tunu* 'burn'	*anak*	*hanyut*

[1] The reflex medial /l/ in place of medial /d/ (the normal reflex of **ñ*) is here probably conditioned by the following /l/.

Chart IV A: Correspondences of **n*, **l*, and **ñ* illustrated in Chart IV

reconstructed phoneme	*n	*l	*ñ
Tsou	n	r	h
Rukai	n	r	l
Bunun	n	h	n
Paiwan	n	l	ly
Amis	n	l	d[1]
Tagalog	n	l	n, l[2]
Malay	n	l	ñ, n, l[2]

[1] Amis /d/ represents a palatalised phoneme with two allophones distributed according to environment (but differently in different dialects) (1) a voiceless palatal [ɬ] or (2) [d] (possibly palatalised [d'] — the literature is unclear).

[2] Malay and many languages of the Philippines and western Indonesia have three reflexes depending on the environment (but redistributed by analogical changes — cf. Wolff 1993). Other languages of the Philippines and Indonesia independently merged earlier **ñ* and **n*.

As for the phonetic characteristics of **ñ*, first, we note that in most of the languages of Taiwan (exemplified here by Paiwan and Amis), the reflex of **ñ* has a feature of palatalisation. In Rukai, this feature was lost (although there is not enough information available about the phonetics of the various Rukai dialects to enable me to state that this feature is not in fact preserved in some of the Rukai dialects). In Tsou it is clear that /h/ derives from a palatalised [ɬ] which was devoiced and then lost the lateral articulation; in Saaroa, one of the languages most closely related to Tsou, **ñ* is still reflected as /ɬ/. In Malay and other An languages outside Taiwan, palatalisation was a feature preserved only in certain environments (and in most of these languages was independently lost in the environments in which it was retained in Malay — Wolff 1993). Further, **ñ* is reflected as /l/ in certain environments but as [ñ] or [n] in others, at least in the Western An languages

outside of Taiwan. Finally, we note that **ñ* merges with **n* in Bunun and some other languages of Taiwan. In short, the evidence is clear that **ñ* had a palatalised articulation.

On the other hand it is not possible to say that **ñ* was a nasal and not a lateral or vice versa. It clearly has reflexes with a lateral articulation over a wide range of languages, extending minimally as far as western Indonesia. In fact, there are forms found in Eastern Indonesian and Oceanic languages with a lateral reflex in cognates which are derived from a PAn form with **ñ*. However, it is not certain whether or not the forms which evince /l/ as a reflex of PAn **ñ* in eastern Indonesia and in Oceania are in fact directly inherited. They may have been spread secondarily. In any case, the lateral reflex is widespread inside as well as outside of Taiwan. This argues for a hypothesis that **ñ* had lateral allophones. The nasal reflex /n/ is also widely distributed in languages ranging from Taiwan through the Pacific, and this argues that **ñ* had nasal allophones. From an articulatory point of view there is not a great deal of difference between a palatalised [l̡] and a palatalised [ñ]. I conclude that in PAn there were two allophones of **ñ*, both palatalised, one with a lateral articulation and the other with a palatalised articulation, and the distribution may well have been tied to the stress pattern of the root, just as is the case in the western Austronesian languages outside of Taiwan (Wolff 1993).

5 Other articulatory characteristics of the PAn phonology

There are several articulatory characteristics of the PAn phonology which can be deduced from the attested data and which had a role in the development of this phonology to the current time. These processes must have come into play during PAn times because they are found over the entire range of An languages and are still in operation in many of our languages. First is the canonical character of the root as a disyllabic which caused the reformation of monosyllabic roots to disyllabics (see §3 above). It is this factor which enables us to connect disparate forms with similar meanings. A second process is the effect of the stress which caused vowel weakening and syllable loss. PAn did not have any roots greater than three syllables. Reconstructed forms with four or more syllables contain what are clearly petrified affixes (e.g. the prefix **qañi-* found on forms referring to fauna, supernatural beings, etc. and illustrated in the word for 'honey bee' on Chart V below).[4] The general tendency is to transform roots with three syllables into roots with two, but this is a language-by-language process. There are some trisyllabics, which show no or very little syncopation anywhere, nor do they show vowel weakening. There are some which are weakened but not disyllabised; there are some which are disyllabised by loss of the initial vowel in one set of otherwise ungrouped languages and are disyllabised by loss of the medial vowel elsewhere; and there are some which are disyllabised in all languages which reflect them. In this last case a legitimate question may be raised as to why they should be reconstructed as trisyllabics at all — a question to be addressed shortly. The reasons for the differential treatment of the trisyllabics are only partly explained so far. It is a detailed question of the histories of the individual languages which remain to be unravelled. Chart V illustrates these phenomena from a range of languages across the board.

---

4 The one exception is the form **qasulipan* 'millipede', which has no recognisable affix. However, it is to be noted that the Bunun reflex treats the root as **qasulip* — Bunun *qapis* 'centipede', as if the **-an* manifested in the reflexes of this etymon in other languages were a suffix.

Chart V: Selected roots of more than three syllables in selected languages

reconstructed meaning	day, sun	night	honey bee	gall	star
PAn form	**qañegaw*	**ɣabii*	**qañiɣuwan*	**qapegu*	**bintuqén*
Paiwan	*qadaw*			*qapedu*	*vitjuqan*
Bunun		*labi-an*		*paqav*	*bintuqan*
Cebuano	*adlaw*	*gabí?i*	*ligwan*	*apdu*	*bitú?un*
Tondano	*edo* 'day'		*nerua*	*peru*	
Manggarai	*leso*	*wié* 'yesterday'		*pesu*	
Bugis	*esso* 'day'	*karawian* 'late afternoon'		*essung*	*wittoeng*
Muna	*ɣoleo* 'day'	*indewi* 'yesterday'	*ka-eniua* 'k.o. bee'	*ɣofei*	
Malay		Toba Batak *robi* 'long ago'	*ñaruan* (Jakarta dialect)	*hampedu*	*bintang*
Ngaju	*andaw*		*ñuan* 'k.o. ant'	*peru*	< Malay
Fijian (Wayan)			*oni* 'k.o. bee'		
Tongan	*?aho* 'day'			*?ahu*	*fetu?u*

In the case of the forms for 'day', 'honey bee' and 'gall' we must conclude that the reflexes derive from forms with three or more syllables, for this is the only good way to connect the illustrated forms with each other. Manggarai and Bugis (and reflexes in many other languages) lose the initial syllable, whereas Ngaju, Tongan, Paiwan, Cebuano, Chamorro and Tondano forms syncopate the medial syllable (and then proceed to make other changes as well). Muna retains the original number of syllables (and in fact adds another syllable in the word for 'bee'). The same is the case for the word for 'gall', except that in the case of this root Paiwan and Malay retain three syllables as does Muna, and the Tondano and Ngaju reflexes lose the initial syllable rather than the medial syllable. Bunun forms a disyllabic root by changing the final **u* to /v/.[5]

It could be argued that the PAn forms were disyllabic and became trisyllabic by vowel epenthesis and then later on reduced. Such a series of changes is certainly possible. However, the form for 'night' and others like it prove that epenthesis is not an explanation for the development of trisyllabic roots. Rather, the trisyllabic roots must have been inherited as such, for in these forms, where a trisyllabic root is retained, the penult is not the reflex of a centralised vowel **e*. Epenthesis would be a possibility if all forms which reflect a medial consonant sequence are cognate with forms which reflect two consonants separated by **e*. But in the form for 'night', for example, languages which reflect three syllables (here Amis, Cebuano, and Tondano) have /i/ in the penult. The most likely conclusion, then, is that the protolanguage had no medial consonant clusters other than sequences of nasal + C (nor for that matter had consonant clusters anywhere within the root). We reconstruct protoforms with a medial **CeC* even in the case of those forms which are reflected only with disyllabic roots having medial consonant clusters. It should be noted that there are few cases of this sort. In most cases in one language or another there is evidence for a vowel separating the medial cluster either by the retention of this vowel or by the fact that the

5 The changes are as follows, beginning with the loss of **-g-*, and they are paralleled in other forms: **qapegu > *qapeu > *paqeu > paqau > paqav*.

process of disyllabisation was accomplished by loss or weakening of the intial syllable in some languages as well as by loss of the medial syllable in others.

In short, there were no consonant clusters within the word. The exception is that stops could be preceded by a homorganic nasal medially. This is illustrated by the word for 'star' in Chart V, which we must reconstruct with a nasal preceding the onset of the penult. The loss of the nasal in reflexes of the word for 'star' in several languages is a function of the distance of the cluster from the stressed final syllable. There was also a process of sporadic inserted nasalisation, but this is attested only in languages outside of Taiwan and may in fact not have been part of the PAn phonological system.

However, there were vowel sequences, as the word for 'night' illustrates. Although Cebuano and Tondano (as well as other languages in the Philippine group) eliminate vowel sequences by inserting a glottal stop in the hiatus between the vowels, Amis shows hiatus. If the word for 'night' had had a **q* before the final syllable, Amis would have reflected this with a glottal stop [ʔ].

Stress in the word was on the penult or on the final syllable. Except for the case of proclitics, there are no examples to my knowledge of the loss of final syllables. If there had been roots of three syllables or more with a strong stress on the antepenult or earlier and no stress on the rest of the root, reflexes manifesting loss of the final syllable would have developed.

References

Bloomfield, Leonard, 1984, *Language*. Chicago: University of Chicago Press. [Reprint]

Dempwolff, Otto, 1934–38, *Vergleichende Lautlehre des austronesischen Wortschatzes*. Beihefte 13, 15 and 17 of *Zeitschrift für Eingeborenen-Sprachen*. Berlin: Dietrich Reimer.

Labov, William, 1994, *Principles of linguistic change*, vol.1: *Internal Factors*. Oxford: Basil Blackwell.

Ross, Malcolm D., 1988, *Proto Oceanic and the Austronesian languages of Western Melanesia*. Canberra: Pacific Linguistics.

1992, The Sound of Proto-Austronesian: an outsider's view of the Formosan evidence. *Oceanic Linguistics* 31/1:23–64.

Wolff, John U., 1975, Proto-Austronesian *r and *d. *Proceedings of the First International Conference on Comparative Austronesian Linguistics. Oceanic Linguistics* 13:77–121.

1982, PAN *c, *z, *g, and *T. In Amran Halim, Lois Carrington and S.A. Wurm, eds *Papers from the Third International Conference on Austronesian Linguistics,* vol.2: *Tracking the travellers*, 1–30. Canberra: Pacific Linguistics.

1988, The PAN consonant system. In Richard McGinn, ed. *Studies in Austronesian Linguistics,* 125–147. Monographs on International Studies, SE Asia Series, No.76. Athens, Ohio.

1991, The Proto-Austronesian phoneme *t. In Robert Blust, ed. *Currents in Pacific linguistics: papers on Austronesian languages and ethnolinguistics in honour of George W. Grace*, 535–549. Canberra: Pacific Linguistics.

1993, The PAN phonemes *ñ and *N. *Oceanic Linguistics*, 32/1, 45–61.

1997, The Proto-Austronesian voiced apical and palatal stops *d and *j. In Cecilia Odé and Wim Stokhoff, eds *Proceedings of the Seventh International Conference on Austronesian Linguistics,* 581–602. Amsterdam/Atlanta: Editions Rodopi B.V.

1999, The monosyllabic roots of Proto-Austronesian. In Paul J.K Li and Elizabeth Zeitoun, eds *Selected papers from the Eighth International Conference on Austronesian Linguistics,* 139–145. Symposium Series of the Institute of Linguistics, Academia Sinica, No.1. Taipei: Academia Sinica,

Final /a/ mutation: a borrowed areal feature in western Austronesian languages

URI TADMOR

1 Introduction

When asked about the differences between Malay and Indonesian, speakers often start by pointing out the difference in the pronunciation of final orthographic <a>: it is pronounced [ɨ] in Malay, but [a] in Indonesian. This is some truth in this simplistic observation; for example, the word <apa> 'what' is realised as [apɨ] in standard Malay, and [apa] in standard Indonesian. However, this represents but a small part of a much larger and more complex phenomenon. In fact, Proto Malayic final **a* has a variety of different reflexes in modern Malay dialects. Moreover, the phenomenon is not limited to Malay; dialects of several other western Austronesian languages, including Javanese, Balinese, and Lampung, exhibit similar phenomena.

What are the geographical and linguistic extents of this phenomenon? Do the changes from **a* into other vowels represent independent innovations within each affected dialect, or perhaps within each language? Or are they interrelated? What was the motivation or trigger for these sound changes? Do they characterise any particular subgroup of Austronesian, or can they be classified as an areal phenomenon? These are some of the questions with which this paper deals.

It should be pointed out from the start that the methodology used was of a non-exhaustive nature. That is, the writer did not randomly collect data from as many as possible of the estimated 1200 Austronesian languages, and see what he came up with. Rather, the starting point was a working hypothesis, whose validity was checked by testing it against evidence from the relevant languages.

2 Final /a/ mutation in Malay

The variety of reflexes of Proto Malayic **-a* in modern varieties of Malay is truly bewildering. Over a dozen phones have been reported, including [a], [ɐ], [ɒ], [ə], [ʌ], [ɯ], [ɨ], [ɑ], [ɔ], [o], [ɛ], [e] and [ɤ]. Indeed, considerable confusion arises from the fact that

John Lynch, ed. *Issues in Austronesian historical phonology*, 15–35.
Canberra: Pacific Linguistics, 2003.

different writers use different symbols to represent the same phone, or use the same symbol for different phones. Moreover, some writers use the symbol [ə] (and the term 'schwa') indiscriminately for any centralised vowel. For these reasons, and in order to avoid getting bogged down in minute but irrelevant phonetic details, it is convenient to group the reflexes of *-*a* under four groups:[1]

1. a-like varieties (e.g. [a], [ɑ])
2. Raised varieties (e.g. [ə], [ɨ])
3. Rounded varieties (e.g. [o], [ɔ])
4. Fronted varieties (e.g. [e], [ɛ])

Table 1 lists a few Malay dialects where these phone groups occur.

Table 1: Final /a/ mutation in Malay dialects

Phone type	Provenance
a-like	Kedah (Malay Peninsula), Brunei (Borneo)
Raised	Johor (Malay Peninsula), Pontianak (Borneo), Tanah Abang (Jakarta, Java)
Rounded	Patani (Malay Peninsula), Palembang (Sumatra)
Fronted	Perak (Malay Peninsula), Jakarta (Java), Sambas (Borneo)

Is the distribution of different final /a/ reflexes completely random, or does it follow a pattern? In order to view things more clearly, reflexes of *-*a* in some Malay dialects were placed on a map (Map 1).

Several generalisations can be made, based on Map 1.

1. The most widespread mutation of /*-a/ is into raised varieties. Such varieties are widely found in all three areas where Malay is spoken natively by ethnic Malays (Sumatra, the Malay Peninsula, and Borneo), as well as on Java, where Malay is spoken natively by members of the Malayicised Betawi ethnic group of Jakarta.
2. The distribution of fronted varieties, while not as widespread as that of raised vowels, also occurs at all four locations.
3. Rounded varieties have a more limited distribution, and have been reported on Sumatra and in the Malay Peninsula, with a possible exception on Borneo.[2]
4. Finally, no final /a/ mutation has been reported in any variety of Malay spoken in eastern Indonesia (east of Borneo and Bali). In all recorded varieties spoken in that region, *-*a* is reflected consistently as [a]. In other words, the application of final /a/ mutation in Malay seems to be restricted to western Indonesia and the Malay peninsula.

---

1 There is, of course, an element of arbitrariness in this division, as a phone may belong to more than one category. Yet overall it was deemed the practical thing to do to facilitate the discussion.

2 The Debak subdialect of Sarawak Malay (Collins 2000) is reported to have final rounding of /a/, but this seems to be a very recent phenomenon, unrelated to the one treated here.

Map 1: The distribution of different reflexes of /*-a/ in Malay dialects

Key: **a**: a-like varieties; **ə**: raised varieties; **o**: rounded varieties; **e**: fronted varieties

3 Final /a/ mutation in other western Austronesian languages

Even if Malay were the only language in the region exhibiting final /a/ mutation, it would be difficult enough to explain. Why would so many different dialects and subdialects independently develop so many different reflexes for *-*a*? However, the situation is further complicated by the fact that the phenomenon also occurs in several other western Austronesian languages. Some examples of the different reflexes of the word for 'five' are listed in Table 2.[3]

Table 2: Reflexes of the word 'five' in dialects of Lampung, Javanese, and Balinese

Lampung		Javanese		Balinese	
Dialect	Form	Dialect	Form	Dialect	Form
Komering	*lima*	Tengger	*lima*	Pedawa	*lima*
Abung	*limo*	Yogyakarta	*limɔ*	Ubud	*limɒ*
Kayu Agung	*limɛ*	Banten	*limʏ*	Denpasar	*limə*

3 Sources are Walker (1976) and Hilman (1994) for Lampung; Smith-Hefner (1983), Horne (1961), and Iskandarwassid et al. (1985) for Javanese; and my own field notes for Balinese.

Again, the question needs to be asked: why would different dialects of various languages, independently of each other, change **-a* to various other phones? Had the changes been confined to contiguous geographical areas — for example, a change to [ɛ] in one area and a change to [ɔ] in another — it may have been easier to explain. Yet there is no possible way, for example, to connect between the origin of [-ɛ] in the varieties of Malay spoken in Perak (on the Malay Peninsula), Jakarta (on Java), Pegagan (on Sumatra), and Sambas (on Borneo). These varieties occur in geographically disparate locations, which were not in contact with each other. So, the mystery here lies not in the fact that final /a/ changed to some other phone, but in that similar changes seem to have taken place independently in dozens of unrelated dialects and subdialects of several languages spoken in one area. A related and equally interesting question is why the phenomenon affected only some dialects of the relevant languages and not others.

Map 2: The area of operation of final /a/ mutation in western Austronesian languages

While it is obvious that the individual reflexes in the many subdialects developed independently, it is equally obvious that it is not a coincidence that **-a* underwent this split in a geographically contiguous area (see Map 2). This is clearly an areal phenomenon. Yet in order to show that a linguistic phenomenon is areal, it is not sufficient merely to point out similarities between geographically close languages or dialects. It must also be demonstrated that the similar phenomena which occur in the relevant languages are indeed related. Moreover, pointing out that a certain feature is 'areal' does not explain much in itself. Linguistic features do not float in the air, as it were, over certain areas, ready to be 'absorbed' into the local languages. For each areal feature, it must be shown that it originated in one particular language, and was transferred into another language at one particular time and under specific circumstances. The transfer of linguistic features is done through contact between languages in the mind of a bilingual speaker, and not, as is often naively thought, merely by virtue of the fact of languages being spoken in the same

geographical area. And while a linguistic feature may occur in many languages, each individual transfer is binary, from a single source to a single recipient. Keeping these facts in mind, we can now examine how final /a/ mutation first arose in western Austronesian languages.

4 The initial step

As we have concluded above, the final /a/ mutation which has affected Malay, Lampung, Javanese, Balinese, and possibly some other languages, must be an areal phenomenon. Yet it is difficult to see how or why [a] would change into so many different reflexes in so many dialects and subdialects. Therefore, it seems reasonable to posit an intermediate stage, in which *-*a* in all the affected dialects changed to a vowel which was phonetically closer to the modern reflexes. For example, it is easier to imagine [a] changing to [ɐ], then to [ə], and from [ə] to either [ɨ], [e], [o], etc., rather than a change directly from [a] to [ɨ], from [a] to [e], or from [a] to [o]. A change from [ə] to [ɨ], or from [ə] to [e], or from [ə] to [o] would be more natural, as the phonetic distance between these vowels and [ə] is smaller. Moreover, this would partially explain how different dialects developed different reflexes of *-*a*. Diagram 1 charts the possible course of change of final */a/ to its modern reflexes in various Malay dialects.

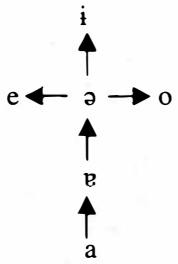


Diagram 1: Phonetic progression of final /a/ mutation

The distributional pattern of fronted varieties in Malay lends some credence to the hypothesis that most reflexes of *-*a* were derived via the intermediate stage of a mid central vowel. In every recorded case of a dialect exhibiting a [ɛ] reflex, there is a nearby dialect exhibiting a [ə] reflex (see Map 1). For example, the Ogan dialect of Malay, spoken in southern Sumatra, has two major subdialects: Ogan Proper and Pegagan. The first has a [ə] reflex, while the second has [ɛ] (Zainul Arifin et al. 1985). On Java, most subdialects of central Betawi (the dialect of Jakarta) exhibit an [ɛ] reflex, but the Tanah Abang subdialect had [ə] (Muhadjir 1981:4; Abdul Chaer 1976:XVIII).[4] This frequent occurrence of raised and fronted varieties in closely related subdialects supports the hypothesis that fronted varieties developed via [ə], and not directly from [a]. In such cases, subdialects which exhibit [ə] are the more conservative ones, and those exhibiting [ɛ] are the more innovative ones.

---

4 During the 1960s and 1970s, most Betawi residents of the city were pushed out to the periphery to make room for new development, resulting in the disappearance of the old Betawi subdialects.

5 The trigger of final /a/ mutation in Javanese

In addition to Malay, another western Austronesian language affected by final /a/ mutation was Javanese. The Old Javanese language was unique in the Austronesian world in the degree to which it was influenced by Sanskrit. According to Blust (1999), '[a]bout half of the vocabulary of Old Javanese texts is of Sanskrit origin'. As Blust pointed out, 'this material clearly reflects the language of the courts and almost certainly would not have been representative of the common people'. As will be claimed below, final /a/ mutation first emerged as a consequence of this wholesale lexical borrowing.

It is true that languages may borrow some words from each other without any affect on their structure. But it would be wellnigh impossible for a language to borrow half of its vocabulary from an unrelated language without any phonological consequences. Moreover, it should be noted that modern Javanese, including the ordinary colloquial speech, still has a very large number of Sanskrit loanwords. This clearly indicates that Sanskrit influence on Javanese, while originating at the courts, eventually reached all speakers.

Of course, it is impossible to say exactly how Sanskrit loanwords in Old Javanese were pronounced. It stands to reason that they were at least partially modified, to suit the phonetics and phonotactics of Javanese. Yet, given the fact that Sanskrit was used extensively in the courts of Java, and that some members of the court were fluent in it, it is also reasonable to assume that the speakers made some effort to pronounce Sanskrit words authentically, as they were pronounced by Indians. It is also quite possible that the source of final /a/ mutation within the courts came from the Indian Brahmans, monks, and priests, who introduced Hinduism and Buddhism to the Javanese, and taught them Sanskrit. One can imagine their Indic-accented, prestigious Javanese being emulated by others in the court, and then disseminated to communities outside the court.

A feature which characterises many Indic languages is the weakening of *ă*. According to Beames (1872:67), 'The short *a* ... is pronounced by the western languages and Hindi — in fact, by all except Bengali and Oriya — as a short dull sound like the final *a* in *Asia*, or that in *woman*'. Actually, the phonological rules under which historical *ă* was changed varied from language to language. For example, in Western Hindi historical *ă* developed into a 'half-open, central, unrounded' sound, similar to the vowel of the English word 'but' (Coulson 1976:5); in final position (where it was never accented) it was deleted, indicating that final /a/ was not only raised but also weakened, until it disappeared altogether. In Sinhalese, historical *ă* was reduced to a mid-central vowel in all open syllables except in word-initial position (Feinstein 1979). The important point is that whatever the original form of this sound change, its origins are ancient, and it has affected the pronunciation of Sanskrit for millennia. In fact, the process was already known to Panini, who composed his great grammar of Sanskrit around 400 BCE (Coulson 1976:5).

The Indianisation of Java did not commence until the first millennium CE (Coedès 1968). Thus there is no doubt that when Sanskrit was first taught to the Javanese, some changes had already affected the original *ă* in Indic languages and in the pronunciation of Sanskrit. In order to understand how the pronunciation of Sanskrit could have influenced the pronunciation of Javanese, it is important to consider two facts. The first is that (modern) Javanese does not have a vowel length distinction. Regarding old Javanese, de Casparis (1975:25–26) stated: 'We do not know whether long vowels were ever pronounced in Old Malay (or Old Javanese), but we know that all the cases in which long vowels are written can be explained in a different manner'. In the case of Old Javanese, long vowel signs were probably used to indicate a sequence of two like vowels, for example <rāh> ([raah], modern

Javanese [rahJ) 'blood', derived from Proto Austronesian *daRaq (with regular deletion of intervocalic /r/). At any rate, even if Javanese did have a vowel-length distinction, it was before the time final /a/ mutation had begun to apply in Javanese.

The second fact that must be kept in mind is that speakers are not normally aware of the etymology of the words in their language. Thus, the average English speakers would not normally realise that the words 'second' and 'language' in the previous sentence were originally borrowed from French, but that the words 'speakers' and 'aware' are of Germanic origin. For the non-linguist, the words 'second' and 'language' are just as English as the words 'speakers' and 'aware'. The same was true for Javanese speakers; at a certain point, the speakers were no longer aware of which words in their vocabulary were originally borrowed from Sanskrit, which were of Austronesian origin. They pronounced all the words in a similar manner.

Taking these two facts into consideration, it is possible to reconstruct the initial stages of final /a/ mutation in Javanese. Since ordinary speakers of Javanese had no way of knowing which words contained an original *ā* and which ones contained an *ă*, final /a/ mutation was generalised to affect originally long final /a/, as well as originally short /a/. Similarly, since the speakers had no way of distinguishing between words of Indic and non-Indic origin, the process spread to all Javanese words, regardless of origin. This process was facilitated by the high prestige of the Indianised pronunciation.

At this stage it necessary to emphasise an important point. The rule affecting the pronunciation of /a/ in Javanese did not apply in exactly the same way as it did in Indic languages or in the pronunciation of Sanskrit by Indians. As mentioned above, even in India itself the rule applied in different ways in the different languages and dialects. The fact that the rule which applied in Javanese was not identical to its counterparts in Indic languages cannot be used as an argument against its ultimate Indic origin. As mentioned above, even among the Indic languages themselves, the details of the *ă*-raising rule varied considerably. Moreover, there is no reason to expect the details of the rule in Javanese to match exactly those of Sanskrit. Thomason and Kaufman (1988:61–62) make this point rather forcefully: 'It has sometimes been claimed that a particular change cannot be due to foreign interference because the putative source language does not exhibit exactly the same structure that has been innovated ... many interference features will in fact *not* be exactly the same as the source-language features that motivated the innovations' (emphasis original). The authors proceed to provide several examples to illustrate this point.

Based on the facts and arguments presented above, it is hypothesised that the court language of Majapahit, under Indic influence, had a rule under which final /a/ was raised. It is interesting to note that Adrian Clynes, whose unpublished thesis was not available to me when I first developed this hypothesis, traced the origin of final /a/ mutation in Balinese (which he referred to as 'R1') to the court language of Majapahit. According to Clynes (1989:159–160):

> Our observations of native speakers from Surabaya, Probolinggo and Madiun in East Java indicate that the East Javanese realisation should probably in most cases be transcribed as [ʌ] (the *unrounded* counterpart of [ɔ]). Soetoko et al. (1984), in their dialect atlas of the area around Surabaya, East Java, close to the site of the former Majapahit capital, record seven words ending in historical /a/. This was realised as [ə] in two items, [kɛʔɔpə] 'why' and [mbəʔtuwə] 'grandmother' *(ibid: 32,* 105). [kɛʔɔpə] was recorded in three villages, [mbəʔtuwə] in four, but in no case did a village show both forms. These would then appear to be 'relict' forms, as yet unaffected by a general change of *a# > ɔ (or ʌ) in this area.

Clynes (1989:160, fn.9) explains that 'many of the forms transcribed in Soetoko et al. as [ɔ] should in fact be [ʌ]. [ʌ] is not used in that work, even though it is clearly common in the area'. Furthermore, he stresses that there are probably more realisations of final [ʌ] than recorded by Soetoko et al. (1989:159, fn.8).

It is well known that extensive lexical borrowing from literary languages may result in some structural interference as well. There are numerous documented cases in the literature. An interesting example is the impact of Arabic on languages of non-Arab Muslim speech communities. Languages like Turkish, Farsi, Urdu, and Malay borrowed heavily from literary Arabic, and this lexical borrowing impacted their respective phonological structures. The phonological interference was no doubt aided by the fact that Arabic is used not only as a written language in these communities, but also as a liturgical language. The more educated (or religious) the speaker, the more authentic his pronunciation of Arabic (and subsequently of Arabic loanwords). The more authentically Arabic loanwords are pronounced, the greater the probability of phonological interference from Arabic upon the vernacular. Through prestige, this interference then gradually percolates down to less educated or less religious members of the community.

An even closer parallel comes from the Dravidian languages of southern India. Dravidian languages borrowed heavily from Sanskrit, and the originally purely lexical borrowing eventually had an impact on their respective structures, including their phonologies. Based on Sridhar (1978:202–206), Thomason and Kaufman (1988:79) report that 'Brahman Dravidian speakers avoid making the usual Dravidian substitutions to nativize Sanskrit loanwords, such as deaspiration of aspirated stops and voicing of intervocalic obstruents; instead, they retain the Sanskrit pronunciations in loanwords and even extend some of them, for example, aspirating stops in borrowed Sanskrit words that did not originally have aspirated stops ...'.

In the case of Javanese-speaking Brahmans of Indian origin, the overgeneralisation of Sanskrit pronunciation would be even more expected than that of Dravidian-speaking Brahmans, since they were not native speakers of Javanese. This Sanskritised pronunciation of Javanese would then be emulated by non-Indian Brahmans, other members in the court, and eventually by members of the general speech community as well.

6 The transfer of final /a/ mutation to other western Austronesian languages

As proposed above, the phenomenon of final /a/ mutation had its beginnings in the Sanskritised court language of the Javanese empire of Majapahit. The Majapahit empire, which was founded towards the end of the 13th century, was a major vehicle for the spread of Indianisation in the Austronesian-speaking world. The empire reached its zenith during the reign of Rejasanegara, better known as Hayam Wuruk (1350–89), and his *patih* ('prime minister') Gajah Mada. A Javanese chronicle of that period, the Negarakertagama, places most of the territory of modern-day Indonesia (with the possible exception of northern Sulawesi — see Coedès 1968:239–240), as well as much of the Malay peninsula, under the rule of Majapahit. In some areas, Majapahit's rule was probably indirect at most, and limited to paying an annual tribute. Yet in other places, Javanese colonies were established, and Javanised courts installed. It was in this period that final /a/ mutation was probably spread to areas outside Java which were under Javanese rule or influence. Strong evidence

for Javanese linguistic influence exists in the form of numerous Javanese loanwords in the languages of Bali, Lombok, southern Sumatra, the Malay peninsula, and southern Borneo.

It would not be possible, of course, to provide a precise chronology of the emergence and spread of final /a/ mutation in all the affected western Austronesian languages. However, it is possible to give a rough framework, showing the approximate time the change could have spread to the various languages which were affected by it.

It is widely known that after the final fall of Majapahit in the 1520s, its royal family and nobility fled to Bali (Hall 1992:227). What is less known is that by that time Bali was already thoroughly Javanised. Coedès (1968:234) notes that 'on Bali ... an expedition of 1343 led to the destruction of the local princely family and a Javanisation of the island that was to be intensified during the reign of Hayam Wuruk'. This was nearly two centuries before the fall of Majapahit, and final /a/ mutation may have taken root in Bali already then.

Likewise, Indo-Javanese influence on Sumatra is ancient (Marwati & Nugroho 1990b: 23–24), and the kingdoms of Malayu (Jambi) and Palembang were under direct Javanese rule for long periods throughout the second millennium CE. Coedès (1968:201) cites '[t]angible proof of the ascendancy of Java over Sumatra' in the 13th century. Javanese influence was perhaps strongest in Palembang, whose Malay dialect borrowed heavily from Javanese (Tadmor forthcoming), and which was ruled by a Javanese-speaking royal house and elite. Javanese influence in Palembang was probably strongest between the mid 16th and mid 17th centuries (Djohan 1995:155), and final /a/ mutation in the local Malay dialects may well have originated then.

Further south, Lampung also came under Javanese influence repeatedly and for extended periods, beginning with the Majapahit period. The transfer of final /a/ mutation to the Lampung language more probably took place later, when Lampung was under direct Javanese rule, as a vassal of the Javanese kingdom of Banten (Marwati & Nugroho 1990b:37; Watson Andaya 1992a:431).[5]

In the Malay peninsula, Javanese influence was also very strong. The court of Malacca, the most important Malay kingdom of the 15th century, was 'thoroughly Javanized' (Clynes (1989:162). The same was true for the court of the most important Malay kingdom on the east coast of the peninsula, Patani. Teeuw and Wyatt (1970:239) discuss the important role of the Javanese in Patani. The number of Javanese slaves there was so great that in 1613 they staged a failed rebellion (Teeuw & Wyatt 1970:251, fn.10). But Javanese influence was also important in the court (1970:239). Peter Floris, who stayed in Patani in 1612–13, attended 'a commedye ... to the manner of Java' performed at the court (1970:257). In fact, traditional Javanese arts such as *wayang kulit* and batik survived in Patani until the 20th century. Singapore, at the southern end of the peninsula, was also under strong Javanese influence, if not direct rule. The old name of Singapore, Tumasik (Temasik), is also of pure Javanese etymology.[6] The famous Singapore inscription, deliberately destroyed by the British authorities, was written in Old Javanese, based on contemporary descriptions, as well as an analysis of surviving fragments done by de Casparis (1975:45).

The very early and very strong Javanese presence and influence in south-eastern Borneo is discussed in detail in Ras 1968 (especially pp.182–200). The linguistic evidence of this presence are still obvious today in the form of a significant number of Javanese loanwords in languages of the area, especially Banjarese, the regional Malay-based koine. Javanese

---

5 'Daerah pengaruh Banten di luar Jawa Barat adalah daerah Lampung yang sejak Hasanuddin telah berada di bawah kekuasanan Banten.'

6 Derived from *tasik* 'sea' and the infix *-um-*.

loanwords in Banjarese include numerous core vocabulary items, as well as several archaisms (Wolff 1988:89). Interestingly, Banjarese itself does not exhibit final /a/ mutation; this may have been due to the fact that it originated as a contact language between speakers of Javanese, Malay, and local Dayak languages.[7] As mentioned above, Malay-based contact languages such as Bazaar Malay and the Malay-based creoles of eastern Indonesia do not exhibit final /a/ mutation. However, other south Borneo languages do exhibit the phenomenon, including Ngaju Dayak, Maanyan, and Malagasy (Adelaar 2000:14). Banjarese loanwords in Ngaju reflect *-*a* as [a] (Dyen 1956), indicating the change in Ngaju was relatively early and had ceased to operate by the time the Banjarese words were borrowed.

The presence of final /a/ mutation in Malagasy may appear problematic, because of the assumed early migration date of the Malagasy to Madagascar. However, Dahl (1951:96–105) has documented the presence of 30 Sanskrit loanwords in Malagasy, clear evidence of at some Indic linguistic influence. Adelaar (1989:32) documented very numerous Malay and Javanese loanwords, which demonstrate that Malay and Javanese 'had a strong influence on Malagasy'. Moreover, Adelaar (1989:32) has shown that all but one of the Sanskrit loanwords in Malagasy are also present in Malay or Javanese, so their presence in Malagasy is most probably due to secondary borrowing, rather than direct Indic influence. Therefore, he concluded that the migration of the Malagasy 'must have take place at a period when the Malays had already undergone Indian influence'. Final /a/ mutation in Malagasy thus could be plausibly due to transfer from Javanese or Malay.

In western Borneo, final /a/ mutation has two possible immediate sources. Java maintained close contacts with western Borneo, and controlled the major cities of western and southern Borneo in the 16th and early 17th centuries (Marwati and Nugroho 1990a:147).[8] The linguistic impact of this occupation has been long-lasting. For example, in the Javanised court of Sambas, a special vocabulary was used to address or speak about members of the royal family (Mustapa et al. 1984:12). Not only was the institution of a special vocabulary itself patterned after Javanese *krama*, but the words themselves were often taken from Javanese. Thus, for example, the personal pronouns /kolɛ/ '1st person singular' and /nikkɛ/ '2nd person singular' were borrowed from Javanese. The presence of final /a/ mutation is Sambas Malay may therefore be attributed to a transfer from Javanese. Further south, in Pontianak, *-*a* is realised as [ə]. Here, the source of the phenomenon may have been Riau Malay, which is the ancestor dialect of Pontianak Malay, or at least heavily influenced it (Adelaar 2000:18). In Riau Malay, *-*a* is realised as [ə] or [i]. Final /a/ mutation did not spread far along the Kapuas river; in the next dialect upriver from Pontianak, that of Sanggau, *-*a* is realised as [a], as it is in the further upriver dialects of Sintang and Putussibau. It did, however, spread southward along the coast to Ketapang, and from there inland with migrating Malays along several rivers.

In Java, the only area where final /a/ mutation operated was Jakarta, among the Malayicised Betawi ethnic group. In the Tanah Abang subdialect, final /a/ was reportedly realised as [ə] (Abdul Chaer 1976:XVIII; Muhadjir 1981:4), although people who have heard this subdialect (it is now extinct) report that the sound was closer to the [ɑ] now heard in

---

7 Ras (1968:8) describes Banjarese, impressionistically yet insightfully, as 'the independent continuation of a rather archaic type of Malay, superimposed on a substratum of Dajak dialects, with an admixture of Javanese'.

8 'Begitu penting kota-kota Kalimantan ini bagi pesisir Utara Jawa sehingga beberapa kali dikirim ekspedisi untuk mendudukinya pada abad ke-16 dan awal abad ke-17. Sumber-sumber dari masa kemudian lebih banyak menyebut hubungan dengan Sambas, Banjarmasin, dan Sukadana ...'.

central Balinese. The influence of Balinese on Betawi has been substantial (Grijns 1991) and long known (van der Tuuk 1867). This is not surprising, in view of the fact that the original ancestors of the Betawi were mostly Balinese slaves. (In the census of 1819, the Balinese were recorded as the largest group both among the slaves and among the free indigenous inhabitants of Batavia; see Ikranagara 1980:2). So final /a/ mutation in Jakarta can be traced to the Balinese slaves who were brought there by the Dutch during the 18th century. The other subdialects of central Betawi show [ɛ], a natural development from a schwa-like sound (see above).

In the Sasak language, spoken on the island of Lombok, final /a/ mutation may have had a multiple origin. The island came under direct Balinese rule in the 16th century (Watson Andaya 1992a:424; Watson Andaya 1992b:526), and Balinese influence there has continued ever since. In addition, Javanese has long been used as a literary language on Lombok. Since both Balinese and Javanese were affected by final /a/ mutation, it is not surprising that this feature has been transferred to Sasak.

Finally, it is interesting to note that Clynes (1989) proposed a connection between court language, the presence of speech styles, and final /a/ mutation, which he referred to as 'R1' ('Rule 1'). Clynes (1989:163) also hypothesised that 'the similar realisations of word-final historical /a/ in Balinese and Malay dialects derive from a common source, the Javanese of the late 15th century'. (As mentioned above, Clynes's thesis was not available to me when I first developed my hypothesis regarding the emergence and spread of final mutation.)[9]

7 The spread of final /a/ mutation within Javanese

As we have seen above, final /a/ mutation spread from Javanese to other languages in the region which came under strong Indo-Javanese influence. It is also interesting to examine how final /a/ mutation spread within Javanese itself. The original Majapahit court was located in the Brantas river basin of eastern Java. When Javanese courts were established in central and western Java, the court language of Majapahit was emulated in these newer courts, including the feature of final /a/ mutation. The feature then spread to areas surrounding the courts, though never reaching the periphery.

In the central Javanese court cities of Yogyakarta and Solo (Surakarta), the modern realisation of final orthographic <a> is /ɔ/. This rounding is apparently of rather recent origin, and has not reached all parts of central Java. In the early 19th century, Raffles (1817:359) reported that the realisation of <a> (he did not specify the phonological conditioning) was 'that of *a* in "water", or of *o* in "homo"; the *o* being at present invariably used at the native courts and their vicinity for the inherent vowel of the consonant, instead of *a*.' An unrounded schwa-like sound is still used in some dialects of central Javanese spoken in areas more distant from the courts or major cities. Nothofer (1982:294) mentions an [ɤ] variety; my data for the Pemalang dialect (see final paragraph, this section) show [ə].

Some Javanese dialects preserve the original [a] sound, and interestingly these dialects are spoken precisely in the areas least affected by Indianised Javanese court culture: Tegal in the west and Tengger in the east. Both the Tegal and Tengger dialects also share another feature: neither of them has speech levels, another feature associated with courts of Java. In addition, the Tengger people are also culturally unique by being the only Javanese group that has not converted to Islam. To this day, the Tengger maintain their traditional belief system, in

---

9 I am very grateful to Prof. Clynes, who was kind enough to make his thesis available to me in response a query I had placed on the Linguist Internet discussion list.

almost total separation from mainstream Javanese culture. The distribution of final /a/ mutation within Javanese, then, clearly demonstrates its close relationship to Indianised court Javanese.

An even more striking example, lending further support to the hypothesis that final /a/ mutation originated in court Javanese, comes from the distant and isolated dialect of Banten. This Javanese dialect is spoken in an area completely surrounded by Sundanese-speaking areas, hundreds of kilometers away from the main Javanese-speaking communities of central and east Java. In Banten Javanese, final /a/ is either maintained as [a], or realised as [ɤ]. The choice of vowel is not random, and determined geographically. Iskandarwassid et al. (1985:20) report that 'the [ɤ] variant is found ... [in] the areas which were in the vicinity of the Banten court in the past',[10] while 'the [a] variant is found in the speech communities... which were rather distant from the old court center'.[11] Banten Javanese thus provides clear evidence linking final /a/ mutation to the court language of Java. Another interesting feature of Banten Javanese is that because of its isolation from the bulk of Javanese-speakers, and its small speech community, /a/ mutation there is preserved in its original form, as reconstructed above for the court language of Majapahit. The processes of rounding (discussed above) and umlauting (see below), which originated in the courts of central Java at a relatively late date, never affected the court language of Banten, which had split off earlier.

As for the main body of Javanese speakers, Clynes (1989:160) makes this interesting observation: '... word-final historical /a/ is variously realised as [ʌ], [ɔ], or [ə] in Central and East Java, with forms with [ə] and [ʌ] being recorded in the area close to the former Majapahit capital'. This constitutes further evidence that /a/ mutation originated in the Majapahit court language, and that the original mutation was to a schwa-like sound.

In addition to rounding, another change that affected the court language of central Java was an umlaut-like assimilatory process, by which /a/ mutation spread from final syllables to open penultimate syllables. For example, <cara> 'manner' is realised as [cɔrɔ] in modern standard Javanese, and <agama> 'religion' is realised as [agɔmɔ]. This umlauting is a relatively recent change, as indicated by the fact that it does apply in all modern central Javanese dialects, let alone the western dialect of Banten (see above) or languages outside Java. Nothofer (1982:294), in a dialect map of central Javanese dialects, records the following forms as reflexes of **mata* 'eye': [mata], [mato], [matɤ], and [mɔtɔ]. The form [mata] exhibits no final /a/ mutation at all. The form [mato] constitutes evidence that the rounding and umlauting of final /a/ mutation were parts of two different process, as it shows rounding without umlauting. The third form ([matɤ]) is the most conservative: it exhibits final /a/ mutation, but with neither rounding nor umlauting. The last form ([mɔtɔ]) is the standard Javanese form,[12] which originated in the courts of Yogyakarta and Solo and their immediate surroundings, but is now used over an extensive area of central Java. This form is the most innovative, as it exhibits final /a/ mutation as well as the subsequent processes of rounding and umlauting.

To complete the picture, the Javanese dialect of Pemalang provides some interesting data. This dialect constitutes a transitional area between central Javanese dialects, where /a/

---

10 'Variasi [ɤ] terdapat ... [di] daerah yang berdekatan dengan lingkungan Keraton Banten pada masa lampau.'

11 'Variasi [a] ditemukan dalam lingkungan penutur di daerah-daerah ... yang letaknya agak jauh dari pusat keraton dulu.'

12 In modern standard Javanese, the word <mata> is considered crude and suitable more for animals than for humans. In reference to humans, it has been replaced by <mripat>, originally a *krama* (honorific) term.

mutation applied to the entire lexicon, and western Javanese dialects (like Tegal and Brebes), where it did not apply at all. In Pemalang, final /a/ mutation occurs in certain lexical items but not in others, without any apparent phonological or other conditioning factors. It seems that the change spread by lexical diffusion, and had only affected a part of the lexicon when it ceased operating. Pemalang is also conservative in that it has not been affected by rounding; in other words, it preserves a sound close to the original result of final /a/ mutation at the court of Majapahit. On the other hand, those forms affected by final /a/ mutation have undergone umlauting when applicable, thus completing the paradigm of forms collected by Nothofer. Thus the reflex of **mata* 'eye' in Pemalang is [mata], with no final /a/ mutation, but the reflex for **gawa* 'bring' is [gəwə], reflecting final /a/ mutation (and subsequent umlauting). Forms like [bisə] 'able to' and [critə] 'story' on the one hand, and [ana] 'there is' and [ləŋa] 'oil' on the other hand, indicated that we are not dealing with a phonetically motivated change that affected vowels following voiced consonants, as is the case with Madurese (see §8 below).[13]

8 Excluding unrelated phenomena

In describing an areal phenomenon, it is important not only to determine which features of which languages were affected by the change, but also to determine which phenomena are unrelated, even if they exhibit superficial similarities. For example, tonogenesis (the emergence of lexical tone) took place in the histories of many contiguous yet unrelated languages in southeast Asia, belonging to the Sino-Tibetan, Austroasiatic, Miao-Yao, Tai-Kadai, and Austronesian language families. It is accepted by virtually all historical linguists of these languages of the region that this is not a coincidence, and that tone is an areal feature of southeast Asian languages. Yet no linguist, to my knowledge, has ever made the claim that tonogenesis in Africa — or even in closer areas, such as New Guinea — is related to tonogenesis in southeast Asia.

In our case, it is certainly not claimed that any change which affected final /a/ in any language is part of the areal phenomenon described here. To illustrate this point, let us consider the case of the Gorontalo language, spoken in northern Sulawesi.[14] In Gorontalo, final /a/ has changed to /o/ in most inherited words. For example, the Gorontalo reflexes of the Proto Austronesian words **maCa* 'eye' and **lima* 'five' are *mato* and *limo* respectively. However, the change did not affect final /a/ in words of ultimate Sanskrit origin, such as *nyata* 'obvious', *kira* 'be of the opinion', *yuta* 'million', *jasa* 'public service', *jiwa* 'spirit', *gowa* 'cave' (examples are from Mansoer 1977).[15] Moreover, there is no evidence that the change /a/ > /o/ went through an intermediate schwa stage. Gorontalo also does not share other features of the languages involved in the areal phenomenon, such as lack of word accent or iambic rhythm (see §10 below). In short, the change *a* > *o*/_# in Gorontalo appears to be totally unrelated to the areal phenomenon discussed in this paper. Indeed, Gorontalo is geographically very distant from the area of operation of the phenomenon discussed in this paper, and has never come under Indo-Javanese influence.

A more difficult yet very illustrative case is that of Madurese. The island of Madura is situated just off the north-east coast of Java, and its language and culture have come under

---

13 I am grateful to Heri Tanujaya, a native speaker of the Pemalang dialect, for these interesting examples.

14 I am grateful to Robert Blust for bringing the case of Gorontalo to my attention.

15 These words were apparently all borrowed via Malay/Indonesian.

heavy Indo-Javanese influence. One might therefore expect to find final /a/ mutation in Madurese. Some Madurese words[16] indeed appear to exhibit final /a/ mutation, in the form of a [ə] reflex of **-a*. Some words also exhibit [ə] as a reflex of **a* in penultimate syllables, thus resembling the Javanese umlauting process described in §7. However, a close examination of the data reveals that the similarity is superficial, and that these vowel changes in Madurese are part of an internal, phonetically motivated process, unrelated to the one with which we are dealing.

The situation in Madurese is partially obscured by two mirror-image processes, one which deleted original final /h/, the other which inserted /h/ at the ends of words which originally ended in vowels. Thus the words [jilə] 'tongue' and [bəbə] 'under' are historically derived from **jilah* and **babah* respectively, while [rajəh] 'big' and [dədəh] 'chest' are derived from **raya* and **dada* respectively. Despite the confusion caused by the application of these mirror-image rules, an analysis of the data yields an important fact: the change of final /a/ to [ə] occurred only after voiced oral consonants (/b, d, g, j/). It did not occur after voiceless stops or after (voiced) nasals. Moreover, the change applied regardless of the position of /a/ in the root, that is in non-final position as well as final ones (even if the final vowel was not /a/). For example, the change did not affect the original final /a/ in the words *apah* 'what' (< **apa*) and *matah* 'eye' (< **mata*), yet it did affect the /a/ in the words [bəgus] 'good' (< **bagus*), [gəli] 'hard' (< **gali*), and [jəriŋ] 'net' (< **jariŋ*), even though the original [a] was not in final position, nor did the word end in /a/. The vowel change after historical voiced oral consonants in Madurese was, in fact, related to the breathy phonation of syllables commencing in voiced consonants, and had nothing to do with final /a/ mutation.

The examples of Gorontalo and Madurese demonstrate the degree of thoroughness which must be striven for when analysing an areal phenomenon. In order to establish that a certain feature is areal, it is not sufficient merely to point out similar features in two (or more) languages. It must also be demonstrated that the features can be related linguistically, and that the transfer of the feature would have been historically and geographically plausible. In the cases cited above, Madurese fails the linguistic criterion, while Gorontalo fails the historical and geographical criteria as well as the linguistic criterion. Moreover, areal phenomena are rarely limited to just one feature; it is more common for several common features to link a group of languages into a *Sprachbund*. Indeed, as shown in §10 below, the languages which underwent final /a/ mutation also share other features, such as iambic rhythm and lack of word-level accent (as well the presence of numerous Indic loanwords). A longer list of phonological features which link western Austronesian languages into a larger *Sprachbund* was discussed in Tadmor (2001).

9 The correlation between final /a/ mutation and Indo-Javanese culture

In earlier parts of this paper I have made the claim that the process of final /a/ mutation in dialects of some western Austronesian languages originated in the Indianised speech of the courts of Java. It would therefore be useful to view the evidence together.

Consider the following facts:

---

16 I am grateful to Faisol Riza, a native speaker of Madurese, for the data and to Antonia Soriente for her help in the analysis of the data.

(a) In western Borneo, the major Malay dialects (Pontianak, Ketapang, Sambas), which form part of Indianised Malay civilisation, underwent final /a/ mutation. Significantly, all three dialects emerged in states with Indianised or Javanised courts.

(b) In south-central Sumatra, which has undergone very heavy Indo-Javanese influence, the only Malay dialects which appear to maintain *-*a* unchanged in some contexts are tribal groups which are not part of (Muslim) Indianised Malay civilisation (David Gil, pers. comm.).

(c) In the Malay varieties of east Indonesia (such as those spoken in Manado, Makassar, Ternate, Ambon, Kupang, and Larantuka), final -*a* is maintained. This fact is important because these varieties differ from the Malay dialects of western Indonesia in several crucial factors, internal as well as external. Unlike true Malay dialects, these Malay-based creoles all have word accent (sometimes phonemic; see Tadmor 2000) and a trochaic (rather than iambic) rhythm, and were not spoken in areas which were part of the Majapahit empire, or underwent extensive Indianisation (see §10).

(d) In the Malay varieties spoken on Java, -*a* is maintained by Peranakan Chinese, but not by the Betawi of Jakarta, who are culturally (as well as linguistically) part of Malay civilisation.

(e) None of the varieties of Bazaar Malay known to the writer exhibit final /a/ mutation. Bazaar Malay (which has many varieties) served as an interlanguage between native speakers of different languages, and especially in contacts between native Indonesians and speakers of Chinese and European languages. It is widely considered to be the 'least refined' form of Malay. It is therefore not surprising that it has not been influenced by the language varieties which form the opposite end of the spectrum, viz. the refined language of the Malay courts.

The fact that standard Indonesian does not exhibit final /a/ mutation is related to the fact that it is not the home language of any speech community or the first language of any speaker. Rather, it is used exclusively as a second language or dialect. This is a feature it shares with the Malay-based creoles (which were initially used purely as a second language) and with Bazaar Malay (which is still only used as a second language). Colloquial versions of Indonesian, which have been developing in urban centres throughout Indonesia, also do not exhibit final -*a* mutation, for the same reason.

Outside the Malay-speaking area, the correlation between these linguistic and cultural factors is just as remarkable.

(a) In the Javanese speech island of Banten, final -*a* changed to [ɤ] around the location of the old court, but was retained as [a] in more distant areas.

(b) In the Tengger speech community of eastern Java, a group which has been least affected by Indo-Javanese court culture, final -*a* is maintained.

(c) The Tegal dialect of central Java, situated furthest away from the courts of Yogyakarta and Solo, and least affected by their culture, has maintained final -*a*.

(d) In Bali, the indigenous Bali Aga, the only group not to have been incorporated into the Javanised-Indianised culture of the Balinese kingdoms, maintain final -*a*. They are the only group to maintain the religious beliefs and other traditions which presumably prevailed in Bali before the Javanisation of the island.

(e) On Lombok, the Sasak language, which was heavily influence by Balinese, and whose speakers use Javanese as their literary court language, exhibits final /a/ mutation.

There is hardly need to reemphasise that all these facts could not be due to coincidence. The close relationship between final /a/ mutation and Indianisation, as well as the direct link between Javanese court language and the spread of final /a/ mutation, are obvious.

10 Possible other causes for final /a/ mutation

Phonological change can be brought about by many factors. First, there are internal factors, such as analogy, levelling, systemic pressure, and various natural processes which can be grouped together under 'phonetic motivation'. Then there are external causes attributable to language contact, such as substratum influence and borrowing. 'Sporadic' (that is, unmotivated) change is also mentioned in the literature, although it is debatable whether language change can occur spontaneously, without any internal or external triggers. In fact, 'sporadic change' does not explain anything, and is sometimes used in lieu of an explanation when the real cause or trigger for change cannot be found.

Finally, there is also the possibility of multiple causation. Some linguists believe that if an internal explanation is available, that automatically excludes an external one. Yet often internal factors interact with external ones to produce change. A well-known example is the phonemicisation of voiced fricatives in English, which was caused by an influx of French loanwords containing voiced fricatives into English, as well as by some internal factors (Thomason & Kaufman 1988:61). It is therefore possible that in addition to the external factors discussed in this study, internal factors also played a role in the process of final /a/ mutation in western Austronesian languages.

Final /a/ mutation is reminiscent of accent-related processes in languages where word accent plays an important role, such as Germanic languages. One may thus put forward the hypothesis that final syllables in the relevant languages were unaccented, and subsequently final /a/ was reduced to schwa in such syllables. Later — so the hypothesis would continue — an accent shift to the last syllable took place, resulting in the 'restrengthening' of schwa to a 'full' vowel in the relevant languages or dialects. This possibility was considered at length by Tadmor (1997). However, as pointed out there, in Malay/Indonesian — at least in standard varieties of the languages and dialects spoken by ethnic Malays — there is no fixed word accent, hard as this may be for Eurocentric linguists to accept.[17] This has been convincingly shown by several phoneticians, including Odé (1994), van Heuven and van Zanten (1994), and Yong (1998).[18] In fact, none of the languages or dialects affected by final /a/ mutation have fixed word accent. Moreover, the Malay varieties which do have a fixed word accent, such as the Malay-based creoles of eastern Indonesia, do not exhibit any final /a/ mutation at all. Final /a/ mutation and the presence of word accent therefore appear to be mutually exclusive.

The rhythm of Malay, as well as that of the other languages affected by final /a/ mutation, is iambic. Rather than having a word-level accent, there is a tendency for accent to occur on the last syllable of the phonological phrase (Tadmor 2000). There appears to be no reason to believe that when final /a/ mutation started to operate in these languages, the

---

17 Practically all grammars of Malay/Indonesian wrongly state that the language does have word accent, although there is no agreement about the position of this alleged word accent (for a full discussion, see Tadmor 2000).

18 My Russian colleagues assure me that the fact that Indonesian does not have penult stress or accent has been common knowledge among Russian scholars since the 1960s.

situation was much different from what it is today.[19] Therefore, it is difficult to view the original change (of /a/ to a higher or more centralised vowel) as an accentuation-related process.

Indeed, it appears that the various 'full-vowel' reflexes of *-*a* emerged as a restrengthening of a centralised vowel of some sort. Such a centralised ('reduced' or 'weakened') vowel would be unnatural in a syllable that was often accented. Thus, it was restrengthened independently in many dialects. This led to the emergence of multiple reflexes of *-*a*, as each dialect changed the reduced vowel into a different full vowel. Yet the question remains: if the final syllable was accentuable, why was final /a/ weakened to a schwa-like sound there in the first place? Vowel weakening in an accented (or accentuable) syllable would be highly unusual. The answer is that the initial change of *-*a* to a centralised vowel was not a phonological process of weakening or reduction. Rather, it was a borrowed feature, which lacked phonetic motivation.

11 A similar case: the spread of uvular /r/ in Europe

In order to put the spread of final /a/ mutation in western Austronesian languages in perspective, it may be useful to compare it to a similar and well-known phenomenon, the spread of uvular *r* in western Europe. Trudgill (1974:160–163) provides a clear and concise description of this process:

> It is thought that up until at least the sixteenth century all European languages had an *r*-type sound which was pronounced as *r* still is pronounced today in many types of Scots English or Italian: a tongue-tip trill (roll) or flap. At some stage, though, perhaps in the seventeenth century, a new pronunciation of *r* became fashionable in upper-class Parisian French ... Starting from this limited social and geographical base, the uvular-*r* pronunciation has during the last 300 years spread, regardless of language boundaries, to many other parts of Europe ... It is now used by the overwhelming majority of urban or educated French speakers, and by most educated Germans. Some Dutch speakers use it, as do nearly all Danes, together with a majority in the south of Sweden and parts of the south and west of Norway.

There are several similarities between the spread of uvular *r* in European languages, and the spread of final /a/ mutation in western Austronesian languages. Both processes lack phonetic motivation; both started with a very limited and well-defined social and geographical distribution; both spread due their high prestige; both operated in a contiguous geographical area, rather than a genetic subgroup; both applied to only certain dialects of the affected languages and not others. Indeed, it would not be unreasonable to assume that the ultimate source of the uvular realisation of *r* was in the French court, which in the 17th century was a source of emulation throughout Europe, just as the ultimate source of final /a/ mutation in the Malay archipelago was in the Majapahit court.[20]

---

19 It has sometimes been claimed that the long vowel symbols used (inconsistently) in Old Malay and early Classical Malay in penultimate syllables are in fact an indication of penultimate word accent. In Tadmor (2000) I discussed this theory, and concluded that the language represented in the Old Malay inscriptions and Classical Malay manuscripts was not representative of the natural spoken language. Rather it was an artificial court language, replete with Sanskrit (and later Arabic) loanwords, whose writing system was devised by foreigners, and whose pronunciation was affected by them. It did not reflect the prosodic structure of the colloquial language as spoken by native speakers.

20 This would also be reminiscent of the change /z/ > /θ/ in Castilian Spanish, which also originated in the 17th century. Legend has it that the ultimate origin of the change was a lisping prince. This detail may not

As Trudgill (1974:161) himself pointed out, the change from an alveolar *r* to a uvular or velar *r* in Europe is not limited to the contiguous area described above. Phonetic realisations of /r/ vary greatly in languages the world, and changes from one realisation to another have occurred unrelatedly outside the limited area of operation of this areal phenomenon. For example, uvular *r* is used in parts of Northumberland and Durham (1974:161), yet this phenomenon is apparently not related to the spread of uvular *r* on the European mainland. Likewise, there are other Austronesian languages where final /a/ has undergone changes (see §8), independently from the final /a/ mutation treated here. As Thomason and Kaufman (1988:59) noted: 'If a reasonable external explanation for a change is available, it must not be rejected merely because similar changes have occurred under different antecedent conditions'. They also dismiss the untenable view that 'a plausible internal motivation is preferable to, as well as exclusive of, an external motivation' (1988:43).

Obviously, the changes to uvular *r* in dialects of several languages spoken in a contiguous area of western Europe are interrelated. Yet not all similar changes anywhere else in Europe are related to this areal phenomenon. The same can be said for final /a/ mutation. Obviously, the changes that affected final /a/ in dialects of several languages spoken in a contiguous area of the Austronesian-speaking world are interrelated. Yet not all similar changes anywhere else in the region are related to this areal phenomenon.

12 Conclusions

In this paper I have shown that an areal feature of Indic languages, the raising of *ă*, was transferred to some western Austronesian languages, where it developed into the phenomenon which I call final /a/ mutation. Specifically, I have demonstrated that:

- The changes that affected final /a/ in many dialects of Malay, and resulted in a large variety of modern reflexes, were interrelated.
- In turn, the phenomenon in Malay is related to similar phenomena in some other languages spoken in contiguous areas, such as Javanese, Balinese, Lampung, and Sasak.
- The languages affected by the change were under strong Indo-Javanese influence.
- It is difficult to show a phonetic motivation for this sound change. Neither systemic pressure nor accent-related phenomena can be convincingly demonstrated. And even if internal causation is eventually demonstrated, that would not exclude the role of external causation.
- The isolects affected by final /a/ mutation exhibit geographical proximity or contiguity, suggesting that this is an areal phenomenon.

The linguistic, geographical, and historical origin and progression of final /a/ mutation can be summarised as follows:

- Short /a/ was realised as a schwa-like vowel in final syllables of Sanskrit loanwords in the Indianised court language of Majapahit, in imitation of Indian pronunciation.
- Because of the confusion (or indeed lack of distinction) between short and long vowels in Javanese, the rule affecting the pronunciation of /a/ applied regardless of historic length distinction.

---

be true, but it is entirely possible that the change indeed originated in the Castilian court, whence it spread to surrounding areas, eventually encompassing most of Castile.

- Due to the large number of Sanskrit loanwords in Javanese, the prestige of Indianised pronunciation, and the ignorance of most speakers regarding the historical origin of different words, the rule was generalised to include words of non-Sanskrit origin.
- The change spread from the Majapahit court to later Javanese courts, and from these courts to the surrounding areas, eventually reaching most Javanese-speaking areas.
- The change was also transferred from Javanese to some dialects of languages spoken in areas under Javanese control or Indo-Javanese cultural influence, such as Lampung, Balinese, Sasak, and Malay.
- The distribution patterns of final /a/ mutation in Bali and in the Malay-speaking world were similar to those of Javanese: the communities which were closer to the political centres of power, and thus most affected by Indianisation, underwent final /a/ mutation, while more isolated communities (geographically and/or culturally) did not.
- Since the speaking rhythm of the affected languages was (as it still is) iambic, the schwa-like final vowel was independently restrengthened to a full vowel in many isolects. This is the reason behind the great variation of **-a* reflexes in modern languages and dialects. In more conservative dialects, a schwa-like sound has been preserved until today, but it often has a different phonetic realisation from the 'true' phonological schwa (if one indeed exists in the relevant language).
- An umlaut-like process extended /a/ mutation to open penultimate syllables in central Javanese. This process had only a limited geographical distribution.
- Later, central Javanese acquired its characteristic rounding of mutated historical /a/. This, too, has had only a limited distribution.

The scenario presented above is tentative, of course. It is quite possible that when more information becomes available — for example about the history and nature of the original phenomenon in India, or about the provenance of the Indians who introduced Sanskrit to Indonesia — some of the details will need to be changed. However, I believe that the main premises of the hypothesis presented here will stand: that final /a/ mutation is an areal phenomenon, that it was related to Indianisation, and that it originated in the courts of Java.

References

Abdul Chaer, 1976, *Kamus Dialek Melayu Jakarta – Bahasa Indonesia.* Ende: Nusa Indah.

Adelaar, K.A., 1989, Malay influence on Malagasy: linguistic and culture-historical Implications. *Oceanic Linguistics* 28/1:1–46.

2000, Borneo as the homeland of the Malay language: fifteen years down the road. Paper presented to the colloquium Borneo as the Homeland of Malay: the Implications for Research, Universiti Kebangsaan Malaysia, Bangi.

Beames, John, 1872, *A comparative grammar of the modern Aryan languages of India*, vol. 1: *On sounds.* London: Trübner & Co.

Blust, Robert A., 1999, Austronesian languages. *Encylopaedia Britannica.* http://www.britannica.com/.

Casparis, J.G. de, 1975, *Indonesian palaeograhpy.* Leiden/Köln: E.J. Brill.

Clynes, Adrian, 1989, Speech styles in Javanese and Balinese: a comparative study. MA thesis, The Australian National University.

Coedès, G., 1968, *The Indianized states of Southeast Asia.* Honolulu: University of Hawai'i Press.

Collins, James T., 2000, Kepelbagian dialek Melayu Sarawak: Tinjauan di Sungai Saribas dan di Sungai Rejang. *Dewan Bahasa* 44:2–17.

Coulson, Michael, 1976, *Sanskrit: an introduction to the classical language.* Hodder and Stoughton: Teach Yourself Books.

Dahl, Otto Chr., 1951, *Malgache et maanjan: une comparaison linguistique.* Oslo: Egede-Instituttet.

1991, *Migration from Kalimantan to Madagascar.* Oslo: Norwegian University Press.

Djohan Hanafiah, 1995, *Melayu-Jawa: citra budaya dan sejarah Palembang.* Jakarta: PT RajaGrafindo Persada.

Dyen, Isidore, 1956, The Ngaju-Dayak 'Old Speech Stratum'. *Language* 32:83–87.

Feinstein, Mark, 1979, Prenasalization and syllable structure. *Linguistic Inquiry* 10/2, 245–278.

Grijns, C.D., 1991, *Jakarta Malay: a multidimensional approach to spatial variation.* Leiden: KITLV Press.

Hall, Kenneth R., 1992, Economic history of early Southeast Asia. In Tarling, ed. 1992:183–275.

Heuven, Vincent J. van and Ellen van Zanten, 1994, Effects of substrate language on the localization and perceptual evaluation of pitch movements in Indonesian. In Cecilia Odé and Wim Stokhof, eds 1994:63–79.

Hilman Hadikusuma, 1994, *Kamus Bahasa Lampung.* Bandung: Penerbit Mandar Maju.

Horne, Elinor C., 1961, *Beginning Javanese.* New Haven: Yale University Press.

Ikranagara, Kay, 1991, *Melayu Betawi grammar.* NUSA 9.

Iskandarwassid, Yoyo Mulyana, Agus Hudari and Tjem Kartimi Sjachrul Sjarif, 1985, *Struktur bahasa Jawa dialek Banten.* Jakarta: Pusat Pembinaan dan Pengembangan Bahasa.

Mansoer Pateda, 1977, *Kamus bahasa Gorontalo–Indonesia (Gorontalo–Indonesian dictionary).* Jakarta: Pusat Pembinaan dan Pengembangan Bahasa.

Marwati Djoened Poesponegoro and Nugroho Notosusanto, 1990a, *Sejarah Nasional Indonesia*, vol. 2. Jakarta: Balai Pustaka.

1990b, *Sejarah nasional Indonesia*,vol. 3. Jakarta: Balai Pustaka.

Muhadjir, 1981, *Morphology of Jakarta dialect: affixation and reduplication.* NUSA 11.

Mustapa Kamal, Y. Tri Mantomo GH, Asmadi M. T. Raz and Mochtar Umar, 1984, *Struktur bahasa Melayu Sambas.* Jakarta: Pusat Pembinaan dan Pengembangan Bahasa.

Nothofer, Bernd, 1982, Central Javanese dialects. In Amran Halim Lois Carrington and S.A. Wurm, eds *Papers from the Third International Conference on Austronesian Linguistics*, vol.3:287–309.

Odé, Cecilia, 1994, On the perception of prominence in Indonesian: an experiment. In Odé and Stokhof, eds 1994:151–166.

Odé, Cecilia and Wim Stokhof, eds, 1994, *Proceedings of the Seventh International Conference on Austronesian Linguistics.* Leiden University: Department of Languages and Cultures of South East Asia and Oceania.

Raffles, Thomas S., 1817, *History of Java*, vol. 1. London: Black, Parbury, and Allen.

Ras, J.J., 1968, *Hikajat Bandjar: a study in Malay historiography.* The Hague: Martinus Nijhoff.

Smith-Hefner, Nancy J., 1983, Language and social identity: speaking Javanese in Tengger. PhD dissertation, University of Michigan.

Sridhar, S.N., 1978, Linguistic convergence: Indo-Aryanization of Dravidian languages. *Studies in the Linguistic Sciences* 8/1:197–215.

Tadmor, Uri, 1997, Accent in Austronesian. Paper presented to the Austronesian Circle, Honolulu.

2000, Rekonstruksi aksen kata bahasa Melayu. *PELBBA* 13, 153–167.

2001, Phonological convergence in the Malay Archipelago. Paper presented at the Annual Meeting of the Deutsche Gesellschaft für Sprachwissenschaft, Leipzig.

Tarling, Nicholas, ed., 1992, *The Cambridge history of Southeast Asia*, vol. 1. Cambridge: Cambridge University Press.

Teeuw, A. and D.K. Wyatt, 1970, *Hikayat Patani: the story of Patani.* The Hague: Martinus Nijhoff.

Thomason, Sarah G. and Terrence Kaufman, 1988, *Language contact, creolization, and genetic linguistics.* Berkeley: University of California Press.

Trudgill, Peter, 1974, *Sociolinguistics: an introduction.* Harmondsworth: Penguin Books.

Tuuk, H.N. van der, 1867, Foreword to Homan 1867 (pp.V–VIII).

Walker, Dale F., 1976, *A grammar of the Lampung language: the Pesisir dialect of Way Lima.* NUSA 2.

Watson Andaya, Barbara, 1992a, Political developments between the Sixteenth and Eighteenth Centuries. In Tarling, ed. 1992:402–459.

1992b, Religious developments in Southeast Asia, c. 1500–1800. In Tarling, ed. 1992:508–571.

Wolff, John, 1988, The contribution of Banjar Masin Malay to the reconstruction of Proto-Malay. In Mohd. Thani Ahmad and Zaini Mohamed Zain, ed. *Rekonstruksi dan cabang-cabang bahasa Melayu Induk.* Kuala Lumpur: Dewan Bahasa.

Yong, Janet, 1998, Word accent in Malay. Paper presented to the Second International Symposium on Malay/Indonesian Linguistics, Ujung Pandang.

Zainul Arifin Aliana, R.M. Arif, Muslim Tuwi, A. Malian Erman and Ahmad Rozi Zakaria, 1985, *Kamus Ogan-Indonesia.* Jakarta: Pusat Pembinaan dan Pengembangan Bahasa.

*Raising of PMP *a in Bukar-Sadong Land Dayak and Rejang*

RICHARD McGINN

1 Introduction

This paper is a report on field work in progress on two Austronesian language groups: the Rejang, spoken in Bengkulu and South Sumatra Provinces of Indonesia; and the Bukar-Sadong dialects of Land Dayak (called Bidayŭh /bidayəh/ [bidayɨh]) spoken in the Serian District, Sarawak, Malaysia.[1] Of particular interest is a change raising PMP **-aC* > -/ʌC/ in Bukar-Sadong, but not before a final velar: e.g. **bulan* > /burʌn/ 'moon' but **anak* > /anak/ 'child' in Tibakang, Mĕntu [mʌntu] (cf. Court 1967), and neighbouring dialects. This change is typologically interesting because a comparable change occurred in the Rejang language of Sumatra (Blust 1984; McGinn 1997). Other factors inviting comparison include: 30% shared vocabulary in the Swadesh 200-list; some unusual lexical and grammatical items; nasality features (pre-ploded final nasals, contrast between simple and 'barred' nasals prevocalically); and a few possibly shared phonological innovations such as the following.

(1) CV:CV(C) > CVCV:(C) (both Sadong and Rejang are 'oxytone' languages)

(2) Loss of PMP **qa-* in trisyllables (ubiquitous in Borneo (Blust 1990:240))

(3) Neutralisation of PMP prepenultimate **a* (cf. Sad. prefixes /bi-, ti-/ = Rej. /bə-, tə-/)

(4) Shared retention of PMP **uy* and comparable changes affecting the other diphthongs

[1] Abbreviations used are: PAn = Proto Austronesian, PBS = Proto Bukar-Sadong, PMP = Proto Malayo-Polynesian, PR = Proto Rejang. I am grateful for assistance from Mr Saudi Haji Narani of Serian District, Sarawak, Malaysia; and from Dr Zainubi Arbi and Mr Sabidin Ishak of Kabupaten Rejang-Lebong, Sumatra, Indonesia. Language assistants for Bidayŭh-Sadong dialects were as follows. Tibakang dialect: Tuan Peter, Tuan Steward, and Puan Patsi; Tapŭ-Mentu dialect: Tuan Leon, Puan Puni, Tuan Kehing, and Hilda; Mawang-Mentu: Puan Chalom and Tuan Buŭn; Ranchang dialect: Tuan Jerom and Tuan Tapoh; Mujat dialect: Tuan Angĕw and Tuan Jataw; Bĕdup dialect: Tuan Arip. For the Rejang dialects the language assistants were: Pak Ismael (Kebanagung); Ibu Baima (Lebong); and Mo. Hj. Daud, Pak Ibrahim, Ibu Kartila, Alamsyah, and Mariam (Rawas).

John Lynch, ed. *Issues in Austronesian historical phonology*, 37–64.
Canberra: Pacific Linguistics, 2003.

(5) Last-syllable (stressed) schwas reflecting PMP **e* except before **-q*: **p-inzem* > /minjəm/ 'borrow' in both languages; **taneq* > *tanaq 'earth' in both languages.

(6) PMP **-a* > **-ə*, e.g. **mata* > Sadong /batəh/, Rejang /matəy/ 'eye'.

(7) PMP **-aC* > Sadong -/ʌC/, Rejang -/əC/ except before velars;[2] for example **taŋan* > Sadong /tʌŋʌn/, Rejang /taŋən/ 'hand'; but **anak* > Sadong, Rejang /anak/ 'child'.

The question is whether such comparisons, especially (6) and (7), are due to chance, borrowing, or inheritance from a lower-order protolanguage (subgroup). My paper will argue against chance ('drift') to explain most or all of these comparisons, and argue for a mixed tree- and wave-theoretical account based on some early shared innovations followed chronologically by a few conspicuous borrowings. If accepted, the hypothesis places the pre-Rejangs in Borneo until around 1200 BP, when they migrated to their present location in Sumatra. On the basis of other changes not shown above (e.g. PMP **l* > *r* in Sadong but not Rejang — and not all Bidayŭh languages, either (see §5.2 below)) — I argue (a) that change (1) above occurred later than (2)–(7) and spread by borrowing; and (b) that (7) cannot reflect a shared innovation. The unusual nature of the conditioning in (7) suggests borrowing.

The paper is in three parts. Section 2 reviews relevant aspects of Rejang historical phonology based on Blust (1984) and McGinn (1997). Section 3 presents some previously proposed subgrouping hypotheses for Rejang and shows them to be untenable. Section 4 introduces new evidence that Rejang's closest linguistic relationship might be the Land Dayak group in Western Borneo, and this is evaluated in §5. Phonological, lexical, and grammatical evidence is presented that is consistent with a mixed tree- and wave-theoretical subgrouping hypothesis locating the Rejangs in Borneo prior to migrating to Sumatra around 1200 BP.

2 Aspects of Rejang historical phonology

2.1 Vocalic change

Robert A. Blust (1984) has demonstrated that Rejang exhibits more changes in the vowels than any other known Austronesian language. The following chart illustrates 27 splits and 21 mergers of the original four PMP vowels reflected in the Musi dialect.

Table 1: Rejang-Musi outcomes for PMP vowels:
mergers = 27 (horizontal); splits = 21 (vertical) (Blust 1984)

*a (9)	∅	a	ə	e	o	i	u	əa	əy			
*e (7)	∅	a	ə	e	o			əa		oa		
*i (5)			ə	e		i			əy		ea	
*u (6)	∅		ə		o		u			oa		əw

[2] I assume, following Blust (1990:223), that **q* was a back velar. Thus the environment 'before velars' includes 'before **q*'.

2.2 Fate of PMP **a* in Proto Rejang: ten changes

Of the seven comparisons listed in the Introduction, the most important ones for subgrouping purposes involve PMP **a*. As Table 1 shows, this etymon underwent nine mergers in the history of Rejang. However, the number of changes is even greater, since some of the mergers (e.g. PMP **a* > /ə/) occurred in more than one environment (see §2.3). Eleven outcomes which affected PMP **a* directly are illustrated below.

Outcome 1: Prepenultimate neutralisation: **a* > **ə*

PMP	Proto Rejang	
**balaŋa*	**bəlaŋi*	'pot'
**salambaw*	**səlambəw*	'trap'
**maŋ-*	**məŋ-*	(verbal affix)
**maR-*	**bə-*	(verbal affix)

Outcomes 2–5: PMP Penult **a* > **o*, **ä*, **u*, **i* (root harmony)

	PMP	Proto Rejang	Kebanagung	
2.	**manuk*	**monok*	*monok*	'chicken'
3.	**laŋit*	**läŋät*	*leŋet*	'sky'
4.	**sapu*	**supu*	*supəw*	'broom'
5.	**tali*	**tili*	*tiləy*	'rope'

Outcomes 6–9: PMP ultimate **-a* > **ə*, **i*, **o*

	PMP	Proto Rejang	Kebanagung	
6.	**kita*	**kitə*	*itə*	'we (incl.)'
	**niʔe*	**nə*	*nə*	'he/she'
7.	**duha*	**dui*	*dui*	'two'
	**tua*	**tui*	*tui*	'old'
8.	**mata*	**mati*	*matəy*	'eye'
	**naŋa*	**naŋi*	*naŋəy*	'fork of river'
9.	**depa*	**dəpo*	*dəpo*	'fathom'
	**teka*	**təko*	*təko*	'come'

Outcome 10: PMP **a* neutralised in 'diphthongs': **aw*, **ay* > **əw*, **əy*

PMP	Proto Rejang and Rejang		
**Danaw*	*danəw*	(Lebong)	'lake'
**punay*	*punəy*	(Lebong)	'dove'

compare

**qatey*	*atəy*	(Lebong)	'liver'
**hapuy*	*apuy*	(Rawas)	'fire'
**kahiw*	*kiwi*	(Rawas)	'wood'

Outcome 11: **-aC* > **-əC* except before velars

PMP	Proto Rejang	Kebanagung	
**bulan*	**bulən*	*bulən*	'moon'
**quzan*	**ujən*	*ujən*	'rain'
**tawaD*	**tawəh*	*tawəh*	'haggle'
**anak*	**anak*	*anak*	'child'
**hisaŋ*	**isaŋ*	*isaŋ*	'gills'
**hasaq*	**asaq*	*asah*	'sharpen'

Outcome 12: PMP **a* reflected as /a/ in monosyllables and in etyma with schwa in the penult

PMP	Proto Rejang	Kebanagung	
**ba*	**ba*	*ba*	(particle)
**hekan*	**kan*	*kan*	'fish'
**tebas*	**təbas*	*təbas*	'clear-cut'

2.3 Pre-Rejang word stress

In McGinn (1997, 1999) it was demonstrated that, given an appropriate (internal) reconstruction of certain pre-Rejang prosodic features, all changes that directly affected PMP **a* occurred in unstressed syllables. 'Appropriate' in this context means that pre-Rejang's stress system was virtually identical to that of contemporary Malay: stress was final when the penult was **e*, otherwise penultimate. Consider in this light the naturalness of Outcomes 1 and 6 in pre-Proto Rejang, whereby **a* neutralised in unstressed syllables (twice).

**bala:ŋa*	>	**bəla:ŋə*	> ...	'cooking pot'
**ma:ta*	>	**ma:tə*	> ...	'eye'

Nowadays, however, in all contemporary Rejang dialects the stress falls uniformly upon the final syllable of the word. To account for the contemporary data, McGinn (1997) assumed that by the time of Proto Rejang the stress had shifted, so that diphthongisation and other changes affected (newly) stressed schwas from PMP **a*, whereas 'root harmonisation' affected (newly) destressed reflexes of **a*. These assumptions are illustrated below.

PMP	pre-Rejang			Proto Rejang	Kebanagung	
**mata*	**ma:tə*	>	**matə:*	**mati:*	*matəy*	'eye'
**talih*	**ta:li*	>	**tali:*	**tili:*	*tiləy*	'rope'

The complete derivation of the word for 'eye' illustrates a series of changes whose ordering is reconstructible from internal evidence, including conspicuous 'archaic residues' in the language which, according to McGinn (1997), are actually systematic and not true exceptions. Consider the following data.

Table 2: Sample derivation: Kebanagung *əy* from **-a* in word for 'eye'

Outcome	PMP	pre-Rejang	Proto Rejang	Kebanagung	
6	**kita*	**kitə*	**itə*	*itə*	'we (incl.)'
	**niʔa*	**niʔə*	**nə*	*nə*	'he/she/it'
7	**duha*	**du:ə*	**dui:*	*dui:*	'two'
	**tua*	**du:ə*	**tui:*	*tui:*	'old'
8	**mata*	**ma:tə*	**mati:*	*matə:y*	'eye'
	**naŋa*	**na:ŋə*	**naŋi:*	*naŋə:y*	'fork of river'
9	**depa*	**dəpa:*	**dəpo:*	*dəpo:*	'fathom'
	**teka*	**təka:*	**təko:*	*təko:*	'come'
12a	**ba*	**ba*	**ba*	*ba*	(particle)
	—	**bi*	**bi*	*bi*	(particle)
12b	**hekan*	**kan*	**kan*	*kan*	'fish'
	**daqan*	**dan*	**dan*	*dan*	'branch'

As Table 2 illustrates, pronouns were affected by Outcome 6; content words with **u* in the penult were affected by Outcome 7 (two changes, including Outcome 6); the 'elsewhere' set underwent Outcome 8, which is the most complex, subsuming Outcomes 6, 7, and 8. Finally, none of these outcomes is reflected in two other classes of etyma: (a) when the penult was PMP **e* (presumed to be schwa), **-a* resisted neutralisation and eventually changed to /o/ (Outcome 9); and finally (b) in monosyllables **-a* was unaffected (reflected as -/a/).

McGinn (1997) explained all of these outcomes by first reconstructing aspects of pre-Rejang metrical structure, and then deriving the attested vowels in an array of stressed and unstressed syllables. In particular, neutralisation of **-a* (Outcome 6) affected unstressed syllables, and all other changes affected stressed syllables. Finally, monosyllables by definition are 'unfooted', and hence lack metrical structure. As Outcome 12 indicates, the unfootedness of monosyllables seems to have played a role not only in the history of PMP **-a*, but also of **-aC* in Rejang.

2.4 Rejang and Bukar-Sadong-Bidayŭh

The above examples provide sufficient background to commence the comparative part of this paper. Clearly, any language outside the Rejang area that exhibits some or all of the richness and subtlety of the above system of rules merits further examination as a possible subgrouping partner with Rejang. To begin with, consider the following formula summarising the conditions under which PMP **a* underwent neutralisation in pre-Rejang. (Readers interested mainly in the comparative analysis are urged to skip the next section and turn directly to §4.)

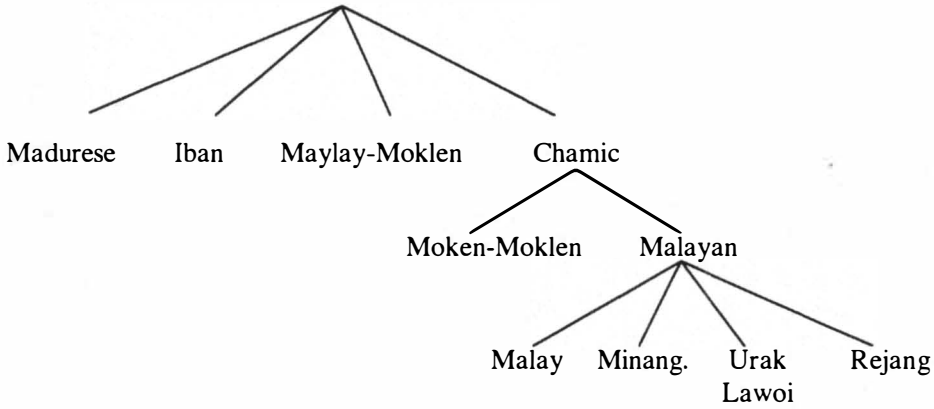
**a* > **ə* / V:C__(C[-velar])#

pre-Rejang			Kebanagung	
**ma:ta*	>	**ma:tə* ...	*matə:y*	'eye'
**da:naw*	>	**da:nəw*	*danə:e*	'lake'
**tawaD*	>	**ta:wəh*	*tawə:h*	'haggle'
**anak*	>	**a:nak*	*ana:k*	'child'

3 Rejang historical phonology: in search of an interpretation

There are at least two 'uses' for historical phonology: to provide data for the study of sound change, and to contribute to language classification. In the previous section we outlined a few of the more interesting sound changes in Rejang. In this section, all previously proposed subgrouping hypotheses for Rejang (none of them satisfactory) are reviewed.

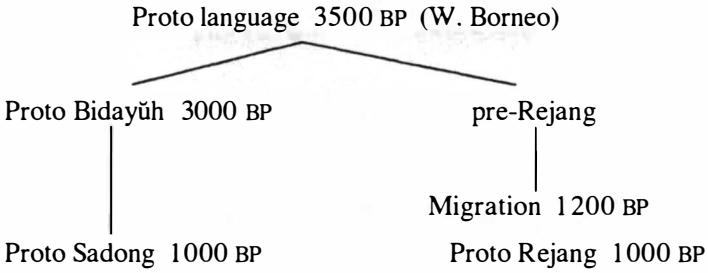
Blust (1981) attempted to classify Rejang together with Malay, Sundanese, and Maloh on the basis of shared exceptional vocabulary, in particular, the numerals 'seven', 'eight', and 'nine', which are clearly uninherited (borrowed) in all four languages; moreover, all correspond closely with Malay *tujuh, delapan, sembilan*, e.g. Rejang-Musi /tojoa?/ 'seven', /lapən/ 'eight', /səmilən/ 'nine'. Merritt Ruhlen (1987) used the same evidence to include Rejang under Chamic and Malayan, and coordinate with Malay, Minangkabau, and Urak Lawoi.



More recently, however, Adelaar (1991) and Blust (1992) argued against any close relationship between Rejang and Malay. In fact, Rejang shares none of the diagnostics (apart from numerals 7, 8 and 9) with Malay, Acehnese, and the Chamic languages. For example, in Rejang, word-initial PMP **w-* exhibits not weakening to zero (sometimes /h/) as in Malay *(h)ari* 'day', *air* 'water', but rather strengthening to /b/-, as in Rejang-Musi /biləy/ 'day', /bioa/ 'water'. Blust concluded that Rejang's numerals 7, 8 and 9 must be borrowings from Malay, thus leaving Rejang unclassified, with no known close relatives. McGinn (1999) attempted to account for a number of morphophonological comparisons involving Rejang, Malay, and Mukah Melanau. However, subsequent field work in Sarawak failed to support a subgrouping hypothesis. As matters now stand, therefore, the position of Rejang is unknown, apart from the fact that it belongs in the PMP subgroup of the Austronesian family.

4 In search of the homeland of the Rejangs

In the remainder of this paper, I introduce evidence suggesting that early Rejang evolved in Borneo and might possibly be grouped as a coordinate member of a subgroup that includes the Land Dayak languages. The following is a general outline of the hypothesis.



The major piece of evidence for this idea, and what drew me to study the Bidayŭh, involves the fate of PMP **a* in final syllables in Bukar-Sadong dialects. The Bukar-Sadong dialects are spoken in numerous villages along the Sadong river in the First District, Sarawak, Malaysia, in and to the north of the city of Serian.

4.1 Data and analysis

Recently I spent seven weeks in Sarawak and three weeks in Sumatra collecting data based on two lists. The first was a finderlist consisting of 300 PMP etyma and 107 additional terms, all presented with Malay equivalents (useful for eliciting from bilingual speakers); the second consisted of 200 sentences developed by Amran Halim from the Swadesh 200-word list, presented in the form of full sentences (in Malay for elicitation purposes). For example, the meaning ALL (English *all*) was elicited by asking for a translation of the Malay sentence *Semua manok kami mati* ('All our chickens have died'). The 407-word finderlist and the 200-sentence list were recorded for the Tibakang, Tapŭ, Bedŭp, Mujat and Ranchan dialects of Sadong and the Lebong, Kebanagung and Rawas dialects of Rejang. I had previously collected data based on the same lists for two other Rejang dialects, namely Lebong and Pasisir. The Appendix displays over 200 reconstructed Proto Rejang forms (based on five dialects) alongside a similar number of Proto Bukar-Sadong forms. All comparisons shown in the remainder of this paper are based on the reconstructed protolanguages; attested forms represent contemporary dialects.

4.2 Summary of PMP last-syllable *a Raising in pre-Rejang

Before beginning the comparative part of the paper, consider again the following formula, which represents three of the earliest changes affecting the historical phonology of pre-Rejang.

PMP	>	pre-Rejang
**a*	>	**ə* / V:C__(C[-velar])#

4.3 PMP last-syllable *a Raising in pre-Bukar-Sadong

What is interesting in the context of this paper is that the set of pre-Rejang changes shown by the formula in §4.2 *almost* works for reconstructed pre-Bukar-Sadong as well. Consider the following set of changes, to be described in full in this section.

	PMP	>	pre-Bukar-Sadong
a.	**a*	>	**ə / V:C __ #*
b.	**a*	>	**ʌ / V:C __ (C[-velar]) #*
c.	**-aw*	>	**əw . . . > u*
	**-ay*	>	**əy . . . > i*

The next display illustrates the range of changes represented above, which will be described in detail below.

	PMP	pre-Bukar-Sadong	Tibakang		Section discussed
a.	**duha*	**du:ə*	*duə:h*	'two'	4.3.1
b.	**Danaw*	**da:nəw*	*danu:*	'lake'	4.3.2
	**punay*	**pu:nəy*	*puni:*	'dove'	
c.	**taŋan*	**ta:ŋʌn*	*tʌŋʌ:n*	'hand'	4.3.3
	**hepat*	**u:mpʌt*	*umpʌ:t*	'four'	

To help explain all of these changes, I assume that pre-Bukar-Sadong (like pre-Rejang) had a Malay-type stress system: i.e. *the accent fell on the ultimate when the penult was schwa; otherwise on the penult.* Another assumption is that all contemporary Bukar-Sadong dialects have ultimate stress, again like Rejang; certainly, all those which have been investigated show this pattern.

*4.3.1 Neutralisation of PMP word-final *a in open final syllables*

Both languages show evidence of early neutralisation of PMP **-a* in open final syllables.

PMP	Pre-Rejang	Pre-Sadong	Tibakang	
**duha*	**du:ə*	**du:ə*	*duə:h*	'two'
**mata*	**ma:tə*	**ma:tə*	*batə:h*	'stone'
**naŋa*	**na:ŋə*	**na:ŋə*	*naŋə:h*	'fork of river'
**lima*	**li:mə*	**li:mə*	*limə:h*	'five'
**niʔa*	**ni:ʔə*	**ni:ʔə*	*niʔə:h*	'he/she'

*4.3.2 Neutralisation of PMP word-final *-a in pre-Bukar-Sadong diphthongs*

Both languages show evidence that **a* raised to **ə* in PMP **aw* and **ay.*

PMP	Pre-Bukar-Sadong	Proto Rejang and Rejang		Proto Bukar-Sadong and Tibakang	
**Danaw*	**danəw*	*danəw*	(Lebong)	*danu*	'lake'
**punay*	**punəy*	*punəy*	(Lebong)	*puni*	'dove'
**qatey*	**atəy*	*atəy*	(Lebong)	*ati*	'liver'
**hapuy*	**apuy*	*apuy*	(Rawas)	*apuy*	'fire'
**kahiw*	**kaiw*	*kiwi*	(Rawas)	*kayu*	'wood'

*4.3.3 Raising of PMP *a in closed final syllables 'except before velars'*

The data in this section is what first drew my attention to the comparison of Rejang and Bidayŭh.

PMP	Rejang (Kebanagung)	Bukar-Sadong (Tibakang)	
**bulan*	*bulə:n*	*burʌ:tn*	'moon'
**quzan*	*ujə:n*	*ujʌ:tn*	'rain'
**tawaD*	*tawə:h*	*tawʌ:r*	'haggle'
**anak*	*ana:k*	*ana:k*	'child'
**hisaŋ*	*isa:ŋ*	*insa:kŋ*	'gills'
**hasaq*	*asah*	*ŋ-asaʔ*	'sharpen'

This comparison offers the strongest evidence of a greater-than-chance relationship between Rejang and Bukar-Sadong (see §5.2.2).

4.4 More phonological evidence

In addition to the above evidence for relating Rejang and Bukar-Sadong dialects, consider the following phonological comparisons. Many of these types of changes are common elsewhere in the Austronesian family, and therefore may seem to have little subgrouping value, as would certainly be the case if each were evaluated individually. In the aggregate, however, they seem to add up, if not to a fully verified subgroup, at least to an indication that the Rejangs originated in Borneo (rather than, say, Taiwan, the Philippines, Sulawesi, Sumatra, or the Malay peninsula), for almost all of the resultant features are particularly widespread in Borneo.

Rejang and Bukar-Sadong	Widespread in Borneo	Shared by Malay
**qa- > Ø* in trisyllables	YES	NO
**Ca- > *Cə-* in trisyllables	YES	YES
**-q > *-ʔ*	YES	NO
**z > *j* (except Rejang *d-* in 'road' and 'needle')	YES	YES
**-mb-, -nd- > -m^b-, -n^d-* ('barred nasals')	YES	NO
**-m, *-n > -^{b}m, -^{d}n* (pre-stopped nasals)	YES	NO
stress shifted to final syllable	YES	NO

4.5 Grammatical comparisons

In addition to the phonological evidence just reviewed, there are a few grammatical comparisons that point in the same direction. Owing to the paucity of inflections in either language, the grammatical comparisons involve grammatical function words. (Rejang has only two inflections, the infixes -/ən/, -/əm/-, both inherited from PAn/PMP; Bukar-Sadong has only -/in/- corresponding to Rejang -/ən/- in both form and meaning.) Possibly shared changes include the following three:

(1) Suffixes are unknown in both Rejang and Bukar-Sadong.

(2) Case distinctions in the pronouns are virtually non-existent (shared with Malay).

(3) Similarities among the following grammatical function words may be significant. (Note: the symbol ~Rawas means 'all Rejang dialects except Rawas'.)

Bukar-Sadong-Tibakang	Rejang dialects		Malay	
aŋ	*taŋ*	Rawas	*di*	'at'
kai?	*coa*	~Rawas	*tidak*	'not'
api	*ipə*	~Rawas	*mana*	'where?'
kudu	*kədəw*	all	*berapa*	'how many?'
mbəh	*bi*	all	*sudah*	'already'
kelek	*kəlak*	all	*mau, hendak*	'want'
boh, mah	*ba*	all	*-lah*	'imperative particle'

Two caveats are in order with respect to the list of function words, however. First, apart from PMP **ba* (imperative particle), the PMP etyma for these words have not been reconstructed, so it is not really known at present whether the data represent shared innovations or simple retentions. Second, it must also be acknowledged that the similarities are merely impressionistic. Whether these comparisons will eventually prove valid must await the results of future research.

4.6 Lexical comparisons

Finally, a few apparently shared irregularities turned up during the search for shared cognates. Consider the following data.

	Bukar-Sadong	Rejang		
PMP	Tibakang	Rawas	Malay	
**bali*	*jaji*	*jijəy* < PR **jaji*	*jadi*	'become'
**kutu*	*gutu*	*gutəw*	*kutu*	'head louse'
**tuqelan*	*tʌrʌtn* 'Adam's apple'	*təlan*	*tulaŋ*	'bone'
**tisuk*	(Mujat *-ujak*)	*tujah*	*tikam*	'to stab'
(7-8-9)	(borrowed)	(borrowed)	(borrowed)	'7, 8, 9'
**pitu*	*iju?*	*tojoh*	*tujuh*	'seven'
**walu*	*mahi*	*lapən*	*dəlapan*	'eight'
**siwa*	*piri?i*	*səmilən*	*sambilan*	'nine'

Shared forms from unknown sources are potentially significant as evidence of early shared borrowings. Thus Tibakang /jaji/ corresponds well with Rawas /jijəy/ < PR **jaji* in form and meaning; so also /gutu/ = /gutəw/ from PMP **kutu* (although **k-* > /g/- is widespread in Western Austronesia). The third form, Bukar-Sadong-Mujat /tʌrʌtn/ 'Adam's apple' corresponds with Rejang /təlan/ 'bone', but, if these are cognates, the Tibakang form has undergone a semantic shift. The words for 'stab' and 'seven' are obviously borrowed in both languages, with a strong resemblance in form and meaning. Finally, the numerals 'seven', 'eight' and 'nine' are obviously

borrowed in both languages: 'seven' probably from Malay; 'eight' and 'nine' from different sources; yet in both languages it is striking that just these three numerals are borrowings.

5 Lexicostatistics and glottochronology

Although the case for a Rejang-Bidayŭh subgroup is far from proven, it is nonetheless helpful to consider some of the consequences that would follow from the assumption that the hypothesis is true. It is in this spirit that I propose to explore some further evidence based on lexicostatistics and glottochronology. Although discredited if taken as exact sciences, these two methods nevertheless constitute useful tools for the extraction of two kinds of information from a body of data: the one to quantify relative linguistic 'distance' between two or more languages; the other to assign tentative dates to language splits. Among the standard caveats, it is perhaps also necessary to point out that the two methods are interdependent in the sense that the relative distance between, say, languages A, B, and C remains constant no matter what dates are assigned. For example, the relative distance between languages A, B, C is the same no matter whether t = 1000 years or 10,000 years. Thus there is no contradiction in adjusting the value of t in order to conform to other lines of evidence, such as archaeological evidence, and even cultural evidence — whether the speakers tend more to linguistic conservatism or the reverse, leading to widespread rapid borrowing.

5.1 Rejang and Bukar-Sadong

Given this much as introduction to the use of statistical methods in historical phonology, consider the following table:

Table 3: Cognate percentages with tentative dates

	% shared homosemantic cognates	(r = 84%)
Rejang dialects:	70–94%	= 1000 years
Sadong dialects:	70–88%	= 1000 years
	(Topping 1990)	
Bidayŭh dialects incl. Lara':	33–36%	= 3000 years
	(Kroeger 1998)	
Tibakang and Kebanagung:	30%	= 3500 years
(Sadong)	(Rejang)	(my field work)

There are three observations to be made about Table 3. First, the maximum spread within Rejang dialects and within Bukar-Sadong dialects is about the same — around 70% shared basic vocabulary (McGinn field work conducted in 2000 and 2001; cf. Topping 1990). Second, the maximum separation within the entire Bidayŭh group, which includes the Lara' language of west Kalimantan, is 33–36% (Kroeger 1998). Third, Rejang's nearest cousin is unknown, unless indeed this turns out to be the Bidayŭh group itself.

Next, for concreteness, it has been useful to assign relative dates to these figures (with the usual caveats). When *r* (assumed rate of replacement of basic vocabulary per thousand years) = 84%, then 70% shared basic vocabulary = 1000 years and 30% shared basic vocabulary = 3500 years. Therefore, both Proto Rejang and Proto Bukar-Sadong began their dialect splits 1000 years ago, and Proto Bidayŭh began splitting into different languages around 3500 years ago. The lowest-order protolanguage containing both pre-Bidayŭh and pre-Rejang began separating earlier than 3500 years ago.

5.2 Problems

The statistical evidence introduced above is consistent with a subgrouping hypothesis for Rejang and Bidayŭh, but problems remain. Here I will mention what I consider to be the two most serious objections to the hypothesis. First, the change **l* > /r/ affected Bukar-Sadong dialects but not Rejang. Second, the Bukar-Sadong version of PMP **a* Raising (**a* > ʌ / *__C# except before velars) — which so temptingly resembles the Rejang version — is (apparently) not found in any other Bidayŭh dialects. Taken together, these two facts undermine any supposed subgroup at the level of Proto Rejang and Proto Bidayŭh. The case would be permanently closed if **l* > /r/ were assumed to be diagnostic for membership in the Bidayŭh language group, but as we shall see in the next section, any such conclusion would be incorrect. Nevertheless, it is probable that **l* > /r/ preceded PMP **a* Raising in Bukar-Sadong, and if so, the Rejang version of PMP **a* Raising cannot be a shared innovation with Bukar-Sadong. These two objections notwithstanding, there still exists one more possible scenario open for uniting these two languages at some level lower than PMP.

*5.2.1 How widespread was PMP *l > /r/ in Bidayŭh?*

As mentioned, /r/ from PMP **l* is reflected in Bukar-Sadong and many other Bidayŭh languages, but not in Rejang. The crucial point, however, is that this change is likewise not attested in some other Bidayŭh languages, a number of which regularly show PMP **l* as /l/ (e.g. Grogo, Sau, Milikin). Consider the Grogo forms below taken from Ray (1913).

PMP	Proto Bidayŭh 3000 BP	Bidayŭh-Grogo (Ray 1913)	Proto Bukar-Sadong 1000 BP	
**laŋit*	**laŋit*	*laŋit*	**raŋit*	'sky'
**laud*	**laud*	*laud*	**laut* (irregular *l-*)	'sea'
**silun*	**silun*	*silun*	**siruh* (irregular *-h*)	'fingernail'
**tuqelaN*	**tulaŋ*	*tulaŋ*	**turaŋ*	'bone'
**talih*	**talih*	*toli*	**tarih*	'rope'
**bulan*	**bulan*	*bulan*	**burʌn*	'moon'

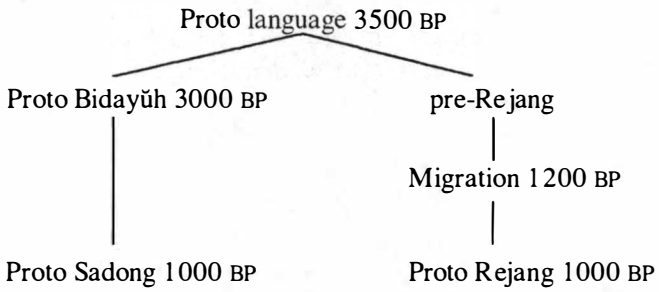
On this the evidence, **l* > /r/ cannot be assigned to Proto Bidayŭh.

*5.2.2 PMP *a raising in Bukar-Sadong and Rejang (revisited)*

Let us again consider the comparative data of PMP **-aC* Raising shown in §4.3.3 above. As far as I know, the Bukar-Sadong version of PMP **-aC* Raising is not found in other Bidayŭh dialects, in contrast to **l* > /r/ which is fairly widespread. It follows that **l* > /r/ must have preceded **a* Raising in Bukar-Sadong; and therefore no version of **a* Raising can possibly be assigned to any subgroup containing Proto Rejang and Proto Bukar-Sadong as members. Our most interesting comparison, therefore, must be due to borrowing (language contact) or chance (phonetic drift). But the likelihood of chance must be considered extremely low given the unusual nature of the conditioning (**-aC* underwent raising 'except before velars') in exactly these two languages. Therefore, I shall argue for borrowing as the more likely explanation.

5.2.3 'Saving the hypothesis'

If the hypothesis is to stand up against the two objections mentioned in §5.2.1 and §5.2.2, then the only way to save the hypothesis is to introduce a 'mixed' theory based on an orderly application of tree-theory and wave-theory assumptions. The following display outlines the temporal and geographical requirements of the revised hypothesis.



The final display below indicates in more detail the set of assumptions consistent with the hypothesis.

A plausible sequence of events

before 3500 BP	Shared innovations defining pre-Rejang and pre-Bidayŭh subgroup (e.g. **-a* > **-ə*)
3500 BP	Language split into pre-Rejang and Proto Bidayŭh
3500–1200 BP	Pre-Rejang in contact with pre-Sadong in Sarawak
1500–1200 BP	1. **-V:CaC*[-velar] > **-V:CʌC* spread by borrowing 2. Final syllable stress spread by borrowing 3. **l* > /r/ in Sadong (did not spread to Rejang)
1200 BP	Proto Rejang migration
Rejang:	**ʌ from *a merged with /ə/
Sadong:	**ʌ from *a became new (7th) vowel
1000 BP	Rejang dialect split Bukar-Sadong dialect split

6 Epilogue: a role for 'shared retentions' in language classification

It is widely assumed that only shared innovations are relevant for subgrouping, and that shared retentions have no subgrouping value. This assumption has served as a guiding principle in this paper up to this point. A moment's reflection, however, should be enough to convince anyone that this principle has no real theoretical status. At best it is a caveat advising students of language not to waste their time looking for subgrouping hypotheses in unlikely places. But consider the matter theoretically. Assume the case of a subgrouping hypothesis that is well established on the basis of a reasonable number of shared innovations, e.g. Maanyan and Malagasy (Dahl 1951). It almost goes without saying that any Maanyan and Malagasy cognates that are assumed to be inherited from the protolanguage 'directly' (without change) would be expected to be just as regular in their vacuous development as are the shared innovations in their altered development. And indeed, in practice such retentions are always regarded as backgrounded information in relation to the set of innovations, and for this reason their status is easily overlooked. What these retentions actually do theoretically (which is not usually noted) is: first, to bear witness against alternative hypotheses; and second, to add positive weight to (the preponderance of) the standard evidence in favour of the hypothesis. This holds true despite the fact that, in the simplest cases, no alternative hypotheses are under consideration, and the added weight is not needed.

But what about the more difficult cases, where such additional evidence might actually perform useful work? Such cases can and do arise, I suggest, in attempts to establish subgrouping hypotheses for isolated language groups like the Rejang. In such cases, there may be a legitimate use of evidence from shared retentions. When such evidence is examined, and considered alongside other evidence, it can help to refute a false hypothesis; and by the same token, it can add weight in support of a hypothesis.

Consider the following three classes of retentions that are found scattered among many Austronesian languages, including Proto Rejang and Proto Bidayŭh.

(a) PMP diphthong **uy* inherited as /uy/ in all known Rejang and Bidayŭh dialects.

(b) PMP infix **-in-* 'past tense' inherited as -/in/- (Bidayŭh) and -/ən- (Rejang) — reanalysed as the passive morpheme.

(c) PMP **-eC* inherited as **-əC* except before **-q*, where **-eq* > **-aq*: e.g. PMP **asəp* 'smoke' > PR, PBS **asəp* 'smoke' alongside PMP **taneq* > PR **tanaq* (not ***tanəq*), PBS **tanah* (not ***tanəh*) 'earth'. (This retention is also found in Jakarta Malay.)

The question to be asked is: can retention facts such as these, admittedly a distraction during the *initial* stages of research, nonetheless be useful at some point in the later stages of research, i.e. to *support* or *refute* an as yet unproven subgrouping hypothesis? I suggest that such facts can and should be brought to bear in cases like the hypothesis of this paper, which does have other facts to recommend it — facts that may be insufficient in number and quality to establish the hypothesis once and for all.

Consider a possible alternative hypothesis that situates Proto Rejang within some other reasonably established subgroup, such as Proto Malayic (Adelaar 1992). Clearly, two of the three retention facts mentioned above, namely that Rejang retains PMP **uy* (as /uy/) and the infix **-in-* (as -/ən/-), do not favour any close Rejang–Malayic link; rather, Rejang and Malay

must have split well beore the Malayic group underwent certain changes in the relevant etyma. Then what about the retention of **-eC* as /əC/ except before **-q*, where **e* > /a/? This retention is shared not only by Rejang and Bidayŭh, but also by at least one Malay dialect: Jakarta Malay. But since we already know that Malay does not belong in a lower-order subgroup close with Rejang, this particular comparison can be safely ignored. But the conclusion does not apply with the same force in the comparison of Rejang and Bidayŭh. It may not be totally vacuous to consider positively, in relation to the hypothesis of this paper, that Rejang and Bidayŭh have preserved all three of these features of PAn/PMP (mostly) unchanged for at least 5000 years, against the hundreds (or perhaps thousands) of changes that affected neighbouring subgroups, and indeed, against all of the possible changes that could have occurred, but did not.

Appendix: Proto Rejang and Proto Bukar-Sadong reconstructions

The reconstructed forms presented below are based on five Rejang dialects and five Bukar-Sadong dialects; data were collected by the author using Malay equivalents for elicitation purposes. For example, bilingual speakers were presented with a Malay form (e.g. *tangan* 'hand') and asked to produce the Rejang or Bukar-Sadong equivalent, which was duly taped and transcribed by the author. Rejang data were obtained in April 2001 for Rawas, Lebong, and Kebanagung; Musi and Pasisir data are from McGinn (1997). Bukar-Sadong data were obtained in December 2000 and April 2001. Unfortunately, space limitations do not permit displaying the data of all ten dialects surveyed. Included are the reconstructed protolanguages with one example from a contemporary dialect. Unless marked otherwise, the Rejang data are from the Rawas dialect, and the Bukar-Sadong data are from Tibakang. PMP forms are taken without modification from Robert A. Blust's online *Austronesian comparative dictionary* (n.d.), an invaluable resource which is hereby gratefully acknowledged.

Phonetic notes: 1. Rejang Rawas /ä/ = low fronted vowel contrasting with low back /a/. 2. In all contemporary dialects below (both languages), nasal phonemes represented orthographically as /mb/, /nd/, /nj/, and /ŋg/ are distinguished acoustically from plain nasals /m/, /n/, etc., in that whereas the latter are followed by nasalised vowels, the former are followed by oral vowels. See Scott (1964) for a description of Sadong nasal phonemes; see Coady and McGinn (1983) for the corresponding Rejang nasal phonemes.

PMP	1. *anay	2. *aɲin	3. *anak
PR	*anəy	*aɲin	*anak
PBS	*ani dbl. *riŋga	*mahu	*anak
Rawas	makak (Keb. anəe-anəe)	aɲin	anak
Tibakang	ani ani	aɲin (Mujat mahu)	anak
GLOSS	TERMITE	WIND	CHILD

PMP	4. *ajeŋ	5. *arep	6. *hasaq
PR	*ahaŋ (irreg. *e > a)	*ahəp	*asaq
PBS	*buhə	*ar[ə, ʌ]p	*ŋ-asaʔ, *[n, ŋ]-ulik
Rawas	aʔaŋ	ndak (Keb. ahəp)	asah
Tibakang	bahan apuy (Mujat buhə)	arəp	midʌn (Mujat ŋasaʔ)
GLOSS	CHARCOAL	HOPE	SHARPEN

PMP	7. *asep	8. *qatep	9. *qatey
PR	*asəp	*atəp	*atəy
PBS	*asəp	*iraw	*ati
Rawas	asəp	atəp	atuy
Tibakang	asəp	irʌw	ati
GLOSS	SMOKE	ROOF	LIVER

PMP	10. *hawak, *tubuq	11. *bahu	12. *bales
PR	*awak	*bau	*baləs
PBS	*tibu?	*sikəh	*baləs
Rawas	kəw	baəw	baləs
Tibakang	tibu?	səkəh	maləs
GLOSS	BODY	ODOUR	REPLY

PMP	13. *uRat 'vein; root'	14. *bapa-q	15. *bataŋ
PR	*balət	*bapak	*bataŋ
PBS	*uhʌt	*amaŋ	*bataŋ
Rawas	bania/akəa (Keb. balət)	bapak	bataŋ
Tibakang	uhʌt	amaŋ	təŋən kayuh (Mujat bʌtaŋ)
GLOSS	ROOT	FATHER	TREE TRUNK

PMP	16. *bibiR	17. *baniŋ	18. *babaq
PR	*bibiR; MOUTH *mus dbl. *ŋus	*beneŋ	*baq
PBS	*bibih	*kura?	*sagu?
Rawas	ŋus, bibia	labəy ~ ku?aw	pi-bah
Kebanagung	bebea 'lower lip'	beneŋ	bah
Tibakang	bibih	kura?	sagə?
GLOSS	LIPS	TORTOISE	BELOW

PMP	19. *balik	20. *baqeRu	21. *binehiq
PR	*bäläk	*bəlu	*biniq
PBS	*mari[ŋ, ?] 'pulang'	*ba[?, Ø]uh	*bini?
Rawas	bäläk	bələw	bənäh
Tibakang	balik, mari?	bauh	bene?
GLOSS	RETURN	NEW	SEED

PMP	22. *bener	23. *benaqi	24. *beReqat
PR	*bənəh	*bənəy	*bənəg dbl. bəhət
PBS	*mənə [?, Ø]	non cognate	*bahat
Rawas	bənəa	bənuy	bənəg
Tibakang	mənə	kirasik	bahat
GLOSS	CORRECT	SAND	HEAVY

PMP	25. *beRuk	26. *bitiqis	27. *betul
PR	*bəhuk	*bətis dbl. käkäl	*bətul dbl. bənəh
PBS	*baruk dbl. *ʌluk	*bites	*bʌtul
Rawas	bu?uk	käkäl (Keb. bətis)	bənəa
Tibakang	kara? (Mujat ʌluk)	betes	mənə
GLOSS	MONKEY, APE	CALF OF LEG	TRUE, CORRECT

PMP	28. *bahi	29. *qalejaw, *waRi	30. *biluk
PR	*bey	*bili	*(b)ilok
PBS	*sue?	*andu	
Rawas	ana? səlawəy	biləy	belok
Tibakang	sue? icɔk	andu	nyimpaŋ
GLOSS	CHILD	DAY	TURN

PMP	31. *baRani	32. *wahiR	33. *bituqin
PR	*bini	*biol	*bitaŋ
PBS	*pʌgʌn	*umon	*bintə?
Rawas	binəy	biol	bitaŋ
Tibakang	pʌgʌn	omon	bintə?
GLOSS	BRAVE	WATER	STAR

PMP	34. *buaq	35. *bunuq	36. *buqaya
PR	*buaq	*unuq	*buəy
PBS	*bua?	*kabəs	*bu[Ø, ?]ʌy
Rawas	buah-buah	onoh	mouy
Tibakang	bua?	kinabəs	buʌy
GLOSS	FRUIT	KILL	CROCODILE

PMP	37. *bukid	38 *bulan	39. *bulat	40. *bulu
PR	*tebə dbl. *bukit	*bulən	*bulət	*bulaw
PBS	*d[a, ʌ]rəd	*burʌn	*burŭŋ	*buruh
Rawas	təbaw	bulən	bulət	buləw
Tibakang	kajuh 'hill'	burʌn	bərəŋ	buruh
GLOSS	HILL	MOON	ROUND	FEATHER

PMP	41. *buŋa	41. *buhek	42. *buRuk
PR	*buŋi	*buk	*buhuk, dbl. *kidek
PBS	*buŋa[Ø, ?] irreg.	*buruh	*madam.
Rawas	buŋəy	buk	kedek 'bad person'
Tibakang	buŋa	buruh	madam
GLOSS	FLOWER	HEAD HAIR	UGLY; WORN OUT

PMP	43. *batu	44. *(d)aRaq	45. *lalej
PR	*butu	*dalaq	*daləj
PBS	*batuh	*d[a, ʌ]ya?	*narəd
Rawas	butəw	dalah	dalət
Tibakang	batuh	dʌya?	tura? (Mujat narʌd)
GLOSS	STONE	BLOOD	HOUSEFLY

PMP	46. *zalan	47. *daqan	48. *Danaw
PR	*dalən	*dan	*danew
PBS	*jʌrʌn	*da?an	*danu
Rawas	dalən	dan	daniw
Tibakang	jʌrʌn	da?an	danu
GLOSS	PATH, ROAD	BRANCH	LAKE

PMP	49. *dahun	50. *dilaq	51. *debu
PR	*daun	*dilaq	*dəbu
PBS	*da?un	*jile[h, ?]	*dʌbu
Rawas	daun	lidah	dəbəw
Tibakang	dawə?	jeleh	dʌbu
GLOSS	LEAF	TONGUE	DUST

PMP	52. *zaRum	53. *dapuR	54. *Duha
PR	*dolom	*dopol	*dui
PBS	*jarum		*duəh
Rawas	dolom	dopol	duəy (Keb. dui)
Tibakang	jarum	abuh	duəh
GLOSS	NEEDLE	KITCHEN	TWO

PMP	55. *dukut, *udu	56.	57.
PR	*dukut	*das	*dahət
PBS	*uduh	sʌmbu	*dayəh
Rawas	dukut	das	da?ət
Tibakang	uduh	sʌmbu	dayəh
GLOSS	GRASS	(ON) TOP	INLAND

PMP	58. *deRes	59. *hiket	60. *Rakit
PR	*dəhəs	*äkät	*häkät
PBS	*dərəs	*kabət	*lantiŋ
Rawas	də?əs	äkät	äkät (Keb. heket)
Tibakang	dərəs	kabət	lantiŋ
GLOSS	SWIFT CURRENT	TO TIE	RAFT

PMP	61. *qiliR	62. *ipen	63. *isep
PR	*iliR	*äpän	*äsäp
PBS	*saba? dbl. *mamʌn	*jipəh/jip[u, ə]n	*sihəp
Rawas	pilot	äpän	(ŋ)äsäp
Tibakang	mamʌn 'flow'	jipəh	nyəhəp
GLOSS	DOWNSTREAM	TOOTH	SUCK

PMP	64. *embun	65. *enem	66. *gatel
PR	*əmbun; awan	*num	*gatal (irreg. *e > a)
PBS	*ambun/*ramaŋ	*ənəm	*gatəl
Rawas	mbun/awən	num	gatal
Tibakang	ramaŋ	ənəm	gatəl
GLOSS	CLOUD	SIX	ITCH

PMP	67. *gilap	68. *genep	69. *quDip
PR	*gələp (spor. *i > ə)	*gənəp	*idup
PBS		*gənəp	*m-udip
Rawas	no data	gənəp	idup
Tibakang	klap-klip (borr.)	gənəp	mudip
GLOSS	FLASH	COMPLETE	ALIVE

PMP	70. *ikuR	71.	72.
PR	*ikoR	*quq	*rimbə
PBS	*uŋkuy	*joho?	*tu[?, Ø]an
Rawas	iku? (borr. Palembang)	jaoh (Keb. hoah)	imbaw
Tibakang	uŋkuy	joho?	nuan
GLOSS	TAIL	FAR	FOREST

PMP	73.	74. *hisaŋ	75. *isi
PR	*indok	*isaŋ	*isi
PBS	*[a, i]ndə	*su?op dbl. *sa?ap	*isih dbl. *abih
Rawas	indok	isaŋ	isəy
Tibakang	andə	so?op (Mujat sa?ap)	isih (Mujat abih)
GLOSS	MOTHER	GILLS	CONTENTS

PMP	76. *kita	77. *qituŋ	78.rejaŋ (name)
PR	*(k)itə	*ituŋ	*tun həjaŋ 'Rejangs'
PBS	*kita?	*itoŋ dbl. *iyəp	
Rawas	kitə	rikin (borr. English)	mərəjaɪŋ 'migrate'
Tibakang	kita?	niəp	(suŋi) rʌjaŋ 'Rejang river'
GLOSS	1PL.INCL	COUNT	REJANG

PMP	79. *zari	80. *kabut	81. *kaka-q
PR	*jihi	*kabut	*kakak
PBS	*bua? tʌŋʌn		*umbu?
Rawas	ji?əy	kabut	kakak
Tibakang	bua? tʌŋʌn	kabus	umbu?
GLOSS	FINGER	FOG	ELD. SIBLING

PMP	82. *hikan	83. *ka-wanan	84. *kasaw
PR	*kan	*kanən	*kasəw
PBS	*eke?, *lauk	*[Ø, n]ta?uh	*kasu
Rawas	kan	kanən	kasiw
Tibakang	eke?	ta?uh	kasu
GLOSS	FISH	RIGHTSIDE	RAFTER

PMP	85. *kami	86. *kawil	87. *kawit
PR	*kami	*kawil	*kait
PBS	*ami[?, Ø]	*minti?	*ka?it
Rawas	kämäy	paciŋ (Keb. kewea)	kait
Tibakang	ami	minti?	ka?it
GLOSS	1PL.EXCL	TO FISH	HOOK

PMP	88. *kutkut	89. *kempu	90. *keRiŋ
PR	*gahut	*kəpu	*kəhiŋ
PBS	*g[a,ʌ]yʌs	*suŋkuh	*b[a,ʌ,ə]də?
Rawas	kaut	kəpəw	ki?iŋ
Tibakang	gʌyʌs	suŋkuh	badə?
GLOSS	SCRATCH	GRANDCHILD	DRY

PMP	91. *esak/*tanek	92. *tawa	93. *kilat
PR	*k-əsak	*tawi	*kilət
PBS	*n-anək	*nʌtʌw	*kilʌt
Rawas	k-esak	tawəy	kilət
Tibakang	ʌsak	nʌtʌw	kilʌt
GLOSS	COOK	LAUGH	LIGHTNING

PMP	94. *kahiw	95. *kahu	96. *kamu
PR	*kiiw	*kau	*kumu
PBS	*kayuh	*amu[ʔ]	*amuʔ
Rawas	kiiw	kabən (Keb. ko)	kuməw
Tibakang	kayuh	amuʔ	amuʔ
GLOSS	WOOD	2SG	2SG/2PL

PMP	97. *kena	98. *kutu	99. *asu
PR	*kəno	*gutu	*kuyuk
PBS	*udog	*gutu	*kasuŋ
Rawas	kənaw (Keb. kəno)	gutəw	kucak (Keb. kuyuk)
Tibakang	odog	gutu	kasuŋ
GLOSS	STRIKE	LOUSE	DOG

PMP	100. *laŋaw	101. *lahud	102. *lawaq
PR	*laŋəw	*laut	*la[w, b]aq
PBS	*rʌŋu	*laʔut	*kʌkaʔ
Rawas	laŋiw	laut	ləlabah
Tibakang	turaʔ dʌbiru	laut	apək
GLOSS	HORSEFLY	SEA	SPIDER

PMP	103. *laŋit	104. *lain	105. *lebiq
PR	*läŋät	*leyn	*ləbiq
PBS	*raŋit	*bukən	*lʌbih
Rawas	läŋät	lain ~ landuman	ləbäh
Tibakang	raŋit	bəkən	lʌbih
GLOSS	SKY	OTHER	EXCESS

PMP	106. *lem	107. *lima	108. *lesuŋ
PR	*(pi)lem	*lemo	*ləsuŋ
PBS	*u[Ø, ʔ]ah	*rimәh	*risoŋ
Rawas	piləm ~ oləm	ləmaw	ləsuŋ
Tibakang	tərəp	rimәh	risoŋ
GLOSS	INSIDE	FIVE	MORTAR

PMP	109. *libeR	110. *laRiw > *laRi	111. *beRey
PR	*libəh	*lili	*ləy
PBS	*libər	*buhu[ʔ, Ø]	*jug[o, ʌ]n
Rawas	libəa	liley	luy
Tibakang	kahi (Mujat lambʌr)	buhu	nyogon
GLOSS	WIDE	RUN	GIVE

PMP	112. *basəq (*ləcəq)	113. *qali-metaq	114.
PR	*ləcaq	*litaq	*luaq
PBS	*bisaʔ dbl. *lʌcaʔ	*mataʔ dbl. *mutək	*[b, m]adaʔ dbl. *ŋancak
Rawas	ləcah	litah	titah
Tibakang	raʔus (Mujat bisaʔ)	mʌtaʔ	ŋancak
GLOSS	SOAKED	LEECH	COMMAND

PMP	115. *lurus	116. *mama-q	117. *mata
PR	*luhus	*mamak	*mati
PBS	*bujog	*ambah	*matəh
Rawas	luʔus	wak (Keb. mamak)	mati
Tibakang	bojog	biradik kawʌn	batəh (Mujat mʌtəh)
GLOSS	STRAIGHT	MO.BRO.	EYE

PMP	118. *matey	119. *embun	120. *um-inem
PR	*matəy	*-mbəm	*mänäm/*mbuk biol
PBS	*kabəs	*salak dbl. *sahu	*nyihəp
Rawas	matuy	nəmbəm	mbuk biol
Tibakang	kabəs	nyalak	nyəhəp
GLOSS	DIE	BURN	DRINK

PMP	121. *ma-iRaq	122. *mi-hepi	123. *emis
PR	*milaq, *abaŋ (borr.)	*mipi	*mis
PBS	*calak	*pi[a,ʌ]məh	*sidiʔ
Rawas	abaŋ	mipəy	mis
Tibakang	cʌlak	piaməh	sidiʔ
GLOSS	RED	DREAM	SWEET

PMP	124. *manuk	125. *ma-anyud	126. *um-utaq
PR	*monok	*monot	*mutaq
PBS	*siok	*mamʌn	*ŋ-uteʔ (/e/ irreg.)
Rawas	monok	anyut (Keb. monot)	mutah
Tibakang	siok	mamʌn	ŋuteʔ
GLOSS	CHICKEN	DRIFT	VOMIT

PMP	127.	128. *nahik	129. *ni-a
PR	*nak, *taŋ, *lə	*näk	*nə
PBS	*aŋ dbl. *dʌ	*nyumak, dbl. *[g, ŋ]atuh	*niʔəh ~ *nəh
Rawas	taŋ ~ lə	kə-näk	nə
Tibakang	dʌ	nyumak	nəh
GLOSS	AT	CLIMB	3SG.POSS

PMP	130. *niuR	131. *nipis	132. *ni-hu
PR	*nioR	*mipis	*nu
PBS	*buntʌn	*lideʔ	*amuʔ
Rawas	nioa	məlipis	kabən
Tibakang	buntʌn	ledeʔ	amu
GLOSS	COCONUT	THIN	2SG.POSS

PMP	133. *nawa	134. *namuk	135. *huluR
PR	*nyabi (irreg. -b-)	*nyomok	*uluh
PBS	*asəŋ, *nyawa	*piruŋgʌt	*pi-tuhun
Rawas	nyabəy	nyomok	ulua
Tibakang	nyʌwʌy	piruŋgʌt	nulur
GLOSS	SOUL	MOSQUITO	TO LOWER; EXTEND S.T.

PMP	136. *qulej	137. *qapuR	138. *hapuy
PR	*uləj	*kapuh	*upuy
PBS	*urəd	*binyuh	*apuy
Rawas	ulət (Keb. olog)	upua	upuy
Tibakang	ərəd	binyuh	apuy
GLOSS	CATERPILLAR	CHALK, LIME	FIRE

PMP	139. *qutek	140. *Ratus	141.
PR	*u:tək > *utə:k > uta:k	*hotos	*p-adaq
PBS	*[i,ə]ntək	*ratus	*s-ʌnda?
Rawas	utak	otos	p-adah
Tibakang	əntək	ratus	sʌnda?
GLOSS	BRAIN	HUNDRED	SAY

PMP	142. *pajey	143. *panas 'hot' *hapejes 'spicy'	144. *panaw
PR	*paəy	*panəs	*panəw
PBS	*pʌdi	*pʌrʌs	*panu
Rawas	pay	panəs	paniw
Tibakang	pʌdi	pʌrʌs	panu
GLOSS	RICEPLANT	HOT (HEAT)	WALK

PMP	145. *panzan	146. *hepat	147. *pataq
PR	*panjaŋ	*pat	*patiq
PBS	*ambuh	*umpʌt	*p[a,ʌ]ta?
Rawas	panjaŋ	pat	patäh ~ paŋəa
Tibakang	ambuh	umpʌt	pata?
GLOSS	LONG	FOUR	BREAK

PMP	148. *pahak	149. *piliq	150. *paqit
PR	*pahak	*(p)iliq	*pät
PBS	*sindək	*milih	*pa?it
Rawas	kədət (M pa?a?)	päläh/mutia	pät
Tibakang	sindək	milih	pa?it
GLOSS	NEAR	CHOOSE	BITTER

PMP	151. *qapeju	152. *pegeŋ	153. *palaqepaq
PR	*pəgu, *ahəy-ahəy	*goŋ	*pələpaq
PBS	*puduh	*[t,m]agəh	*kilapa[?,h,Ø]
Rawas	pəgəw	goŋ	pələpah
Tibakang	puduh	magəh	kilapa buntʌn
GLOSS	GALL	HOLD	PALMFROND

PMP	154. *penuq	155. *peRes	156. *p-inzem
PR	*pənuq	*pehes	*injəm (contrast *ipen)
PBS	*puno?	*pərəs	*m-inj[ə, ʌ]m
Rawas	pənoh	nəcit	injəm
Tibakang	pono?	pərəs	minjʌm
GLOSS	FULL	SQUEEZE	BORROW

PMP	157. *pisaw	158. *puluq	159. *punay
PR	*pisəw	*puluq	*punəy
PBS	*piso[Ø,?]	*puru?	*puni
Rawas	pisiw	poloh	punuy
Tibakang	piso	simənəŋ duəh puru? 'twenty'	puni
GLOSS	KNIFE	TEN	DOVE

PMP	160. *pandak	161. *pusej	162. *puket
PR	*pəndak ~ pədak	*pusej	*pukət
PBS	*kidəg	*pasid	*puk[ə,ʌ]t
Rawas	pədak	pusət	pokot
Tibakang	kədəg	pasid	jariŋ
GLOSS	SHORT	NAVEL	DRAGNET

PMP	163. *pulut	164. *puqun	165. *putiq *budaq 'foam'
PR	*jala,*pulut	*pun,*bataŋ	*putiq
PBS	*purut db. putək	*pu?un	*buda? dbl. *mupo?
Rawas	pulut	bataŋ	putäh
Tibakang	purut	pu?un kayuh	buda?
GLOSS	BIRDLIME	TREE	WHITE

PMP	166.	167. *kizep	168. *silun
PR	*sahəp	*-kijəp	*sələn
PBS	*ur[ə,ʌ]s	*kisəp dbl. *buləp	*siruh
Rawas	sa?ep	goa ~ kədip	kukəw
Tibakang	urəs	kisəp	siruh
GLOSS	TO LITTER	BLINK	FINGERNAIL

PMP	169. *ma-Ruqanay	170. *sempit/*kepit	171. *silu
PR	*manəy	*səpit	*silu
PBS	*dari[Ø,?]	*sʌmpit	
Rawas	sə-manuy	səpit	siləw
Tibakang	dari-dari	sampit	sasəh
GLOSS	MALE	NARROW	RHEUMATIC PAIN

PMP	172. *qasiRa	173. *sintak	174. *sabuŋ
PR	*sili	*sitak	*soboŋ
PBS	*gulo?		*sabuŋ, *taŋko?
Rawas	ga?em (borr. Malay)	sita?	soboŋ, nyoboŋ
Tibakang	golo?	nyintak	nyabuŋ siok
GLOSS	SALT	JERK	COCKFIGHT

PMP	175. *surat	176. *sapu	177. *susu
PR	*suhət	*supu	*susu
PBS	*surʌt	*sapu, *-adus	*sisoŋ
Rawas	su?ət	supew	kajut 'milk' puan 'nipple'
Tibakang	surʌt	nyapu	sisoŋ
GLOSS	WRITE	BROOM	MILK/BREAST

PMP	178. *tazem	179. *tales	180. *taneq
PR	*tajəm	*taləs	*tanaq
PBS	*rʌja?	*tana?	*tanaq
Rawas	tajəm	keladəy (borr.)	tanah
Tibakang	rʌja?	–	tana?
GLOSS	SHARP	TARO	EARTH

PMP	181. *tanem	182. *taŋan	183. *taqun
PR	*tanəm	*taŋən	*taun
PBS	*puruh	*tʌŋʌn	*sʌwa?
Rawas	tanəm	taŋən	ton
Tibakang	puruh	tʌŋʌn	sʌwa
GLOSS	TO PLANT	HAND	YEAR

PMP	184. *tawaD	185. *teka	186. *tuqelaN
PR	*tawəh (K)	*təko	*tələn
PBS	*tawʌr	*mʌndəg	*turaŋ
Rawas	tawəa cf. libəa	təkaw	tələn
Tibakang	tawʌr	mʌndəg	turaŋ tʌrʌn 'adam's apple'
GLOSS	HAGGLE	COME	BONE

PMP	187. *telu	188. *tinaqi 'stomach'	189. *qateluR
PR	*təlu	*tənəy	*tənol
PBS	*taruh	*na?ih dbl. *putuŋ	*[ø, n]tulo?
Rawas	tələw	tənuy	tənoa ~ tənol
Tibakang	taruh	na?ih cf. *tʌni 'intestines'	tolo?
GLOSS	THREE	STOMACH	EGG

PMP	190. *deŋeR	191. *tanda	192. *takebas
PR	*təŋoa, *tihuk (see EAR)	*tandə	*təbas
PBS	*diŋah	*tanda dbl. *indih	*tʌbʌs dbl. *nʌhu?
Rawas	ti?uk	tandə	təbas
Tibakang	kidiŋah	tanda	nʌhu?/tʌbʌs
GLOSS	HEAR	MARK, SIGN	CLEAR-CUT

PMP	193. *tebaŋ	194. *tektek	195. *tiDuR
PR	*təbaŋ	*tətok	*tiduR (K tiduh)
PBS	*tabəŋ	*kapəg	*bu? əs
Rawas	təbaŋ	tətok	tidua
Tibakang	tabəŋ	kapəg	bə?əs
GLOSS	FELL (TREE)	CHOP, HACK	SLEEP

PMP	196. *taqi	197. *tikam	198.
PR	*təy	*tikəm dbl. *tujaq	*tihuk
PBS	*t[a,ʌ]ki?		*kapiŋ
Rawas	tuy	tujah	ti?uk
Tibakang	taki?	naŋkʌt/nabək	
Mujat	kapiŋ	ŋ-ujak	
GLOSS	FAECES	TO STAB	EAR

PMP	199. *tirus	200. *talih	201. *timba
PR	*tihus	*tili	*timbo
PBS		*tarih	*timb[a,ʌw]
Rawas	cituŋ	tiləy	timbaw
Tibakang	tiruk	tarih	timba ~ timbʌw
GLOSS	TAPERING	ROPE	WELL-PAIL

PMP	202. *timeRaq	203. *timbak	204. *tupelak
PR	*timaq	*timbak	*tulak
PBS	*timah (borr. Ml.)	*timbak	
Rawas	timah	nimbak	tulak
Tibakang	timah	nimbak	numbuk
GLOSS	TIN	TO SHOOT	REJECT

PMP	205. *hiup	206. *tuzuq	207. *tuŋked
PR	*t-iup	*tujuq	*tokot
PBS		*iju? (cog.)	*tuŋkət, *siŋkuhud
Rawas	tiup	tojoh	tokot
Tibakang	ŋompo	iju?	tuŋkət
GLOSS	BLOW	SEVEN	CANE, STAFF

PMP	208. *tupul	209. *tuqah	210. *tutup
PR	*tupul	*tui	*tutup,*təkəp
PBS	*tajə	*tu?uh	*tutu[Ø, p]
Rawas	topol	tuəy (Keb. tui)	tutup
Tibakang	tajə	tu?uh	tutu
GLOSS	DULL, BLUNT	OLD	TO CLOSE

PMP	211. *TukTuk	212. *qubi	213. *quDaŋ
PR	*tutuk	*ubi	*udaŋ
PBS	*ŋumpah	*banduŋ	*andaŋ
Rawas	tutuk	ubəy	udaŋ
Tibakang	ŋəmpəh	banduŋ	andaŋ
GLOSS	POUND RICE	YAM	SHRIMP

PMP	214. *quzan	215. *aku	216. *qulu
PR	*ujən	*uku	*ulu
PBS	*ujʌn	*aku?	*ba?ak 'head'
Rawas	ujən	kəw	uləw
Tibakang	ujʌn	aku	ba?ak
GLOSS	RAIN	1S PRONOUN	HEAD

PMP	217. *Rumaq	218. *busuk	219. *qayam
PR	*humaq	*(b)usuk	*yam(-yam)
PBS	*ramin	*sikəh	*rubi
Rawas	umah	busuk	mainan
Tibakang	ramin	səkəh	birubi
GLOSS	HOUSE	ROTTEN	TOY

PMP	220. *walu	221. *siwa	222.
PR	*dəlapən	*səmilən	*cua, *laŋ
PBS	*mahi	*piri?i	*kai? dbl. kadə?
Rawas	lapən	səmilən	laŋ
Tibakang	mahi	piri?i	kai?
GLOSS	EIGHT	NINE	NOT

PMP	223.	224. *ati	225. *-
PR	*iso, *bukən	*ati	*daŋ, jibaq
PBS	*bukən	*bayuh	*aba?
Rawas	bukən 'chicken paunch'	əlum	jibah
Tibakang	bəkən	bayuh	aba?
GLOSS	NOT A	NOT YET	DON'T

PMP	226. *-	227. *-	228. *-
PR	*may	*nak	*di
PBS	*ndəg	*dʌ dbl. *aŋ	*a?ih
Rawas	may	taŋ ~ lə	dəy
Tibakang	ndəg	dʌ	a?ih
GLOSS	TO	AT	THERE

PMP	229. *-	230. *ŋajan	231. *tahiq/zaqit
PR	*(p)iyə	*gän	*mə-ndät
PBS	*ati?	*gʌnʌn	*ji?it
Rawas	iyə	gän	mə-ndät
Tibakang	ati?	gʌnʌn	nyi?it
GLOSS	HERE	NAME	SEW

PMP	232. *buka'	233. *nipay dbl. hulaR	234. *ma-Raya
PR	*buka?	*nipi	*li
PBS	*buka?	*nipəh dbl. *ulʌr	*rayəh dbl. bahas
Rawas	bukak [buka?]	ular /r/ voiceless	ləy
Tibakang	buka?	nyipəh	rayəh
GLOSS	TO OPEN	SNAKE	BIG

PMP	235. *ma-kapal	236. *si-ia	237. *si-ida
PR	*kʌpa[ø, ?]	*si	*si, *tobo ?ə
PBS		*inya si?en	*balainya
Rawas	kəbol	səy	si ~ tobo ?ə
Tibakang	kʌpa	inya se?e	inya se?en
GLOSS	THICK	HE/SHE	THEY

PMP	238. *apa/anu	239. *i-sai	240. *kua/kuja
PR	*jano dbl. *igän	*api dbl. *sapo (?)	
PBS	*anih	*asih	*muŋ anih
Rawas	igän	(s)apaw ~ apəy (hon.)	ci inan caʔəy nə ~ cinan caʔəy nə (Leb. awəy ipə)
Tibakang	anih	asih	muŋ anih
GLOSS	WHAT?	WHO?	HOW?

PMP	241. *esa/isa	242. *balanak	243. *balaŋa
PR	*do	*bəlanak	*bəlaŋi
PBS	*indiʔ	*bi[r, ∅]anak	*b[a, i]laŋaʔ(-l- irreg.)
Rawas	daw (M do)	bəlanak	bəlaŋəy
Tibakang	indiʔ	baranak	balaŋaʔ
GLOSS	ONE	MULLETFISH	CLAYPOT

References

Adelaar, K. Alexander, 1991, On the internal classification of the Malayic subgroup. Paper delivered at the 4th International Conference on Austronesian Linguistics, Honolulu, 4–8 June 1991.

1992, *Proto Malayic: a reconstruction of its phonology and part of its morphology and lexicon.* Canberra: Pacific Linguistics.

Blust, Robert A., 1981, The reconstruction of Proto Malayo-Javanic: an appreciation. *Bijdragen tot de Taal-, Land- en Volkenkunde* 137/4:456–469.

1984, On the history of the Rejang vowels and diphthongs. *Bijdragen tot de Taal-, Land- en Volkenkunde* 140:422–450.

1990, Linguistic change and reconstruction methodology in the Austronesian language family. In Phillip Baldi, ed. *Linguistic change and reconstruction methodology*, 133–153. Berlin and New York: Mouton.

1992, The Austronesian settlement of mainland Southeast Asia. In Karen L. Adams and Thomas John Hudak, eds *Papers from the second annual meeting of the Southeast Asia Linguistics Society*, 25–83. Tempe, Arizona: Program for Southeast Asian Studies.

n.d., Austronesian comparative dictionary. Electronic manuscript.

Coady, James and Richard McGinn, 1983, On the so-called implosive nasals of Rejang. In Raine Carle et al. eds *Festschrift für Hans Kähler.* GAVA' 17:437–449. Berlin: Dietrich Reimer Verlag.

Court, Christopher, 1967, A distinctive feature analysis of the phonemes of Měntu Land Dayak. *Phonetica* 17:202–207.

Dahl, Otto Christian, 1951, *Malagasy and Maanyan.* Avhandlinger utgitt av Instittet 3. Olso: Egede Instituttet.

Kroeger, Paul R., 1998, Language classification in Sarawak: a status report. *The Sarawak Museum Journal* LIII 74:137–173.

McGinn, Richard, 1997, Some irregular reflexes of PMP vowels in Rejang. *Diachronica* XIV/1:67–106.

1999, The position of the Rejang language of Sumatra in relation to Malay and the 'ablaut' languages of Northwest Borneo. In Elizabeth Zeitoun and Paul Jen-Kuei Li, eds *Selected Papers from the Eighth International Conference on Austronesian Linguistics*, 205–226. Taipei: Symposium Series of Linguistics, Academia Sinica, Number 1.

Ray, Sidney H., 1913, The languages of Borneo. *The Sarawak Museum Journal* 1/4:1–196.

Ruhlen, Merritt, 1987, *A guide to the world's languages.* Stanford: Stanford University Press.

Scott, N.C., 1964, Nasal consonants in Land Dayak (Bukar-Sadong). In David Abercrombie et al., eds *In honour of Daniel Jones*, 432–436. London: Longmans.

Topping, Donald M., 1990, A dialect survey of the Land Dayaks of Sarawak. In James T. Collins, ed. *Language and oral traditions in Borneo. Selected Papers from the first extraordinary conference of the Borneo Research Council, Kuching, Sarawak, Malaysia August 4–9*, 247–274. Williamsburg, VA: The Borneo Research Council Proceedings Series, vol. 2.

The Saluan-Banggai microgroup of eastern Sulawesi

DAVID MEAD

1 Introduction

In this paper, I focus on five languages of eastern Sulawesi, namely Saluan, Bobongko, Andio, Balantak and Banggai (see Map 1). Saluan, Balantak and Banggai are relatively well-known languages, each spoken by some thousands of speakers, while Bobongko and Andio are smaller, sometimes overlooked languages spoken in small enclaves.

In the 19th century and into the 20th, it was usual for the Saluan language to be referred to as Loinan (also spelled Luinan, Loinang, or Loindang). Because Loinan carries certain negative connotations, in recent decades this term has given way to Saluan. Loinan and its variants, and later Saluan, have *also* been used to refer to the group which includes Saluan proper, Balantak, and (when recognised) Bobongko and Andio. Researchers, however, have differed regarding the position of Banggai: some have included it within this group, but others have excluded it. The term 'Saluan' is thus doubly ambiguous, referring either to a single language or a group of languages, which group may or may not include Banggai. To resolve this ambiguity, I use 'Saluan' solely for the language, and 'Saluan-Banggai' when referring to the group.[1]

I thus fall on the side of those who include Banggai in a grouping with Saluan, Bobongko, Andio and Balantak. I base this conclusion on evidence from historical sound change.

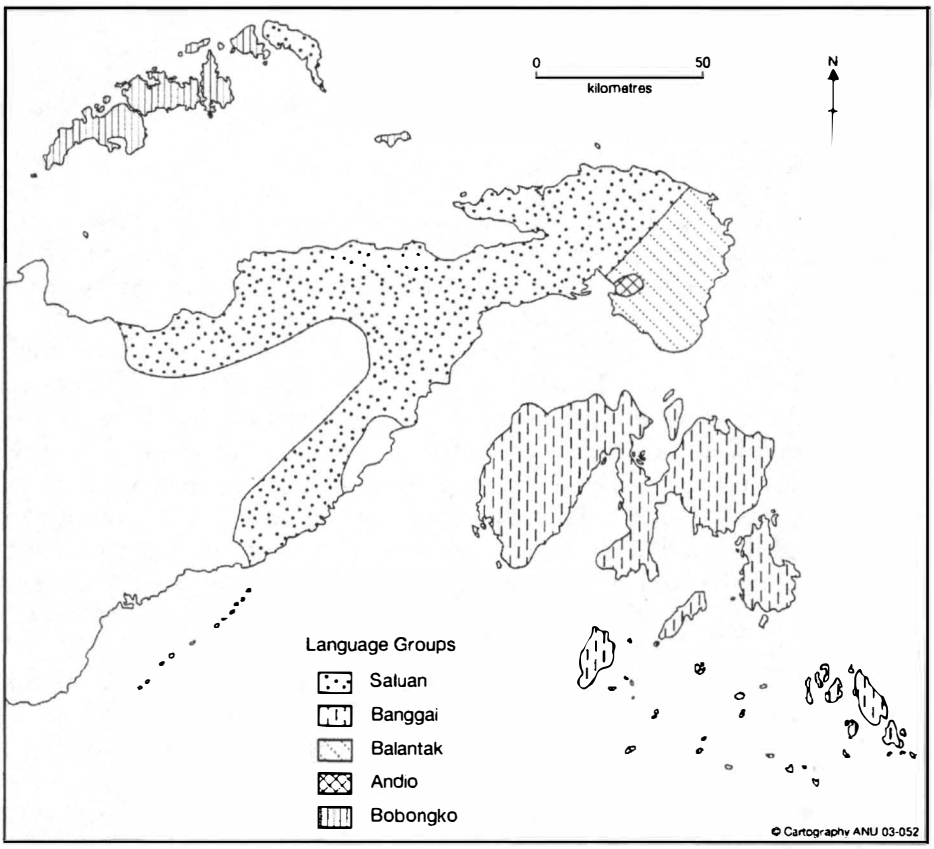
All told, there are twelve sound changes which distinguish Saluan, Bobongko, Andio and Balantak from their Proto Malayo-Polynesian ancestor. Together these changes constitute our best evidence for grouping these four languages into a single microgroup, that is to say a set of languages which share a relatively immediate common ancestor, and which are more closely related to each other than to any surrounding language. At the same time, however, these same twelve changes are also exhibited in Banggai, which perforce must be included in

1 Another reason for not using 'Saluan' as the group label is that Banggai is the largest and most prestigious of the five languages. Formerly the Banggai held sway over the Saluan and Balantak regions and not the other way around, a fact reflected, for example, in that the administrative area where Saluan and Balantak are spoken is named the Banggai Regency.

John Lynch, ed. *Issues in Austronesian historical phonology*, 65–85.
Canberra: Pacific Linguistics, 2003.

any such group. Furthermore, within this group Banggai and Balantak share in two further changes, and in our picture of internal subgrouping these two languages must stand together against a western group composed of Saluan, Bobongko and Andio.

Section 2 discusses previous classifications of the Saluan-Banggai languages. Section 3 lays out the twelve sound changes exhibited by all five Saluan-Banggai languages. Section 4 describes additional sound changes, some of which are exhibited by two or more languages, but none of which is found in all five languages. The implications which these changes have for subgrouping are addressed in §5.



Map 1: The Saluan-Banggai languages

2 Previous classifications of the languages of eastern Sulawesi

In Nicolaus Adriani's seminal work on the language situation of Sulawesi (Adriani & Kruyt 1914), the Loinan language family was presented as consisting of four languages: the Bobongko language spoken on the north coast of Togian Island, the Loinan and Balantak

languages spoken on the mainland of the eastern arm of Sulawesi, and the Banggai language spoken in the nearby archipelago of the same name just to the south.[2]

Unknown to Adriani was the small Andio language, spoken today by about 1700 people in two villages in an otherwise Balantak-speaking area. Although this language came to the attention of linguists two decades ago, through unfortunate circumstances[3] both Barr and Barr (1979) and Sneddon (1983) initially equated this isolect with the Bobongko language spoken in the Togian Islands. Later Bobongko and Andio word lists appeared together in an Indonesian publication (Wumbu et al. 1986), which made it clear that these were different languages, but it remained for Noorduyn (1991a:103, 1991b:140) to bring this to the attention of outside scholars. The affiliation of Andio with Saluan, Bobongko and Balantak has never been questioned.

This is not true of Banggai. Even prior to Adriani, Brandes (1894) had used two primary criteria — the presence/absence of tensed verbs and the so-called reversed genitive (see below) — to classify languages of Sulawesi, and by these criteria Banggai and Saluan fell into separate language families. Although Adriani argued against using either criterion to classify languages (Adriani & Kruyt 1914:284ff., see also Noorduyn 1991b:142), nonetheless Adriani's reasons *for* choosing a particular classification were often impressionistic, leaving subsequent researchers to bolster his claims, or in some cases dispute them. Already by 1938, Adriani's understudy and colleague Esser had given Banggai a position coordinate to his Loinang group, while Salzner (1960), who generally followed Esser in regard to Sulawesi languages, removed Banggai completely from it.[4] Similarly, when Sneddon (1983) compiled his language map of Sulawesi, he placed Banggai as a group-level isolate within his Central Sulawesi supergroup. Later in a discussion concerning Sulawesi microgroups he commented simply, 'Banggai appears to be sufficiently different from other Sulawesi languages to deserve independent status' (Sneddon 1993:2).

The divergent position of Banggai has been confirmed by recent lexicostatistical comparisons in Barr and Barr (1979), Kaseng et al. (1979) and Lauder et al. (2000). According to the Barrs' calculations, for example, while Banggai indeed scored highest in lexical similarity with Saluan (47%), Andio (48%) and Balantak (51%) — compare this to a range from 35 to 45% lexical similarity with all other Central Sulawesi languages — nevertheless, they considered these scores low enough to place Banggai in its own group. My own calculations, based on a comparison of 206 items and displayed in Figure 1, produced lexical similarity scores for Banggai roughly ten percentage points across the board even *lower* than those calculated by the Barrs (Bobongko was not included in their comparison).

---

2 In some 19th century writings one also reads of a Mandono or Mondonu language belonging to this group. The postulation of a Mandono language distinct from Saluan, spoken on the coast opposite Peleng Island, apparently originated with Riedel (1868:44, fn.8). I take the position, also adopted by Holle (1894) and Brandes (1894), that Mandono referred either to a dialect of Saluan or some portion of the Saluan-speaking area. The origin of the term Mandono, and whether it is still used locally today, is a matter for further investigation.

3 First, as reported by Barr and Barr (1979:36), an alternative (but derogatory) name for the Andio language is Bobongko. Second, in the guide to his language map of Sulawesi, Adriani listed the negative term for Bobongko as *andioo* (Adriani & Kruyt 1914:352).

4 Salzner's (1960:14) inclusion, on the other hand, of certain Bajau dialects in his Saluan group can only be regarded as an egregious error.

Bobongko				
53	Saluan			
44	62	Andio		
39	51	66	Balantak	
31	37	37	41	Banggai

Figure 1: Saluan-Banggai lexical similarity matrix

Besides word stock, another oft-cited divergent feature of Banggai is that — unlike other languages of Sulawesi, but like many languages of the Malukus — Banggai speakers make use of the so-called 'reversed' or preposed genitive construction. Compare for example the following expressions for 'the hand (*lima*) of the man (*mian*)'. As noted by Adriani, Banggai actually has three distinct genitive constructions; in all three the possessor precedes the possessum.

Bobongko:	*limanu mian*
Saluan:	*limanu mian*
Andio:	*limanu mian*
Balantak:	*limana mian*
Banggai:	*mian limano, miano limano, miano lima*

Similarly, Banggai speakers also make use of possessed classifiers which precede the possessum, as in *konggu bonua* 'my house'; compare Saluan *bonuangku*, in which the possessive suffix *-ngku* directly follows the possessed noun *bonua* 'house'. Although Adriani downplayed these features as a basis for subgrouping, nonetheless there have been intimations that on such grounds Banggai should be grouped with languages further east (Kanski & Kasprusch 1931).

As I argue below, however, Banggai is a Sulawesi language, and a member of the Saluan-Banggai microgroup. In fact from a consideration of shared sound changes, the primary split of the Saluan group must be between an eastern group composed of Balantak and Banggai on one hand, and a western group comprising Bobongko, Saluan and Andio on the other.

3 Sound changes shared by the Saluan-Banggai languages

In all, present data[5] allow us to reconstruct twelve sound changes which distinguish all five Saluan-Banggai languages from their known common ancestor, Proto Malayo-

---

5 My information about Bobongko comes from my own field notes (Mead in press and in preparation). For Balantak, I consulted a prepublication copy of the *Kamus Balantak–Indonesia–Inggeris* (Balantak–Indonesian–English Dictionary), which is to be published locally in the Balantak area (Bradbury in press). For Banggai I consulted van den Bergh's (1953) *Spraakkunst van het Banggais* (Grammar of Banggai), which also contains a twenty-two page lexicon.

Information about Saluan and Andio was harder to obtain. For word stock I have relied primarily on 488-item word lists collected by Robert Busenitz. His one Andio word list and four Saluan lists were used in a lexicostatistical comparison (Busenitz 1991), but the lists themselves were never published. Robert Brown (2001: pers.comm.) provided me with a similar list for the Saluan isolect spoken in the interior village of Simpang, along with an overview of Saluan dialects. Erik Zobel (2002: pers.comm.) shared notes on Saluan phonology and grammar, for which I am particularly indebted in regard to long vowel reflexes of final **R* and **j*.

Polynesian (PMP). These twelve changes are presented and discussed in turn in the following subsections. Collectively they are our best evidence for establishing, on the basis of historical sound change, a close link between Saluan, Bobongko, Andio and Balantak. On the same evidence, though, Banggai must also be admitted as a member of this group.

3.1 Reduction of consonant clusters

PMP consonant clusters, other than a nasal plus a following consonant, were redued in Saluan-Banggai languages via loss of the first consonant of the cluster. Sequences of nasal plus stop were retained, with the nasal assimilating to the point of articulation of the following stop (if it had a different point of articulation). In Table 1, as throughout this paper, non-cognate forms are underlined, while a dash (—) simply indicates a lack of data. For false cognates, which result from borrowing, I also give the source language, either Pamona (<Pam) or Gorontalic (<Gtl). Metathesis is indicated by (<met.). Language abbreviations are Bgg = Banggai, Bal = Balantak, And = Andio, Sal = Saluan and Bbk = Bobongko.

Table 1: Reflexes of PMP consonant clusters

	'forge' **tuktuk*	'suck' **sepsep*	'rub' **gisgis*	'dark' **demdem*	'horn' **tanduk*
Bgg	*tutuk*	*sosop*	—	*londom/i*	*tanduk*
Bal	*tutuk*	*sosop*	*geges*	*rondom*	*tanduk*
And	—	—	*geges*	—	*tanduk*
Sal	—	*sosop*	*geges/i*	—	*tanduk*
Bbk	—	*sosop/i*	*geges*	*morikoyom*	*tondu* (<Pam)

3.2 Loss of PMP *h

PMP *h is invariably reflected as zero in all Saluan-Banggai languages.

Table 2: Reflexes of PMP **h*

	'fish gills' **hasaŋ*	'head hair' **buhuk*	'two' **duha*	'old' **tuqah*
Bgg	—	*buuk*	*lua*	*langkai*
Bal	*ansang*	*wuuk*	*rua?*	*tu?a*
And	*ansang*	*ulu*	*rua?*	*tu?a*
Sal	*ansang*	*ubak*	*o/hua?*	*tu?a*
Bbk	*ansang*	*bulu*	*doluo* (<Gtl)	*langkai?*

Other sources of information on these languages consist of a brief description of Bobongko word stock (Adriani 1900), a Saluan text with grammar notes (Gobée 1929), a Banggai Holle word list (Stokhof 1985:265–278), a Balantak Holle word list (mislabelled as Banggai) (Stokhof 1985:251–263), and the several lists published in Barr and Barr (1979), Wumbu et al. (1986), and Lauder et al. (2000). I have used these other sources sparingly and with caution because of their frequent errors in transcription, especially regarding vowel quality, vowel length, and glottal stop in word-final position.

3.3 Merger of PMP **d*, **D* and **r*

Although the status of **d* as a valid PMP phoneme is not in doubt, both **D* and **r* have come under question. There is no evidence from Saluan languages that these were distinct protophonemes, as all three (if there were three) became **r* in the ancestor to the Saluan languages (but remained **d* when immediately preceded by a nasal; see Table 1 above). This protophoneme is still reflected as *r* in Bobongko, Andio, and Balantak, but via subsequent changes it became *l* in Banggai and *h* in Saluan (zero in final position).[6]

PMP **z* is also a questionable protophoneme. In one known case PMP **z* appears to have merged with **d/D/r*; in a second case mentioned in §4.5, **z* is reflected the same as PMP **Z*.

Table 3: Reflexes of PMP **d/D/r*

	'near' **zani*	'hear' **DeNeR*	'leaf' **Dahun*	'two' **duha*	'scratch' **karut*
Bgg	*lo/lani*	*longol*	*loon*	*lua*	—
Bal	*ka/rani?*	*rongor*	*roon*	*rua?*	*karut*
And	*a/rani*	*rongo*	*roon*	*rua?*	*karut*
Sal	*o/hani*	*hongo, hongoo*	*hoon*	*ohua?*	*kahut*
Bbk	*i/rani*	*rongo*	*ron*	*doluo* (<Gtl)	*karut*

	'side' **biriŋ*	'prawn' **qudaŋ*	'sew thatch' **pawed*	'tree stump' **tuqeD*	'knee' **tuhud*
Bgg	—	—	*paul*	*tuul*	*tuul*
Bal	*biring*	*urang*	*paur*	*tu?or*	*tuur*
And	*biring* 'ear'	—	—	—	*utur* (<met.)
Sal	*bihing* 'ear'	*uhang*	*pawot* (<Gtl)	*tu?o*	*buku*
Bbk	*biring* 'ear'	*gale*	—	—	*buku*

3.4 Monophthongisation of PMP **-ay* and **-ey*

PMP final diphthongs **-ay* and **-ey* are consistently reflected as *e* in modern Saluan languages (see §4.2 regarding reflexes of PMP **-uy*).

Table 4: Reflexes of PMP **-ay, *-ey*

	'chin' **qaZay*	'use' **pakay*	'die' **patey*	'liver' **atey*	'paddle' **beRsay*
Bgg	*ade*	*pake*	*mate*	*ate*	*bose, bosi*
Bal	*asi*	*pake*	*pate*	*ate*	*bose*
And	*ade*	—	*mate*	*ate*	*bose*
Sal	*aje*	—	*mate*	*ate*	*bose*
Bbk	*aje*	*pake*	*mate*	*ate*	*bose*

6 I am aware of a single datum which suggests that some Saluan dialects might have retained final /h/ in some contexts, namely the recording of *sanggohnyo* 'its name' by Brown (1988:1) (compare Bobongko and Andio *sanggor* 'name' with final *r*). A word list from the same village thirteen years later, however, gives only *sanggo* 'name'.

3.5 Monophthongisation of PMP **-aw* and **-ew*

In a parallel change, PMP **-aw* was monophthongised and is reflected as *o* in all Saluan languages. On scant evidence, the same apparently could be said of PMP **-ew*.

Table 5: Reflexes of PMP **-aw*, **-ew*

	'clear (water)' PPh **linaw*	'above' **babaw*	'rafter' **kasaw*	'odour, stink' **behew*
Bgg	*mo/lino*	*babo*	*kaso*	*boo*
Bal	*mo/lino*	*wawo*	*kaso*	*woo*
And	—	*babo*	*kaso?*	—
Sal	—	*bawo*	*kaso?*	—
Bbk	*mo/lino*	*bafo*	*kaso?*	—

3.6 PMP **j* > **y*

PMP **j* is reconstructed only in medial and final position. It is thought to have been a voiced velar fricative or palatalised velar stop (Blust 1990; Ross 1992). It became **y* in a language ancestral to the Saluan-Banggai languages. It is still reflected as *y* in Banggai, but further weakened in the other languages.

Table 6: Reflexes of PMP **j*

	'sun' **qalejaw*	'sniff' **hajek*	'navel' **pusej*	'snake, worm' **ulej*
Bgg	*oloyo*	*oyok/i* 'kiss'	*pusoy*	*uloy*
Bal	*ilio*	*ook*	*puse*	*ule*
And	*sina*	*ook*	*puse*	*ulo*
Sal	*sina, sinaa*	*ook*	*puso, pusoo*	*ulo, uloo*
Bbk	*dolag*	*ook*	*puso*	*bintana?*

A fuller set of data is considered in §4.2, where reflexes of PMP **j* are taken up again. On the basis of counteradditive reasoning, the change PMP **j* > **y* must have occurred after the monophthongisation of PMP **-ay* and **-ey* > **-e* (see §3.4).

3.7 PMP **e* > **o*

PMP *e is regarded as having been a mid central vowel (schwa). It is regularly reflected as the back rounded vowel *o* in all Saluan languages.

Table 7: Reflexes of PMP **e*

	'brain' **utek*	'deaf' **beŋel*	'low tide' **eti*	'roofing thatch' **qatep*
Bgg	—	*bongol*	*o/oti*	—
Bal	*utok*	*bongol*	*mo/oti*	*atop*
And	*utok*	*bongol*	—	*atop*
Sal	*utok*	*bongol*	*mong/oti*	*atop*
Bbk	*utok*	*bongol*	*mo/oti*	*atop*

3.8 Lowering of PMP **i* preceding final **q*

PMP **i* was lowered to **e* (mid front vowel) preceding final **q*. A parallel lowering of **u* before final **q* also occurred, but only in Banggai and Balantak; see §4.1.

Table 8: Reflexes of PMP **-iq*

	'choose' **piliq*	'white' **ma-putiq*	'seedling' **binehiq*
Bgg	*ile/i*	*moute*	—
Bal	*ruruki*	*mobulak*	*wine?*
And	*pile?/i*	*mobulak*	*bine?*
Sal	*pile?/i*	*mopute?*	*bine?*
Bbk	*pile?/i*	*mopute?*	*bine?*

3.9 Raising of PMP antepenultimate **a*

PMP **a* in antepenultimate position is reflected as *o* in all Saluan languages.

Table 9: Reflexes of PMP antepenultimate **a*

	'fingernail' **kanuhkuh*	'bat' **paniki*	'gall, bile' **qapeju*	'ginger' **laqia*	'spouse' **qasawa*
Bgg	—	*uniki*	*sopot*	—	*osoa/an* 'to marry'
Bal	*ngurun*	*poniki?*	*(s)opoyu?*	*loiya?*	*samba-samba*
And	*konuku*	—	*poyu?*	*loiya?*	—
Sal	*konuku*	*poniki?*	*pou?*	*loiya?*	*osoa*
Bbk	*koñuku*	*raupa*	*opou?*	*moinit*	*osoa*

3.10 PMP **-awa- > *-oa-*

The sequence **-awa-* is reflected as *oa* in all Saluan languages.

Table 10: Reflexes of PMP **-awa-*

	'spider' **lawaq*	'breathe' **ñawa*	'laugh' **tawa*	'spouse' **qasawa*
Bgg	*loa*	*noa*	*kokumbit*	*soa*
Bal	*loa?*	*mi/noa*	*to/toa*	*samba-samba*
And	*loa?*	—	*molomi*	—
Sal	*loa?*	*miki/ñoa*	*kumojo, kumojoo*	*osoa*
Bbk	*loa?*	*mingusa*	*gumeleng*	*osoa*

Otherwise PMP **w* was retained in Balantak, and therefore could not have been lost in Proto Saluan-Banggai: note here Balantak *walu* 'eight' (< PMP **walu*), *wani?* 'bee' (< PMP **wañi*) and *wiwik* 'swing legs or fidget with feet while seated or lying down' (cf. PPh **wigwig* 'shake').[7]

3.11 Devoicing of final stops

Although PMP **d/D* became **r* in final position (§4.3), there is evidence that the other PMP voiced stops, **b* and **g*, were devoiced word-finally. Devoicing of final stops is only weakly attested by the available data, which is limited to one nearly complete cognate set for PMP **-b*, and two partial cognate sets for PMP **-g*. However, I am not aware of any counterexamples to this claim either.[8]

Table 11: Reflexes of PMP **-b, *-g*

	'yawn' **huab*	'current' **seleg*	'shake' PWMP **wigwig*
Bgg	—	*solok*	—
Bal	*ming/oap*	*solok*	*wiwik* 'dangle'
And	*ming/oap*	—	—
Sal	*um/oap*	—	—
Bbk	*ming/oap*	—	—

Since presently neither Balantak nor Banggai allow voiced stops in word-final position (Busenitz & Busenitz 1991:31; van den Bergh 1953:13) — Adriani also makes this claim for Saluan (Adriani & Kruyt 1914:83) — it stands to reason that PMP final voiced stops must have been lost via some process in these languages.[9]

3.12 PMP **q* > glottal stop

It appears that PMP **q* became glottal stop in a language ancestral to the Saluan-Banggai languages. Word-medially and -finally it is reflected as glottal stop in all languages except

---

7 PMP **w* is retained as *u* in Balantak *uanan* 'right' (PMP **wanan*), *kauri* 'left' (PMP **wiRi*) and *uate?*, *mondoluate?* 'earthworm' (cf. PMP **wati*). Banggai has no bilabial approximant (or for that matter bilabial fricative) in its phoneme inventory (van den Bergh 1953). Banggai *alu* 'eight' suggests PMP **w* > ∅, but I have no other examples to confirm this as a pattern. Bobongko has only recently regained a bilabial approximant through borrowing (Mead in press), but Bobongko and Saluan both retain PMP **w* through resegmentation to *u*, at least in initial position; compare Bobongko *ualu*, Saluan *ualu?* 'eight' (< PMP **walu*), Bobongko, Saluan *uani?* 'bee' (< PMP **wañi*). The Saluan terms for 'right' and 'left' given by Gobée (1929:199) are *koanan* and *koii* (thus with loss of **w* from PMP **wanan* and **wiRi*), but Busenitz records the latter term as *kowi*, *kewe?*.

8 Bobongko has *tingkod* 'heel' (PMP **tiked*), but I consider this to be a later borrowing from Gorontalo. The inherited reflex is to be seen in Bobongko, Andio *tengker* 'foot', Saluan *tengke*.

9 Even in Bobongko where voiced stops occur in word-final position, a majority can be attributed to borrowing from Gorontalo (Mead in progress). See further §4.3 and footnotes 8 and 14.

Banggai, which does not have a glottal stop phoneme[10] and which reflects PMP **q* as zero. Although one could say that word-initial **q* was lost in all Saluan languages, more correctly it merged with the phonetic glottal stop which appears preceding vowel-initial stems. For example, van den Bergh (1953:12) writes about Banggai 'de hamza staat aan het begin van eider open aanvangend woord, maar wordt daar nooit geschreven' (glottal stop stands at the beginning of every vowel-initial word, even though it is never written there). Likewise Busenitz and Busenitz (1991:34) claim for Balantak that 'initial vowels may optionally be preceded by a glottal stop — in these instances the glottal stop is non-phonemic'. Following my sources, I do not write glottal stop in initial position.

Table 12: Reflexes of PMP **q*

	'rain' **quZan*	'scale (of fish)' **qunap*	'thigh' **paqa*	'trunk' **puqun*	'earth' **taneq*
Bgg	*udan*	—	—	*uu*	*tano*
Bal	*usan*	*unap*	*paʔa*	*puʔun*	*tanoʔ*
And	*udan*	*unap*	*paʔa*	*puʔu*	*tanoʔ*
Sal	*ujan*	*sonuku*	*paʔa*	*puʔun*	*tanoʔ*
Bbk	*ujan*	*unap*	*paʔa*	*puʔun*	*tanoʔ*

In two cases, PMP **q* appears to have been lost in all languages: in reflexes of **laqia* 'ginger' (see below Table 14 and the related discussion); and in reflexes of **taqun* 'year' — compare Banggai *taum*, Balantak, Saluan and Andio *taun* (Bobongko has *toʔu*, presumably borrowed from Pamona, which has an identical form).

In some cases, Balantak, Andio, Saluan and/or Bobongko have glottal stop in final position where no **q* (or other consonant) is reconstructed in the corresponding PMP etymon. A final glottal stop in such forms must be regarded as an addition, even if the reasons for this addition remain obscure. One of the most striking examples of this is found in the numerals, where Saluan, Andio and Balantak have added a final glottal stop to all numbers which formerly ended in a vowel (conversely, Bobongko has even lost final glottal stop from the number 'ten' where it would be expected to occur). Numerals for all languages are shown in Table 13.

Table 13: Reflexes of PMP numerals

PMP	Bobongko	Saluan	Andio	Balantak	Banggai	
**esa*	*samba?an*	*isaʔ*	*isaʔ*	*isaʔ*	*meeng*	'one'
**duha*	*doluo* (<Gtl)	*ohuaʔ*	*ruaʔ*	*ruaʔ*	*lua*	'two'
**telu*	*totolu*	*totoluʔ*	*toluʔ*	*toluʔ*	*tolu*	'three'
**epat*	*opat*	*opat*	*paat*	*paat*	*sangkap*	'four'
**lima*	*olima*	*olimaʔ*	*limaʔ*	*limaʔ*	*lima*	'five'
**enem*	*onom*	*anom*	*noom*	*noom*	*noom*	'six'
**pitu*	*pitu*	*popituʔ*	*pituʔ*	*pituʔ*	*pitu*	'seven'
**walu*	*ualu*	*ualuʔ*	*waluʔ*	*waluʔ*	*alu*	'eight'
**siua*	*sio*	*osioʔ*	*sioʔ*	*sioʔ*	*sio*	'nine'
**puluq*	*sampulu*	*sampuluʔ*	*sompuluʔ*	*sompuloʔ*	*songulo*	'ten'

[10] Except as a variant enunciation of the phoneme *k*. In the West Peleng dialect the phoneme *k* is mostly articulated as glottal stop, compare *loʔa* 'banana' (elsewhere in Banggai *loka*), *baʔoʔo* 'machete' (elsewhere

In the numerals and in many other cases, it would seem reasonable to consider the addition of final glottal stop to have occurred as a more recent process; that is, no final glottal stop is to be reconstructed at the level of their common ancestor. On the other hand, Table 14 lists some of the forms known to me where, if anywhere, a final glottal stop might be attributed to Proto Saluan-Banggai.[11]

Table 14: Saluan-Banggai forms with addition of final glottal stop

	'branch'	'mouth'	'ginger'	'rice husk'	'gall, bile'
	**paŋa*	**ŋaŋa*	**laqia*	**qeta*	**qapeju*
Bgg	*panga*	*baba*	—	—	*soput*
Bal	*panga?*	*nganga?*	*loiya?*	*ota?*	*(s)opoyu?*
And	*panga?*	*nganga?*	*loiya?*	*ota?*	*poyu?*
Sal	*sampang*	*nganga?*	*loiya?*	*ota?*	*pou?*
Bbk	*pang?*	*nganga?*	*moinit*	*soka?*	*opou?*
	'thorn'	'clothes louse'	'husband'	'termite'	'knife'
	**DuRi*	**tumah*	**la(ŋ)kai*	**anay*	**pisaw*
Bgg	*sulay*, *tadi*	—	*langkai*	*eakan*	*piso*
Bal	*ruri?*	*tuma?*	*langkai?*	*araka*	*piso?*
And	*rii?*	*tuma?*	*langkai?*	*ane?*	*piso?*
Sal	*hii?*	*tuma?*	*langkai?*	*ane?*	*piso?*
Bbk	*dugi?* (<Gtl)	*tuma?*	*langkai?*	*ane?*	*piso?*

Although it is tempting to think that unexpected final glottal stop might mark a word as having been borrowed — compare *piso?* 'knife', possibly borowed from Malay *pisaw* — this cannot be true of every case. The Balantak, Andio and Saluan forms for 'thorn', respectively *ruri?*, *rii?* and *hii?*, exhibit all the regular reflexes of PMP **D* and **R*, and therefore could hardly be the result of borrowing (only Bobongko *dugi?*, with irregular *d* < **D* and *g* < **R*, can clearly be considered a loan). Second, addition of final glottal stop is not characteristic of borrowing in general, at least certainly not in Balantak, the one language for which we have the most complete information concerning word stock.

There may in fact be no unitary explanation for the occurrence of final glottal stop in these forms. PMP **paŋa* 'fork of a branch, bifurcation' and **ŋaŋa* 'mouth' both contain a root which Blust (1988) reconstructs as **ŋa(q)* 'gaping, wide open'. On more marginal grounds, final glottal stop in reflexes of **laqia* 'ginger' could possibly reflect metathesis of PMP **q* to final position.

At any rate, the addition of final glottal stop has all the earmarks of being a local innovation. There appears to be no correlation with any of the laryngeals reconstructed by Zorc (1982, 1996), nor is there any correlation with Kaili-Pamona languages that have retained final glottal stop. Addition of final glottal stop as a local development has also been noted in the Sangir and Sangil languages of North Sulawesi (Sneddon 1984:51–52).

---

in Banggai *bakoko*). Final *k* is sometimes articulated as glottal stop in the eastern dialect of Banggai as well, e.g. *toik*, *toi?* 'sarong' (van den Bergh 1953:10–11).

11 Tables 13 and 14 are not exhaustive, and the reader will note other examples of added final glottal stops in the other tables. Even if we are to attribute added glottal stop to Proto Saluan-Banggai, this putative phoneme must have had a different character from the reflex of PMP **-q*, since only the latter was able to effect the lowering of **i* described in §3.8, and the later lowering of **u* in Balantak and Banggai (§4.1).

4 Other sound changes

In this section, I describe further sound changes which are exhibited by Saluan-Banggai languages. Some of these changes may be regarded as shared innovations; other changes occurred in individual languages. The significance of these changes for an internal classification of the Saluan-Banggai microgroup is taken up in §5.

4.1 Lowering of **u*

In a parallel change to the lowering of **i* (see §3.8), **u* was lowered preceding reflexes of PMP **q* (presumably glottal stop) in Banggai and Balantak.

Table 15: Lowering of **u* in Banggai and Balantak

	'fall' **Nabuq*	'ten' **sa-puluq*	'full' **ma-penuq*	'penis' **lasuq*
Bgg	*tuong*	*songulo*	*moono*	*laso*
Bal	*nawo?*	*sompulo?*	*buke?*	*lao?*
And	*nabu?*	*sompulu?*	*buke?*	*lasu?*
Sal	*nabu?*	*sompulu?*	*buke?*	*lasu?*
Bbk	*nabu?*	*sampulu*	*moponu?*	*tau*

4.2 Reflexes of PMP **j* and **y*

Since PMP **j* and **y* merged as **y* in a language ancestral to the Saluan-Banggai languages, we may consider their reflexes together. Since neither protophoneme has been reconstructed in word-initial position, this leaves only medial and final instances to be considered. Recall furthermore from §3.4 that all instances of PMP **-ay* and **-ey* had been lost via prior monophthongisation to **-e*.

Table 16: Reflexes of PMP **j* and **y*

	'sun' **qalejaw*	'sniff' **hajek*	'field rice' **pajey*	'name' **ŋajan*	'ant' **sejem*
Bgg	*oloyo*	*oyok/i* 'kiss'	*labue*	*sambu*	—
Bal	*ilio*	*ook*	*pae*	*ngaan*	*soom*
And	*sina*	*ook*	*pae*	*ngaan*	—
Sal	*sina*	*ook*	*pae*	*sanggo*	*soom*
Bbk	*dolag*	*ook*	*pae*	*sanggor*	*loog*

	'inside' **qunej*	'snake, worm' **ulej*	'navel' **pusej*	'fly' **lalej*	'palm, sole' **palaj*
Bgg	*lalong*	*uloy*	*pusoy*	*poos*	*palalap*
Bal	*lalom*	*ule*	*puse*	*laale*	*palaa*
And	*rarom*	*ulo*	*puse*	*laalo?*	*pala*
Sal	*uno, unoo*	*ulo, uloo*	*pusoo*	*lalo, laloo*	*palaa*
Bbk	*uno*	*bintana?*	*puso*	*lalo*	*pala*

	'fire' **hapuy*	'pig' **babuy*	'swim' **ŋaŋuy*	'withered' **layu*	'shy, ashamed' **ma-heyaq*
Bgg	*bilat*	*babui*	*kayok*	*mo/loyu*	*ma/maa*
Bal	*apu*	*bau?*	*langu*	*ma/lau*	*maka/maa?*
And	*apu*	*bau*	*langu*	—	—
Sal	*apu, apuu*	*bau, bauu*	*langu, languu*	—	—
Bbk	*apu*	*bau*	*langoy* (<Gtl)	—	—

As emerges from Table 16, the usual reflexes of **y* (< PMP **y*, **j*) are as follows. In Banggai **y* is reflected as *y* (and as *i* in Banggai *babui* 'pig'; Banggai *mamaa* 'shy' is an exception). In Balantak, **-oy* (< PMP **-ej*) is reflected as *e*, otherwise this phoneme was lost without vowel changes (Balantak *ilio* 'sun' is an exception). In Saluan, Bobongko and Andio, PMP **-j-* was lost without vowel changes; we might assume the same fate befell reflexes of PMP **-y-*, but unfortunately present data are too scant to confirm this.[12] Finally in Saluan, Bobongko and Andio, word-final **y* (from either source) was lost, except that some Saluan dialects reflect **-y* as length on the final vowel.

4.3 Reflexes of PMP **R*

PMP **R* is reflected as *r* in Balantak and *l* in Banggai (via **r*), but was usually lost in Saluan, Andio and Bobongko. Only Saluan has a non-zero reflex and then only in word-final position, where (in at least some dialects) PMP **R* is reflected as vowel length.

Table 17: Reflexes of PMP **R*

	'rib' **Rusuk*	'vein, tendon' **uRat*	'heavy' **beReqat*	'male' **maRuqanay*
Bgg	*lusuk*	*neet*	*ba/balat*	*malane*
Bal	*rusuk*	*urat*	*ma/rawat (<met.)*	*moro/one*
And	*usuk*	*uat*	*ma/buat*	*ma/ane*
Sal	*usuk*	*uat*	*ma/boat*	*mo/ane*
Bbk	*usuk*	*ugat* (<Gtl)	*ma/boat*	*mo/ane*
	'water' **waiR*	'hear' **DeŋeR*	'dry' **tuquR*	'coconut' **ñiuR*
Bgg	*paisu*	*longol*	*mo/tuul*	*potil*
Bal	*weer*	*rongor*	*mo/tu?or*	*nuur*
And	*ue*	*rongo*	*mo/tu?u*	*niu*
Sal	*ue, uee*	*hongo, hongoo*	*mo/ti?i*	*niu, niuu*
Bbk	*ue*	*rongo*	*mogangu?*	*bango?*

12 Perhaps to be added to the list in Table 16 would be Bobongko, Saluan *kau?*, Andio, Balantak *kau*, and Banggai *kau*, *kayung* 'wood', on the assumption that PMP **kahiw* was resegmented to **kayu* after loss of **h*. This interpretation is supported by data from Bungku-Tolaki languages, viz. Bungku, etc. *keu* and Tolaki *kasu* (with fortition of **y*). However, the possibility cannot be ruled out that PMP **kahiw* became PSal **kau*, with Banggai *kayung* a subsequent borrowing from Malay (compare van den Berg 1991:6).

Other examples of the fate of word medial PMP **-y-* in Balantak are *aam* 'animal' (< PMP **qayam* 'domesticated animal'), *lua?* 'vomit' (< PMP **luya*) and — with retention of *y* — *layang* 'fly' (< PMP

Considering the data on the fate of **y* presented in §4.2, this strongly suggests that **R* went through a **y*-stage in Saluan, certainly in final position. The most elegant way of capturing this would be to say that **R* became **y* everywhere, followed by **y* > *Ø-Ø-V: (where V: indicates a lengthened vowel). However, there is less evidence to suggest that **R* went through a **y*-stage before being lost word-medially, and there is no evidence for it in word-initial position, since neither PMP **y* nor **j* have been reconstructed word-initially. I therefore propose the following (but less elegant) set of changes, which occurred in the language ancestral to Saluan, Bobongko and Andio:

**R* > Ø-Ø-*y*

**y* > -Ø-V:

Lengthened vowels were then lost as a drift-like tendency in Bobongko, Andio, and certain of the Saluan dialects.[13]

Although Bobongko reflects PMP **R* as zero in inherited forms, in many cases Bobongko has *g* via borrowing from Gorontalo (or some other Gorontalo-Mongondow language). As illustrated in the following examples, often semantic shifts or other unexpected segmental reflexes give evidence that such forms are loans (Gorontalo and Proto Gorontalo-Mongondow (PGM) forms are from Usup 1986).

Bobongko *dugu?* 'blood', Gorontalo *duhu*, PGM **dugu?*, from PMP **ZuRuq* 'sap, juice, gravy'. Both the change **Z* > *d* and the semantic shift to 'blood' mark this form as borrowed (see further Blust 1991). Compare also Bobongko *juu?* 'honey' via direct inheritance.

Bobongko *kugito?* 'octopus', PMP **kuRita.* Raising of final **a* > *o* is regular in Gorontalo and Buol (Sneddon & Usup 1986).

Bobongko *patig* 'sandbank', Gorontalo *patihu*, PGM **pasig*, PMP **pasiR.* PGM **s* > *t* is regular in Gorontalo, Buol and Suwawa (Sneddon & Usup 1986)

Bobongko *toga?* 'dammar', Gorontalo *tohe*, PGM **soga?* 'lamp', PMP **seRaq* 'fire, roast'. PGM **s* > *t* is regular in Gorontalo, Buol and Suwawa (Sneddon & Usup 1986)

Bobongko *dagat* 'sea', Gorontalo *deheto*, PGM **dagat* 'sea', PMP **daRat* 'surface (of sea or land)'. Initial **d* > *d* is regular in Gorontalo-Mongondow.

Bobongko *dugi?* 'thorn', Gorontalo *duhi*, PGM **dugi*, PMP **DuRi.* Initial **D* > *d* is regular in Gorontalo-Mongondow.

Bobongko *monugang* 'parent/child-in-law', Gorontalo *moluhango*, from a derived form of PMP **tuRaŋ.* Other Saluan languages have reflexes of Proto Celebic **panianan* 'parent-in-law', **manian* 'child-in-law'.

**layaŋ*). Unfortunately data which might give us a clearer picture of the fate of PMP **-y-* in the other Saluan-Banggai languages are are not available to me.

13 The dialect situation in Saluan has never been properly clarified, and thus it is not possible to give an account of which Saluan dialects have long vowel reflexes, and which do not. Zobel (pers. comm.) notes that long vowel reflexes must be fairly widespread, since they are found in both Nambo on the southern coast and in Bunta on the northern coast. Long vowel reflexes are not indicated on my word list from Simpang village (Robert Brown 2001:pers. comm.), and are sometimes indicated, sometimes not, on Busenitz's four Saluan word lists (sometimes even on the same word list an item has two responses indicated, one with and one without long vowel).

Bobongko *tuag* 'answer', Gorontalo *tuahu*, PGM **tubag*, PPh **tubaR*. Buol and Gorontalo exhibit loss of PGM **-b-* in this form.

Although today Bobongko speakers are in contact with Gorontalo speakers who reside in or periodically visit the Togian Islands, sound changes which have occurred in Gorontalo make it clear that these cannot be recent borrowings. The implications for prehistorical relationships are not explored here, but clearly Bobongko represents another case of a 'non-*g*' language that has borrowed extensively from a language where **R* > *g*; see especially Blust (1991:89ff.).[14]

4.4 Weakening of **b*

As exemplified by the following cognate sets, PMP **b* became a bilabial approximant (represented orthographically as *w*) both initially and medially in Balantak. Intervocalic *w* has sometimes further weakened to zero, particularly when one of the continguous vowels is *u*.

Table 18: Lenition of **b* in Balantak

	'bone' **buku*	'buaq' **fruit*	'stone' **batu*	'fall' **nabuq*	'ashes' **qabu*
Bgg	*buku*	*sao*	*batu*	*tuong*	*abuk*
Bal	*wuku*	*woo?*	*watu*	*nawo?*	*awu, au* 'dust'
And	*buku*	*bunga*	*batu*	*nabu?*	*abu*
Sal	*buku*	*bua?*	*batu*	*nabu?*	*abu?*
Bbk	*buku*	*bua?*	*batu*	*nabu?*	*agu?*

There are, however, important exceptions where **b* is retained as *b* in Balantak, among them *bolian* 'shaman' (compare PWMP **balian*), *bitu?on* 'star' (PMP **bituqen*), *bose* 'paddle' (PMP **beRsay*), *bunga* 'flower' (PMP **buNa*), *bau?* 'pig' (PMP **babuy*), *toobuan* 'wasp' (PMP **tabu-an*), *banang* 'string' (PMP **benaŋ*), and *bisul* 'boil' (PMP **bisul*). Borrowing, even from another Saluan language, could account for these forms. The parallel Banggai form *banaang* 'string' (with geminate vowel in the final syllable) makes a strong case for borrowing from Malay *benang*. The fact that dialectally Saluan has both *bisun*, *bisul* 'boil' argues that the Saluan form (and hence the Balantak form) was likewise borrowed from an outside language.

---

14 In all I have accumulated only seven examples of PMP **R* > ∅ in Bobongko, versus twenty examples of PMP **R* reflected as *g* via borrowing. Other instances of borrowing not given in the main text are:

Bobongko *bagu* 'new', Gorontalo *bohu*, PGM **bagu*, PMP **baqeRu*
Bobongko *bagu* 'hibiscus', PGM **bagu*, PMP **baRu*
Bobongko *dagum* 'needle', PGM **dagum*, PMP **ZaRum*
Bobongko *dugian* 'durian', PMP **DuRi-an*
Bobongko *kolofigi* 'left', Gorontalo *?oloihi*, PGM **koloigi*, PMP **wiRi*
Bobongko *bagang* 'molar', PGM **bagaŋ*, PMP **baReqaŋ*
Bobongko *bogaani* 'brave', Gorontalo *buheli*, PGM **bogani*, PMP **beRani*
Bobongko *origi?* 'house post', Gorontalo *wolihi*, PMP **ha-DiRi*
Bobongko *lindug* 'earthquake', Gorontalo *liluhu*, PGM **linug*, PMP **linuR*
Bobongko *layag* 'sail', Gorontalo *layahu*, PGM **layag*, PMP **layaR*
Bobongko *tumolog* 'drip, dribble', Gorontalo *tolohu*, PGM **solog*, PMP **seleR* (or **seleg*)

Another set of data gives evidence that intervocalic **b* weakened (and was often lost) in Saluan and Bobongko. This change is most strongly associated with another **b* elsewhere in the word, though lenition is also to be noted in reflexes of **tubuq* 'live' and **tebuh* 'sugarcane'. Orthographic *f* in Bobongko represents a voiceless bilabial frictative [ɸ], which also has allophones [β] and [h] (Mead in press).

Table 19: Lenition of **b* in Saluan and Bobongko

	'lips' **bibiR*	'all' PSal **ibi?*	'above' **babaw*	'carry' **baba*	'pig' **babuy*
Bgg	*bibil*	*saisa/ibi/no*	*babo*	*baba*	*babui*
Bal	*wewer*	*wiwi?/na*	*wawo*	*wawa*	*bau?* (***wau*)
And	*bibi*	*biibi?/no*	*babo*	*baba*	*bau* (***babu*)
Sal	*biwi, biwii*	*iwi?iwi?*	*bawo*	*boa*	*bau, bauu*
Bbk	*bifi*	*ifiifi?*	*bafo*	*boa*	*bau*

	'fish trap' **bubu*	'female' **ba-b[in]ahi*	'live' **tubuq*	'sugarcane' **tebuh*
Bgg	—	*boine*	*tubo*	—
Bal	*wuwu?*	*wiwine*	*tuo?*	*tombong*
And	—	*bobine*	*tubu?*	*umpol*
Sal	*buu?*	*boune*	*tuu?*	*tumbo?*
Bbk	*buu?*	*boune*	*tuu?*	*tou?*

4.5 Depalatalisation of **Z*

PMP **Z* became *s* in Balantak and *d* in Banggai and Andio. It remained a palatal in Saluan and Bobongko (orthographic *j* represents a palatal affricate). There is one known case of PMP **z* patterning with **Z*.

Table 20: Depalatalisation of PMP **Z*

	'chin' **qaZay*	'rain' **quZan*	'road' **Zalan*	'far' **Zauq-en*	'grain' **zelay*
Bgg	*ade*	*udan*	*loloon*	*o/doon*	—
Bal	*asi*	*usan*	*salan*	*o/loa*	*sole?*
And	*ade*	*udan*	*dalan*	*ma/do?on*	*dole?*
Sal	*aje*	*ujan*	*jalan*	*ma/jo?on*	*binde*
Bbk	*aje*	*ujan*	*jalan*	*o/jo?on*	*binte?*

4.6 Depalatalisation of **ñ*

In the same languages in which **Z* was depalatalised, PMP **ñ* also became *n*. In Saluan and Bobongko, **ñ* was depalatalised only preceding **i*.

Table 21: Depalatalisation of PMP **ñ*

	'weave' **añam*	'breathe' **ñawa*	'3SG POSS' **(n)i-a*	'coconut' **ñiuR*	'bee' **wañi*
Bgg	—	*noa*	*-no*	*potil*	—
Bal	*anam*	*mi/noa*	*-na*	*nuur*	*wani?*
And	*anam*	—	*-no*	*niu*	—
Sal	*añam*	*miki/ñoa*	*-ño*	*niu*	*uani?*
Bbk	*añam*	*mingusa*	*-ño*	*bango?*	*uani?*

4.7 Regressive assimilation of **u* > *i*

In four known cases, **u* in the penultimate syllable regressively assimilated to **i* in the ultimate syllable. Although other vowel changes and assimilations occur in the data, this particular change is notable in that it is reflected in Andio, Saluan and Bobongko, but not in Balantak or Banggai.

Table 22: Regressive assimilation of **u* > *i* in Andio, Saluan and Bobongko

	'skin' **kulit*	'hide' **buni*	'vagina' **puki*	'thorn' **DuRi*
Bgg	*kulit*	—	*uki*	*sulay, tadi*
Bal	*kuang*	*wuni*	*pepek*	*ruri?*
And	*kilit*	*bini*	*piki*	*rii?*
Sal	*kilit*	—	*piki*	*hii?*
Bbk	*kilit*	*bini*	*piki*	*dugi?* (<Gtl)

4.8 Changes in individual languages

Morpheme-initial **p* was sporadically lost in Banggai. Although loss of **p* has been noted most frequently preceding *u*, no conditioning environment can be clearly stated. The following cognate sets have appeared in other tables, but it is worth bringing them together here.

Table 23: Loss of initial **p* in Banggai

	'ten' **sa-puluq*	'trunk' **puqun*	'white' **ma-putiq*	'vagina' **puki*
Bgg	*sangulo*	*uu*	*moute*	*uki*
Bal	*sompulo?*	*pu?un*	*mobulak*	*pepek*
And	*sompulu?*	*pu?u*	*mobulak*	*piki*
Sal	*sampulu?*	*pu?un*	*mopute?*	*piki*
Bbk	*sampulu*	*pu?un*	*mopute?*	*piki*

	'full' **ma-penuq*	'bat' **paniki*	'choose' **piliq*
Bgg	*moono*	*uniki*	*ile/i*
Bal	*buke?*	*poniki?*	*ruruki*
And	*buke?*	—	*pile?/i*
Sal	*buke?*	—	*pile?/i*
Bbk	*moponu?*	*raupa*	*pile?/i*

It is presently unclear whether the loss of **p* was an irregular but independent development in Banggai, or whether it reflects areal influence from Moluccan languages where lenition and often loss of **p* is regular (Adriani & Kruyt 1914:290, 291; Collins 1983:114ff.). Irregular loss of **p* has also been noted in Buginese (Mills 1975:247; Adelaar 1994:12).

Three other changes in individual languages which have been treated incidentally above include **r* > *h-h-Ø* in Saluan (see §3.3), **r* > *l* in Banggai (see §3.3 and §4.3), and the loss of glottal stop in Banggai (see §3.12).

5 Conclusion: shared changes and internal classification

Based on the historical sound changes outlined above, I propose a subgrouping for the Saluan-Banggai languages as shown in Figure 2. This diagram incorporates the changes which together distinguish these languages from Proto Malayo-Polynesian (see §3), and the changes which differentiate these languages from one another (see §4).

From Figure 2 it should be noted that some sound changes must have occurred independently in different branches of Saluan-Banggai. For example, both Andio and Banggai share in the change **Z* > *d*, but I do not regard this change as having occurred during (or as being indicative of) a period of common development. Rather, it occurred independently in both languages, after the break-up of their common ancestor. Were we to assume that **Z* > *d* was a shared innovation, then Andio and Banggai would constitute a first-order subgroup that excludes Balantak (where **Z* > *s*). In that case, however, the changes **R* > **r* and **-u?* > **-o?* in Balantak and Banggai would have to be regarded as parallel (rather than shared) innovations, which in my opinion is a less desirable assumption. By their nature, subgrouping arguments are to some degree circular — our view of shared sound changes both determines and is determined by our subgrouping assumptions — and arguments often hinge on which changes we are willing to posit happened only once, and which happened twice or more.

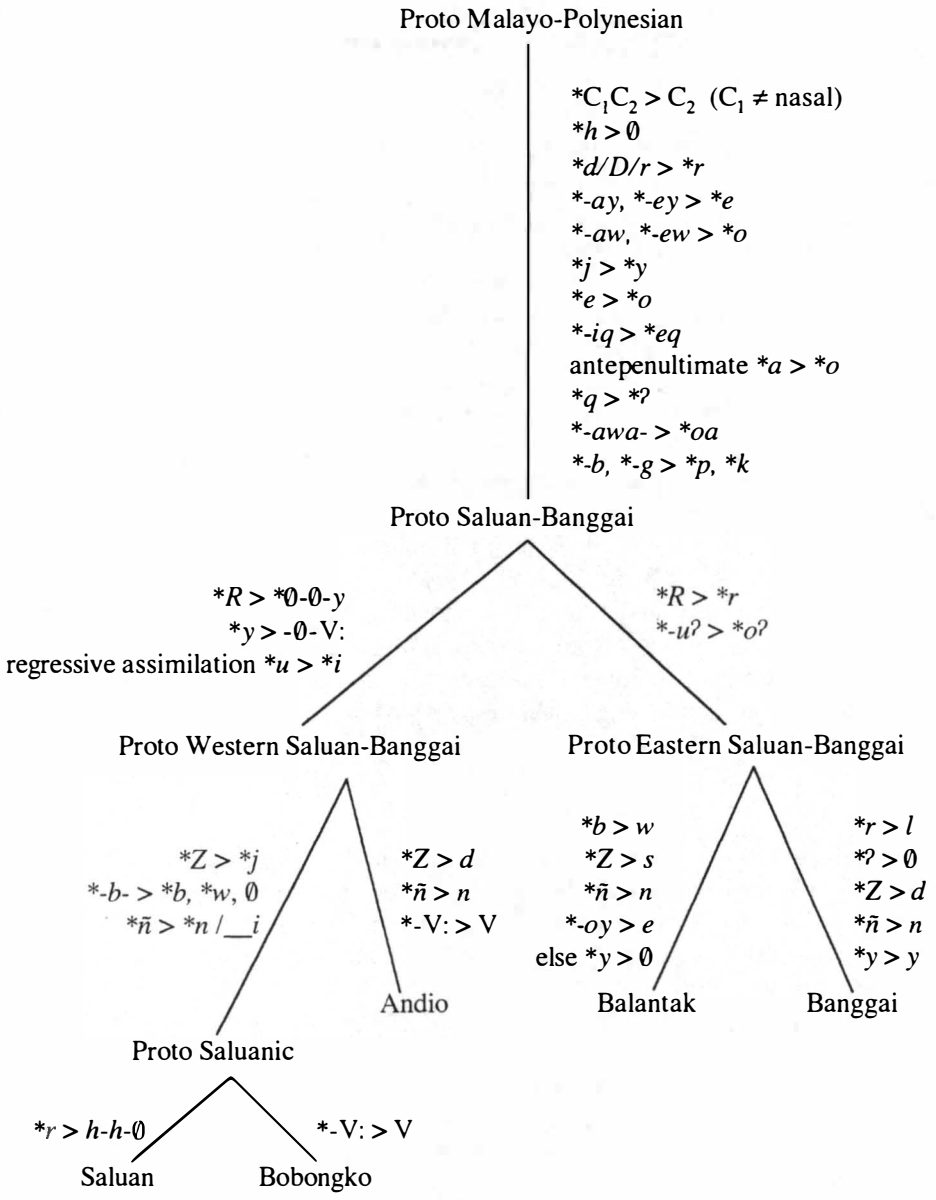


Figure 2: Sound changes in the Saluan-Banggai microgroup

Similarly, although I list twelve changes which distinguish the Saluan-Banggai languages from their common ancestor, Proto Malayo-Polynesian, there is no *a priori* reason to assume that they are all reflective of a period of common development. Loss of PMP **h*, rhotacisation of PMP **d*, PMP **q* becoming glottal stop and even the backing of PMP **e* (schwa) to **o* are found frequently enough elsewhere in Austronesia that such changes could have occurred across the Saluan-Banggai area after the break-up of their common ancestor.

On the other hand, the changes PMP **j* > **y*, PMP **-iq* > **eq*, and PMP **-awa-* > **oa* are sufficiently unusual that it seems best to consider them to be innovations which occurred during a period of common development. The first two of these changes, in fact, are shared by other Sulawesi languages, and are useful for postulating higher-order groupings between the level of Proto Malayo-Polynesian and Proto Saluan-Banggai. The significance of these twelve changes for macrogrouping, then, cannot be fully appreciated without also comparing the Saluan-Banggai languages with neighbouring languages on the island of Sulawesi. While the present paper has laid the groundwork for such a study, I do not make that comparison here. Rather I have made it the topic of a separate paper which also appears in this volume.

This study, then, is only a first look at historical sound change among languages of eastern Sulawesi, and falls short of a complete comparative study in at least three other ways. First, some vowel changes, primarily assimilations, which are localised to particular languages have not been treated. Second, I have not touched on lexical, semantic or grammatical innovations which could bolster the evidence from sound change. Finally, although I consider Proto Saluan-Banggai to be a valid protolanguage, I have not included here any Proto Saluan-Banggai reconstructions.

Nonetheless, the principal results of this study still stand. I have laid out the major sound changes which distinguish the Saluan, Bobongko, Andio, Balantak and Banggai languages from Proto Malayo-Polynesian, and I have also shown that there is no basis in historical sound change for excluding Banggai from a Saluan-Banggai group. Indeed, in the internal classification of these languages, Banggai and Balantak group closely together, and stand apart from an eastern group comprising Saluan, Bobongko and Andio.

References

Adelaar, K. Alexander, 1994, The classification of the Tamanic languages. In Tom Dutton and Darrell T. Tryon, eds *Language contact and change in the Austronesian world*, 1–42. Trends in Linguistics, Studies and Monographs 77. Berlin: Mouton de Gruyter.

Adriani, N., 1900, De Talen der Togian-Eilanden. *Tijdschrift voor Indische Taal-, Land- en Volkenkunde* 42:428–490, 539–566.

Adriani, N. and A.C. Kruyt, 1914, *De Bare'e-sprekende Toradja's van Midden Celebes*, vol. 3: *Taal- en letterkundige schets der Bare'e-taal en overzicht van het taalgebied Celebes–Zuid-Halmahera.* Batavia: Landsdrukkerij.

Barr, Donald F. and Sharon G. Barr, in co-operation with C. Salombe, 1979, *Languages of Central Sulawesi: checklist, preliminary classification, language maps, wordlists.* Ujung Pandang: UNHAS-SIL.

Berg, René van den, 1991, Muna historical phonology. In Sneddon, ed. 1991:1–28.

Bergh, J.D. van den, 1953, *Spraakkunst van het Banggais.* 's-Gravenhage: Martinus Nijhoff.

Blust, Robert, 1988, *Austronesian root theory.* Studies in Language Companion Series 19. Amsterdam: John Benjamins.

1990, Summary report: linguistic change and reconstruction methodology in the Austronesian language family. In Philip Baldi, ed. *Linguistic change and reconstruction methodology.* Trends in Linguistics, Studies and Monographs 45, 133–153. Berlin: Mouton de Gruyter.

1991, The Greater Central Philippines hypothesis. *Oceanic Linguistics* 30:73–129.

Bradbury, Daniel, ed., in press, *Kamus Balantak–Indonesia–Inggeris.* Luwuk: Kantor Bupati Kabupaten Banggai.

Brandes, J.L.A., 1894, Toelichting tot de schetstaalkaart van Celebes samengesteld door den heer K.F. Holle. *Notulen van het Bataviaasch Genootschap van Kunsten en Wetenschappen* 32:xv–xxxiv.

Brown, Cecilia, 1988, Saluan syntax features. Typescript, 6pp.

Busenitz, Robert L., 1991, Lexicostatistic and sociolinguistic survey of Balantak and Andio. In Timothy Friberg, ed. *More Sulawesi sociolinguistic surveys, 1987–1991.* Workpapers in Indonesian Languages and Cultures 11, 1–17. Ujung Pandang: SIL.

Busenitz, Robert L. and Marilyn J. Busenitz, 1991, Balantak phonology and morphophonemics. In Sneddon, ed. 1991:29–47.

Collins, James T., 1983, *The historical relationships of the languages of central Maluku, Indonesia.* Canberra: Pacific Linguistics.

Esser, S.J., 1938, Talen. *Atlas van Tropisch Nederland*, sheet 9b. Amsterdam: Koninklijk Nederlandsch Aardrijkskundig Genootschap.

Gobée, E., 1929, Een Loinansch verhaal. In Walther Aichele, ed. *Feestbundel uitgegeven door het Koninklijk Bataviaasch Genootschap van Kunsten en Wetenschappen bij gelegenheid van zijn 150-jarig bestaan 1778–1928*, vol. 1, 187–201. Weltevreden: Kolff.

Holle, K.F., 1894, *Schets-taalkaart van Celebes.* Map, 1:2000000.

Kanski, P. and P.A. Kasprusch, 1931, Die indonesisch-melanesischen Übergangssprachen auf den Kleinen Molukken. *Anthropos* 26:883–890.

Kaseng, Syahruddin, Masjhuddin Masjhuda, Abdul Muthalib, Indra Bangsawan Wumbu, Amir W. Lumentut, Amir Kadir and Abdul Latif Rozali, 1979, *Bahasa-bahasa di Sulawesi Tengah.* Jakarta: Pusat Pembinaan dan Pengembangan Bahasa.

Lauder, Multamia R.M.T., Ayatrohaedi, Frans Asisi Datang, Hans Lapoliwa, Buha Aritonang, Ferry Feirizal, Sugiyono, Non Martis, Wati Kurniawati and Hidayatul Astar, 2000, *Penelitian kekerabatan dan pemetaan bahasa-bahasa daerah di Indonesia: Propinsi Sulawesi Tengah.* Jakarta: Pusat Pembinaan dan Pengembangan Bahasa.

Mead, David, in press, A preliminary sketch of the Bobongko language. In Wyn Laidig, ed. *Studies in Sulawesi linguistics, part 7.* NUSA Linguistic Studies of Indonesian and Other Languages in Indonesia. Jakarta: Universitas Katolik Indonesia Atma Jaya.

in preparation, A brief Bobongko lexicon. To appear in SIL Electronic Data Series.

Mills, Roger F., 1975, Proto South Sulawesi and Proto Austronesian Phonology. PhD dissertation, University of Michigan.

Noorduyn, J., 1991a, *A critical survey of studies on the languages of Sulawesi.* KITLV Bibliographic Series 18. Leiden: KITLV Press.

1991b, The languages of Sulawesi. In H. Steinhauer, ed. *Papers in Austronesian linguistics* No.1, 137–150. Canberra: Pacific Linguistics.

Riedel, J.G.F., 1868, *Bijdragen tot de kennis der talen en dialekten, voorkomende op de eilanden Luzon of Lesoeng, Panai of Iling-Ilong, alsmede op Noord- en Midden-Celebes.* Verhandelingen van het Bataviaasch Genootschap van Kunsten en Wetenschappen, 33/2. Batavia: Lange.

Ross, Malcolm D., 1992, The sound of Proto-Austronesian: an outsider's view of the Formosan evidence. *Oceanic Linguistics* 31:23–64.

Salzner, Richard, 1960, *Sprachenatlas des Indopazifischen Raumes*. Wiesbaden: Otto Harrassowitz.

Sneddon, James N., 1983, Northern Celebes (Sulawesi). In Stephen A. Wurm and Shirô Hattori, eds *Language atlas of the Pacific area*, part 2: *Japan area, Taiwan (Formosa), Philippines, mainland and insular South-east Asia*, map 43. Canberra: The Australian National University, Australian Academy of the Humanities and The Japan Academy.

1984, *Proto-Sangiric and the Sangiric languages*. Canberra: Pacific Linguistics.

1993, The drift towards final open syllables in Sulawesi languages. *Oceanic Linguistics* 32:1–44.

1991, *Studies in Sulawesi linguistics*, part 2. NUSA Linguistic Studies of Indonesian and other Languages in Indonesia vol.33. Jakarta: Universitas Katolik Indonesia Atma Jaya.

Sneddon, James N., ed., 1991, *Studies in Sulawesi linguistics, part 2* NUSA Linguistic Studies of Indonesian and Other Languages in Indonesia vol. 33. Jakarta: Universitas Katolik Indonesia Atma Jaya.

Sneddon, James N. and Hunggu Tadjuddin Usup, 1986, Shared sound changes in the Gorontalic language group: implications for subgrouping. *Bijdragen tot de Taal-, Land- en Volkenkunde* 142:407–426.

Stokhof, W.A.L., ed., 1985, *Holle lists: vocabularies in languages of Indonesia*, vol. 7/4: *South-East Sulawesi and neighbouring islands, West and North-East Sulawesi.* Canberra: Pacific Linguistics.

Usup, Hunggu Tadjuddin, 1986, Rekonstruksi proto-bahasa Gorontalo-Mongondow. PhD dissertation, Universitas Indonesia.

Wumbu, Indra Bangsawan, Amir Kadir, Nooral Baso and Sy. Maranua, 1986, *Inventarisasi bahasa daerah di propinsi Sulawesi Tengah*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa.

Zorc, R. David, 1982, Where, o where, have the laryngeals gone? Austronesian laryngeals re-examined. In Amram Halim, Lois Carrington and S.A. Wurm, eds *Papers from the Third International Conference on Austronesian Linguistics*, vol. 2: *Tracking the travellers*, 111–144. Canberra: Pacific Linguistics.

1996, The reconstruction and status of Austronesian glottal stop — chimera or chameleon? In Bernd Nothofer, ed. *Reconstruction, classification, description — festschrift in honor of Isidore Dyen*, 41–72. Hamburg: Abera-Verlag.

Mead, D. "The Sahan-Banggai microgroup of eastern Sulawesi". In Lynch, J. editor, *Issues in Austronesian historical phonology*.
PL-550:65-86. Pacific Linguistics, The Australian National University, 2003. DOI:*10.15144/PL-550.65*

The place of Tukang Besi and the Muna-Buton languages

RENÉ VAN DEN BERG

1 Introduction

Until recently, the linguistic position of Tukang Besi in South-east Sulawesi has always been safely within the Muna-Buton group.[1] This classification is based on Esser (1938), who included Tukang Besi within the Muna-Buton group, an entity first postulated by Adriani (1914). In the *Atlas van Tropisch Nederland,* Esser attempted a comprehensive classification of the Austronesian languages of Indonesia, but unfortunately he did not provide any reasons for his decisions or include an evidential basis for his subgrouping. He recognised nine primary subgroups in Sulawesi (then Celebes), one of which (No. XII) is the Muna-Buton group, consisting of the following four languages (with modern names in brackets):

1.	Muna-Buton	(Muna)
2.	South-Buton	(Cia-Cia)
3.	Language of the Tukang Besi islands, Kalaotoa, Karompa and Bonerate	(Tukang Besi)
4.	Wolio and Layolo	(Wolio, Layolo)

Anceaux (1978) added to our knowledge of the area by showing that the language situation on Buton is actually more complex. Leaving the Bungku-Tolaki languages aside for the purposes of this paper, he found three new languages on Buton island, to which he assigned letter codes rather than names: E (now called Lasalimu), G (Pancana) and H (Kamaru). Not much was said about the internal classification of these languages. Based on lexicostatistics, Anceaux (1978:281) notes that B (Tukang Besi) 'scores relatively low with

[1] This paper is a slightly revised version of the one presented at the Ninth International Conference on Austronesian Linguistics, Canberra, January 2002. I wish to thank Robert Blust, Mark Donohue, David Mead, Daniël Vermonden and Erik Zobel for comments and corrections. The following abbreviations are used in the glosses: A.PART = active participle, ART = article, NOM = nominative, OBJ = object, OBL = oblique, REQ = requestive. A list of language abbreviations and sources of data is given at the end of the paper.

John Lynch, ed. *Issues in Austronesian historical phonology,* 87–113.
Canberra: Pacific Linguistics, 2003.

all the others. Still there is reason to believe that all the languages of this area belong to one subgroup, but it is only through further comparisons that the truth of this hypothesis can be tested'.

Subsequent work by Bhurhanuddin (1979), van den Berg (1991a), Donohue (1993, 2000) has shown that the situation is even more complex. The most recent listing of the Muna-Buton languages is found in the 14th edition of the Ethnologue (B.F. Grimes 2000), where the following 17 languages are given, split into four subgroups.

Table 1: Muna-Buton languages with numbers of speakers[2]

a.	Buton	1.	Cia-Cia	61,000
		2.	Kamaru	3,000
		3.	Wolio	50,000
		4.	Wotu	5,000
	Lasalimu-Kumbewaha	5.	Kumbewaha	2,600
		6.	Lasalimu	1,700
b.	Kalao	7.	Kalao	500
		8.	Laiyolo	800
c.	Muna	9.	Busoa	2,300
		10.	Kaimbulawa	1,600
		11.	Kioko	1,000
		12.	Liabuku	>150
		13.	Muna	280,000
		14.	Pancana	6,000
d.	Tukang Besi-Bonerate	15.	Tukang Besi North	120,000
		16.	Tukang Besi South	130,000
		17.	Bonerate	9,500

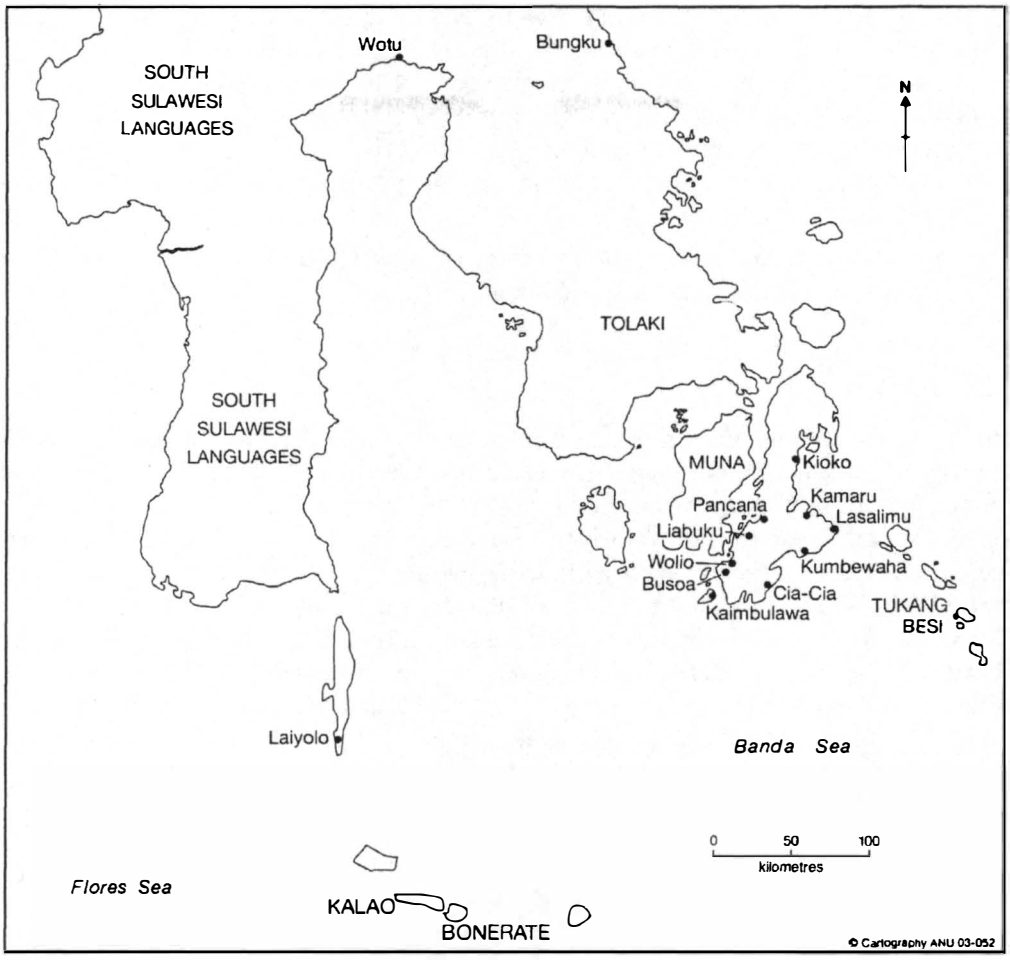
Apart from the classification (to which I return below), a few other points should be raised regarding this list.

1. The status of some of these languages is still tentative, especially Kumbewaha, Kioko, Liabuku and Pancana. Detailed survey work still needs to be done in this part of Buton in order to establish the exact number of languages and dialects (to the extent that this is possible).
2. Donohue (1993) splits up Cia-Cia into three languages: Cia-Cia, Masiri and Island Cia-Cia, but no evidence is given.
3. Bonerate is best considered a dialect of Tukang Besi South (85% cognate with the Binongko dialect, Donohue 2000:57). Also, Tukang Besi North and South are 'often identified as one by the native speakers'. I will follow this usage and refer to the language(s) as Tukang Besi, indicating individual dialects when necessary.

The accompanying map (adapted from Donohue 1993) gives the location of most of these languages.

---

2 These figures are based on *Ethnologue* (14th edition) supplemented by local updates obtained by David Andersen and myself.



Map 1: The Muna-Buton languages

In an important but as yet unpublished paper, Donohue (1993) has questioned the validity of the Muna-Buton group as a unity.[3] He was the first to take up the challenge of putting this putative subgroup on firm comparative ground. Based on a number of sound correspondences he reached the following conclusions:

1. Wolio is to be removed from the Muna-Buton group. Instead, together with Kamaru (eastern Buton) it links up with Laiyolo (southern part of Selayar island) and Wotu (at the tip of the gulf of Bone in South Sulawesi), as suggested earlier by Grimes and Grimes (1987). Presumably Kalao should also be included in this Wolio-Wotu group.

---

3 In spite of its unpublished nature, I refer to this paper as Donohue (1993) as it has been in circulation for several years. A more recent version of this paper (to be published by Pacific Linguistics) was kindly made available to me by the author, but reached me after I had written most of this article. The recent version differs from the earlier one (among other things) in that all references to the position of Tukang Besi have been removed.

2. Tukang Besi is also to be removed from the Muna-Buton group, as it differs in crucial respects from the 'real' Muna-Buton languages. Its subsequent affiliation is unclear, although Donohue mentions that 'certain aspects of linguistic classification point to an origin in the Tomini gulf area in their prehistory' (Donohue 1993:18). This statement must be considered speculative. In his grammar of Tukang Besi (Donohue 1999), the author makes no further reference to the question of the position of Tukang Besi within the wider Sulawesi context.
3. The remaining languages of Muna and Buton do constitute a valid group containing two subgroups, which he calls Munan and Buton.

2 Current hypothesis

Based on further comparative work, I would like to validate Donohue's claim that Wolio is not part of the Muna-Buton group. This would also seem to apply to Kamaru, although the scarcity of data for this language makes it difficult to be dogmatic. Whether or not Wolio, Kamaru, Wotu, Kalao and Laiyolo form a distinct subgroup is a completely different question which has not been properly addressed. Further research may well indicate that the putative Wolio-Wotu group is part of the Kaili-Pamona microgroup.

However, my conclusion for Tukang Besi is different from Donohue's. I propose that this language is indeed part of the Muna-Buton group, but constitutes a primary branch of it. The remainder I call Nuclear Muna-Buton, which splits into Munic and Butonic. In other words, I postulate the following structure (leaving details within Munic and Butonic to be worked out).

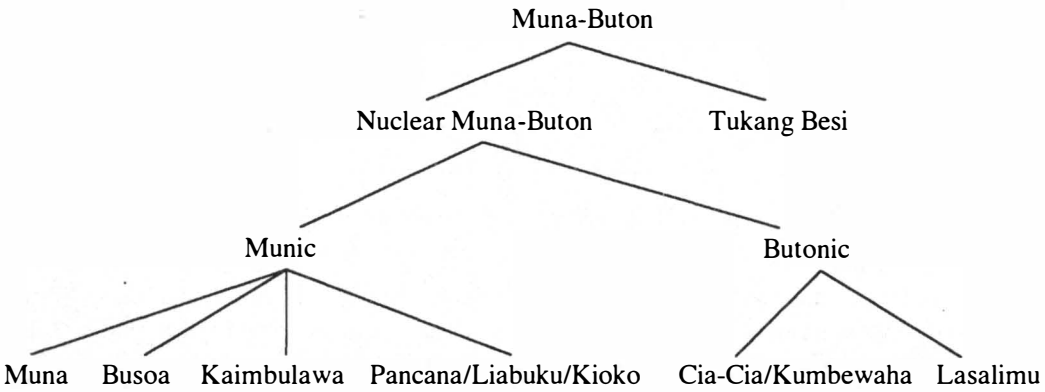


Figure 1: The Muna-Buton group

In this paper I will present the evidence for including Tukang Besi within the Muna-Buton group. This evidence will consist of phonological, grammatical and lexical innovations.

3 Phonological evidence

3.1 Phonological innovations

I start off by repeating Table 5 from Donohue (1993), as this presents in his opinion the crucial evidence for separating off Tukang Besi from the Muna-Buton group. I hope to show that this table is not without its problems, and that another solution is possible.

Table 2: Muna-Buton vs Tukang Besi sound changes

PAn	*b	*e	*Z	*R	*uy	*S	*j	*w
Kaimbulawa	*w*	*o*	*s*	*y	*i*	*Ø*	*y[4]	*o, u*
Muna	*w*	*o*	*s*	*y	*i*	*Ø*	*y	*o, u*
Kumbewaha	*w*	*o*	*s*	*y	*i*	*Ø*	*y	*o, u*
Cia-Cia	*w*	*o*	*s*	*y	*i*	*Ø*	*y	*o, u*
Tukang Besi languages	*w*	*o*	*s*	*h*	*u*	*ʔ*	*Ø*	*Ø*
Wotu-Wolio languages	*b*	*a*	*d*	*Ø*	*o, u*	*Ø*	*y	*o, u, w*

It is obvious that the Wotu-Wolio reflexes diverge considerably from the other languages and hence their exclusion from Muna-Buton is justified. For Tukang Besi the situation is less clear. Of the eight sound changes listed by Donohue, five are not shared between Tukang Besi and the Muna-Buton languages, two of the others (**b* > *w*, **e* > *o*) Donohue considers to be relatively natural changes occuring elsewhere also, which leaves only one (**Z* > *s*). Donohue concludes that a basis of one shared sound change is too weak for subgrouping purposes and, consequently, that Tukang Besi is not part of the Muna-Buton group.

A number of cricital comments can be made here. First, the change **b* > *w* listed above is too rigorous. Many etyma in the Muna-Buton languages have retained PMP **b* as implosive *ɓ* and there is much variation between *ɓ* and *w* even within the group. For example, PAn **baŋun* 'get up' is reflected as Mun *wanu* but as *ɓanu* in Kaimbulawa and as *ɓaŋu* in Busoa, Cia-Cia, Lasalimu and Tukang Besi. The change **b* > *w* is one of the lenition changes that pervades the whole group (see below) and cannot be used for subgrouping purposes.

Second, PAn **S* > *ʔ* in Tukang Besi is problematic. I believe it is much easier to assume that PAn **S* (PMP **h*) became zero in Tukang Besi. Donohue (1993:23) gives the following illustrative etyma:

PAN		Tuk	Cia	Mun
**kaSiw*	'wood'	*kau ~ kaʔu*	*sau*	*sau*
**daSun*	'leaf'	*roʔo*	*roʔo*	*roo*
**i-kaSu*	'you'	*ikoʔo*	*isoʔo*	(*ihintu*)
**Su*	'2SG subject prefix'	*ʔu*	*o-, u-*	*o-*

According to Donohue, the glottals in the Cia-Cia forms are a 'regular non-phonemic insertion between like vowels', whereas Tukang Besi preserves **S* as *ʔ*. It appears, however, that in Tukang Besi the same rule operates in many (but not all) forms. A glottal often appears to break up like vowels where it does not reflect a consonant. Some examples:

PMP		Tuk	Cia	Mun
**pahuq*	'mango'	*poʔo*	*poʔo*	*foo*
**sawa*	'snake'	*saʔa*	*saʔa*	*sa(a)-*
**enem*	'six'	*noʔo*	*noʔo*	*noo*
**epat*	'four'	*paʔa*	*paʔa*	*paa*

This rule is not as strict in Tukang Besi as it is in Cia-Cia. Sequences of like vowels do indeed occur in Tukang Besi, but they appear to be the result of the loss of **j* or (in one case) a sporadic change. Examples are *ʔoloo* 'day' (PMP **qalejaw*); *ŋaa* 'name' (PMP **ŋajan*)

---

4 Proto Muna-Buton **y* is discussed below.

and *mohii* 'left-handed' (PMP **wiRi*, through metathesis from ***ihi*). Sequences of like vowels are also found in loan words e.g. *karajaa* 'work' (Malay *kerja*); *paraluu* 'need' (Malay *perlu*); *koosu ae* 'socks (of feet)' (Dutch *kous*) and words of unclear origin, e.g. *mohoo* 'sick' and *ree* 'cough'. The glottals in the words *ro?o* 'leaf' and *iko?o* 'you' can therefore be considered regular phonemic developments. The second person singular subject prefix has an unexpected initial glottal, but this is almost certainly a later development. The Southern Muna dialect of Gu uses initial glottals on all function morphemes (including pronouns and subject markers). Contrary to regular glottals, these have no corresponding *gh* or *h* in the northern (standard) dialect of Muna. This leaves us with only one etymon from the list (*kau* ~ *ka?u* 'wood') in which the glottal is unexplained. And even there we find a parallel in Lasalimu which has *ka?u* for 'smoke', from Proto Muna-Buton **qahu* (expected ***kau*: **h* regularly goes to zero in Lasalimu). A last piece of evidence is the reflex of **S* as zero in Tukang Besi *dua* 'two' (PAn **duSa*). The argument that Tukang Besi differs from the other Muna-Buton languages in its reflex of PAn **S* can therefore not be maintained.

Third, the claimed change of PAn **w* to Muna-Buton *o/u* is also problematic. Donohue gives the following etyma to prove his point.

PAn		Tuk	Cia	Mun
**walu*	'eight'	*alu*	*oalu*	*oalu*
**siwa*	'nine'	*sia*	*siua*	*siua*
**wanan*	'right'	*mo/ana*	*s/uana*	*s/oana*

I posit instead that all the Muna-Buton languages share the change PMP **w* > *Ø*. Initial evidence that PMP **w* > *Ø* is provided by Muna *sa(a)-* 'snake' (in names of certain snakes and in the place name *Lianosaa* 'snake cave'; from PMP **sawah*); *aa* 'waist' (PMP **hawak*) and *kai* 'fish hook' (PMP **kawit*). The apparent reflexes of **w* as *o/u* can all be explained. As Mead (1998:46) has pointed out, Muna and Cia-Cia *oalu* go back to a reduplicated form **wa-walu*. The developments are then completely regular, with loss of **w* and antepenultimate **a* > *o*. The word for 'nine' is indeed an exception in Cia-Cia and Muna, but it is irregular anyway in that the glide has syllabified (although stress remains on the antepenultimate syllable *si* in Muna, and not — as expected — on the penultimate syllable *u*). Finally, Cia-Cia *soana* and Muna *suana* 'right' go back to PMP **ka-wanan* (rather than **wanan*), with regular loss of **w*, lenition of **k* > *s* and antepenultimate raising of **a* > *o* and even further to *u* in Muna. (For antepenultimate raising of **a* > *u* in Muna, cf. **qateluR* 'egg' > *ghunteli* and **baqeRu* 'new' > *bughou*.)

In conclusion, of the five changes listed as differentiating Muna-Buton and Tukang Besi, two are invalid.

3.2 Reconciling Muna-Buton and Tukang Besi

This leaves three changes in Donohue's chart where Tukang Besi and the remaining Muna-Buton languages actually differ: the reflexes of PMP **R*, **uy* and **j*. I add here the reflex of PMP **ej* (Tukang Besi *o*, Nuclear Muna-Buton *e*), The solution I propose is shown in Table 3.

Table 3: Selected Proto Muna-Buton reflexes

PMP	Proto Muna-Buton	Tukang Besi	Proto Nuclear Muna-Buton	Cia-Cia	Muna
**uy*	**uy*	*u*	**i*	*i*	*i*
**ej*	**oy*	*o*	**e*	*e*	*e*
**R*	**R*	*h*	Ø, vowel coalescence	vowel coalescence	vowel coalescence
**j*	**y*	*Ø*	Ø, vowel coalescence	vowel coalescence	vowel coalescence

The different reflexes of **uy* can be accounted for by assuming the diphthong was retained in Proto Muna-Buton. This is relevant for three etyma:

PMP	PMB		Tuk	Cia	Mun
**babuy*	**wawuy*	'pig'	*wawu*	*wawi*	*wewi*
**hapuy*	**apuy*	'fire'	*apu*	*api*	*ifi*
**(l, n)aŋuy*	**(l, n)aŋuy*	'swim'	*naŋu*	*pika/naŋu*	*leni*

(Cia-Cia *pika/naŋu* irregularly reflects **uy* as *u*; it is probably a loan from Tukang Besi.)

The situation for PMP **ej* is similar. For 'fly' and 'snake' there is disagreement, the Tukang Besi reflex *o* pointing to PMB **oy*. I therefore tentatively reconstruct this diphthong, so far found in only two reconstructed forms. Notice that the reflexes of **ej* differ from those of **ey* in Tukang Besi.

PMP	PMB		Tuk	Cia	Mun
**lalej*	**laloy*	'fly'	*lalo*	*lale*	*lale(mbanua)*
**qulej*	**quloy*	'snake'	*ulo*	*kule*	*ghule*
**qatey*	**qate*	'liver'	*ate*	*hate*	*ghate*
**quey*	**que*	'rattan'	*ʔue*	Las *kue*	*ghue*

The reflex of PMP **R* as Tukang Besi *h* is indeed surprising, but a reconciliation with Muna and Cia-Cia vowel coalescence or zero is possible. I propose Proto Muna-Buton **R*, which at that time still had its original quality, probably a voiced velar fricative /ɣ/ or a voiced uvular fricative /ʁ/. A velar or uvular fricative is still a very common sound in the Muna-Buton area, synchronically attested for Muna (written as *gh*), Busoa and the Pasarwajo dialect of Cia-Cia. While in Tukang Besi **R* > *h* and **j* > Ø, the western Muna-Buton languages merged these sounds as either zero in between high vowels or as **y* between non-high vowels, with subsequent vowel coalescence. Some examples are:

PMP	**DaRaq* 'blood'	**DuRi* 'thorn'	**ŋajan* 'name'	**qalejaw* 'sun'
PMB	**raRa*	**ruRi*	**ŋaya*	**qoloyo*
Tuk	*raha*	*ruhi*	*ŋaa*	*ʔoloo*
PNMB	**rea* (< **raya*)	**rui*	**ŋea*	**qoleo*
Cia	*rea*	*rui*	*ŋea*	*holeo*
Mun	*rea*	*ki/ri* (< *ka/rui*)	*nea*	*gholeo*

At this point it is not entirely clear whether PMP final **R* was completely lost in PMB or whether it was retained as **y*. The distribution of the various reflexes is uneven in the daughter languages, and more data is needed before a firm conclusion can be drawn.

The preceding discussion has shown that Tukang Besi does indeed differ from the other Muna-Buton languages in four of its PMP reflexes, but this can be accounted for by proposing a higher link between Tukang Besi and the Nuclear Muna-Buton languages.

3.3 From PMP to Proto Muna-Buton

In Table 4 below I have listed the reflexes of PMP (Proto Malayo-Polynesian) in seven languages of the Muna-Buton group. My starting point is PMP rather than PAn, as a significant number of PAn phonemes had already merged by the time of PMP. PAn protophonemes such as **C* , **N* and **S* are therefore irrelevant for lower-level reconstruction work. In Table 4, multiple reflexes are listed in decreasing order of frequency; brackets denote rare occurrences; *v.c.* = vowel coalescence; and ? indicates uncertainty.

Table 4: PMP reflexes in PMB and selected Muna-Buton languages

PMP	PMB	Tukang Besi	Muna	Busoa	Kaimbulawa	Cia-Cia	Lasalimu
*a	*a	a	a	a	a	a	a
*i	*i	i	i	i	i	i	i
*u	*u	u	u	u	u	u	u
*e	*o	o	o	o	o	o	o
*ay	*e	e	e	e	e	e	e
*ey	*e	e	e	e	e	e	e
*ej	*oy	o	e	e	e	e	e
*uy	*uy	u	i	i	i	i	i
*aw	*o	o	o, u	o	o	o	o
*p	*p	p, h	p, f	p, f	p	p	p
*t	*t	t	t	t	t	t, c /_ i, u	t, c /_ i, u
*k	*k, *s	k (s)	k, s	?, s	k, s	k, s	k, s
*q	*q	?, Ø	gh (=ʁ)	h	h	h, k	Ø, k
*b	*ɓ, *w	ɓ, w	ɓ, b, w	ɓ, w	ɓ, w	ɓ, w	ɓ, w
*D/d/r	*r, *ɗ	r, ɗ	r, ɗ	ɣ, ?	ʕ, ?	r, ɗ	r, ?
*Z/z	*s, (*ɗ)	s (ɗ)	s, (ɗ)	s, ?	s, ?	Ø (ɗ)	Ø, ?
*j	Ø, *y	Ø	v.c	v.c	v.c	v.c.	?
*m	*m	m	m	m	m	m	m
*n	*n	n	n	n	n	n	n
*ñ	*n	n	n	n	n	n	n
*ŋ	*ŋ	ŋ	n	ŋ, n	n	ŋ	ŋ
*R	*R, Ø, ?*y	h	Ø, v.c	Ø, v.c.	Ø, v.c.	Ø, v.c.	Ø, v.c.
*l	*l	l	l	l	l	l	l
*s	*s, *h	s, h	s, h	s, h	s, Ø	s, Ø	s, Ø
*h	Ø	Ø	Ø	Ø	Ø	Ø	Ø
*w	Ø	Ø	Ø	Ø	Ø	Ø	Ø
*y	Ø	Ø	Ø	Ø	Ø	Ø	Ø

3.4 PMB phonological innovations

I propose the following phonological innovations from PMP to PMB as defining the Muna-Buton subgroup.

1. *Z/z > *s, *ɗ
2. *s > *s, *h
3. *D/d/r > *ɗ, *r
4. *b > *ɓ, *w
5. *h, *w, *y > *Ø*
6. *e,*aw > *o*
7. *ay, *ey > *e*
8. Loss of all final consonants (except in the diphthongs *uy* and *oy).
9. Reduction of all medial clusters.

3.5 Proto Muna-Buton phonology

Based on the preceding discussion, the following picture emerges of Proto Muna-Buton phonology:

Consonants					Vowels		Diphthongs
*p	*t	*k	*q		*i	*u	*uy
*ɓ	*ɗ	*g			*e	*o	*oy
*mp	*nt	*ŋk			*a		
*mb	*nd	*ŋg					
	*s		*R	*h			
	*ns						
*m	*n	*ŋ					
	*l						
	*r						
*w		*y					

The following comments on the phoneme chart need to be made:

1. I assume *q* was a voiceless uvular plosive /q/.
2. There were three voiced plosives, two of which were implosives (*ɓ* and *ɗ*), as well as a regular *g*. In present-day languages the implosive quality of the voiced stops is greatest for the bilabials, somewhat less for the alveolars and rare for velars, with only Tukang Besi having an implosive [ɠ] as an allophone of /g/. Implosives have spread to non-Muna-Buton languages in the area such as Wolio (Anceaux 1952) and Kulisusu (Mead 1998:21); even Tolaki on the Sulawesi mainland has implosives in free variation with plain plosives (Mead & Tambunan 1993).
3. In Muna, Cia-Cia and Lasalimu /ɓ/ does not occur before /u/, but in Tukang Besi it does (according to Donohue 1999). I tentatively reconstruct *ɓ also before /u/, e.g. Proto Muna-Buton *ɓura* 'face powder', based on Muna, Cia-Cia *bura* and Tukang Besi *ɓura*.
4. *R was probably phonetically /ɣ/ or /ʁ/.
5. The seven prenasalised consonants functioned as units: *mp, *mb, *nt, *nd, *ns, *ŋk* and *ŋg*. Some of these also occur word-initially in PMB etyma, e.g. *mbaka* 'delicious', *ndoke* 'monkey', *nturu* 'often'.
6. *w was probably /β/ and could alternatively be listed in the fricative column.
7. Syllable structure was open: (C) V.

8. Sequences of two or three vowels were allowed: **ɗeu* 'needle', **ɗaoa* 'market'.
9. The diphthongs occurred only word-finally.
10. Stress was penultimate.

3.6 Sporadic sound changes within the Muna-Buton group

Several lenitions and other types of changes have pervaded the Muna-Buton subgroup, but they are clearly not diagnostic for the group as a whole. These changes have taken place to a greater or lesser extent in individual languages, but only affecting part of the vocabulary. In other words, they are typically wave-like in character. The most common lenitions are as follows:

PMB		Reflexes in individual languages
**p*	>	*f, h*
**k*	>	*s*
**q*	>	*ʁ, k, ?, h, Ø*
**ɓ*	>	*w*
**ɗ*	>	*r*
**s*	>	*h, Ø*

I discuss these in more detail below.

1. PMB **p* > *f* > *h*. The change **p* to *f* has primarily affected Muna, and to a lesser degree some Pancana dialects. Tukang Besi has gone even further in this lenition process, and often reflects **p* as *h* (actually [ɸ] before *u*). Even /p/ can be realised as [ɸ] or [pɸ] before /a/ and /o/ (Donohue 1999:16). This lenition process is interesting in that the two languages on the geographical edges of the group are affected, but those in the middle are not. However, Muna and Tukang Besi do not always agree with each other in the etyma in which this lenition occurs, giving further evidence to the irregular spread of this sound change. Some examples of PMP and PMB **p* in Tukang Besi and Muna:

PMP	PMB		Tuk	Mun
**epat*	**pato*	'four'	*hato-*	*fato-*
**pitu*	**pitu*	'seven'	*hitu-*	*fitu-*
**paniki*	**poniki*	'bat'	*honiki*	*ponisi*
**Depah*	**ropa*	'fathom'	*sa/roha*	*pute*
**paqah*	**paqa*	'thigh'	*pa?a*	*fagha*
**pakay*	**pake*	'wear, use'	*pake*	*pake*
**pandak*	**panda*	'short'	*me/panda*	*panda*

2. The lenition of PMP **k* > *s* is widespread in Nuclear Muna-Buton (e.g. PMP **kahiw* 'wood' > PNMB **sau*; PMP **kali* 'dig' > Muna *seli*), but in Tukang Besi it has only occurred in the possessive suffix *-su* 'my' (Muna *-ku*) and the dative object suffixes *-nso* and *-nsami* (see below).
3. None of the Muna-Buton languages has retained **q* as a uvular plosive, but only in Tukang Besi is there a split between glottal and zero, with glottal occuring medially (with one exception), and zero or glottal initially, without obvious conditioning factors. In one case a difference in meaning seems to have developed, if the transcription and the glosses are correct. Notice the following reflexes of PMP **q* in Tukang Besi, Cia-Cia and Muna:

PMP	PMB		Tuk	Cia	Mun
**qatep*	**qato*	'roof'	*ato*	*hato*	*ghato*
**qazay*	**qase*	'chin'	*ase*	*hae*	*ghase*
**qabu*	**qabu*	'ashes'	*awu*	*hawu*	*ghabu*
**qenay*	**qone*	'sand'	*ʔone*	*hone*	*ghone*
**qalipan*	**qolipa*	'centipede'	*oliha* 'centipede' *ʔoliha* 'scorpion'	*honipa*	*gholifa*

4. PMB **ɓ* > *w*. This is an ongoing lenition in the whole Celebic group (see van den Berg 1991b:12), especially common before *u*. In some cases PMB had already undergone weakening (e.g. PMP **tebuh* 'sugarcane' > PMB **towu*). Notice the erratic pattern in the following words:

PMP	PMB		Tuk	Cia	Mun
**baŋun*	**ɓaŋu*	'get up'	*ɓaŋu*	*ɓaŋu*	*wanu*
**bisul*	**ɓisu*	'boil'	*ɓisu*	*ka/wincu*[5]	*ka/wisu*
**baqeRu*	**ɓoqou*	'new'	*woʔou*	*wukou*	*bughou*

5. PMB **s* > *h* > zero is again a lenition pattern which had already begun in some words before PMB: note PMP **tasak* 'ripe' > PMB **taha*; PMP **tasik* 'sea' > PMB **tahi* (Mun *tehi*, Cia *tai*).

Other sporadic changes occurring in the Muna-Buton group are as follows:

6. Vowel height assimilations, e.g. **iCa* > *eCa* or *iCe* and even *iCi*, as in PMB **mia* 'person', Cia Tuk *mia*, Mun *mie*, Bus *mii*. Also **aCi* > *eCi* and **uCa* > *uCe*, as in PMB **wuta* 'earth, ground', Tuk *wuta*, Bus *wute*, Mun *wite*.
7. Fronting of back vowels (**u* > *i*; **o* > *e*), e.g. PMB **tau* 'put, place', Mun *tei* (through ***tai*).
8. Antepenultimate raising (**a* > *o* or *u* in pretonic syllables), e.g. PMP **ka-wanan* 'right side', Cia *soana*, Mun *suana*.
9. Final **u* > *o*, e.g. PMB **taqu* 'year', Tuk *taʔo*.
10. Prenasalisation, e.g. PMP **tadɛ* 'stand', Tuk *tadɛ*, Cia *tadɛ* ~ *ntadɛ*, Mun *ntadɛ*.
11. Metathesis (often involving words of VCV shape with identical vowels), e.g. PMB **isi* 'flesh', Cia *isi*, Mun *ihi*, Bus *hii*.

3.7 Irregular phonological developments from PMP to PMB

The following list shows a number of irregular phonological developments in individual etyma which are shared by all the languages in the Muna-Buton group, including Tukang Besi. Again, this is strong subgrouping evidence. For reasons of space the supporting material in the individual languages cannot be given.

---

5 The Masiri/Mambulu dialect of Cia-Cia has *kaɓisu*.

PMP		PMB	
**bahaq*	'flood'	**mawa*	(possibly fossiziled *ma-*)
**baReqaŋ*	'molar tooth'	**ɓaga* 'cheek'	(**Req* > **g*)
**binehiq*	'seed'	**wine*	(contraction and lowering)
**(ma-)beReqat*	'heavy'	**ɓoRa*	(contraction of final syllables)
**baqeRu*	'new'	**ɓoqou*	(loss of **R*)
**dilaq*	'tongue'	**lela*	(assimilation and vowel lowering)
**Dalem*	'inside, deep'	**laro* 'inside'	(metathesis)
		**ndalo* 'deep'	(prenasalisation)
**enem*	'six'	**noo* (free)	(metathesis)
		**nomo* (bound)	(paragogic vowel)
**epat*	'four'	**paa* (free)	(**e* > **a*, metathesis)
		**pato* (bound)	(paragogic vowel)
**isa*	'one'	**sa* (bound)	(loss of **i*)
**kabut*	'fog, mist'	**gawu*	(**k* > **g*)
**kempuŋ*	'intestines'	**kompo*	(lowering of **u* to **o*)
**kita*	'see'	**ita*	(loss of **k*)
**maRi*	'come'	**mai*	(loss of **R*)
**maRuqanay*	'man'	**moqane*	(reduction of first syllable)
**mula*	'plant'	**pembula*	(prefix and prenasalisation)
**penuq*	'full'	**pono*	(lowering of **u* to **o*)
**putiq*	'white'	**pute*	(lowering of **i* to **e*)
**qelet*	'crevice'	**qolota* 'space between'	(irregular paragogic vowel)
**Ribu*	'thousand'	**riwu*	(**R* > **r* irregular; or loan?)
**sa-puluq*	'ten'	**ompulu*	(prenasalisation, loss of **s*)
**sebu*	'temper'	**soropu*	(**b* > **p*; infix)
**ZaRum*	'needle'	**ɗeu*	(via ***dayu*, with **R* > **y*)

Some of these innovations may in fact have occurred in a higher-order subgroup comprising both Proto Muna-Buton and Proto Bungku-Tolaki. Compare PBT **laro* 'inside'; **nomo* 'six'; **pato* 'four'; **gawuQ* 'mist'; **kompoN* 'stomach, intestines'; **mai* 'come' and **pono* 'full'. If this is indeed the case, it would strengthen the hypothesis of a South-eastern Celebic subgroup as put forward by Mead (this volume).

4 Grammatical evidence

In this section I list some apparently exclusively shared grammatical innovations in the Muna-Buton group. This section is far from exhaustive and is probably only indicative of what can be further unearthed.

4.1 The pronominal system

The substantial similarities in form and function of the pronoun systems in three daughter languages indicate that a similar system existed in PMB. Let us first consider present-day languages for which data are available. (In these charts the symbol ' marks a glottal stop).

Table 5: Tukang Besi pronominal forms (terminology from Donohue 1999)

	Free forms	Irrealis Subject	Realis Subject	Possessive	Object	Dative object
1SG	*iaku*	*ku-*	*ku-*	*-su*	*-aku*	*-naku*
2SG	*iko'o*	*ko-*	*'u-/nu-*	*-'u*	*-ko*	*-nso*
3SG	*ia*	*na-/a-*	*no-/o-*	*-no*	*-'e*	*-ne*
1PAUCAL	*ikami*	*ka-*	*ko-*	*-mami*	*-kami*	*-nsami*
1PLURAL	*ikita*	*ta-*	*to-*	*-nto*	*-kita*	*-nggita*
2PL	*ikomiu*	*ki-*	*i-*	*-miu*	*-komiu*	*-ngkomiu*
3PL	*amai*	*na-/a-*	*no-/o-*	*-no*	*-'e*	*(amai)*

Notes to Table 5:

1. 3SG and 3PL are identical in their affixed forms; the free pronoun is different in each case.
2. The pronouns *ikami* and *ikita* are glossed as 'paucal' and 'plural' respectively, but 'they are used with a lot of overlap by most speakers' (Donohue 1999:114). Some younger speakers impose an inclusive–exclusive meaning difference onto these forms (probably due to Indonesian influence), but in the speech of older speakers this is not the case.
3. Respect for the addressee is shown by using the second person plural forms; even greater respect by the use of the first person plural forms (*ikita*).
4. 'The dative object suffixes are observed only rarely in the Wanci dialect, and their use is somewhat archaic' (Donohue 1999:135).

Table 6: Muna pronominal forms (van den Berg 1989)

	Free forms	Subject realis	Subject irrealis	Possessive	Direct object	Indirect object
1SG	*inodi, idi*	*a-*	*a-*	*-ku*	*-kanau*	*-kanau*
2SG	*(i)hintu*	*o-*	*o-*	*-mu*	*-ko*	*-angko*
2SG POLITE	*intaidi*	*to-*	*ta-*	*-nto*	*-kaeta*	*-kaeta*
3SG	*anoa*	*no-*	*na-*	*-no*	*-e*	*-ane*
1DU INCL	*intaidi*	*do-*	*da-*	*-nto*	–	–
1PL INCL	*intaidiimu*	*do-Vmu*	*da-Vmu*	*-ntoomu*	–	–
1PL EXCL	*insaidi*	*ta-*	*ta-*	*-mani*	*-kasami*	*-kasami*
2PL	*(i)hintuumu*	*o-Vmu*	*o-Vmu*	*-Vmu*	*-angkoomu*	*-angkoomu*
2PL POLITE	*intaidiimu*	*to-Vmu*	*to-Vmu*	*-ntoomu*	*-kaetaamu*	*-kaetaamu*
3PL	*andoa*	*do-*	*da-*	*-ndo*	*-anda*	*-anda*

Notes to Table 6:

1. The subject realis forms are given for only one of the three verb classes (named class *a-*, after its 1SG subject marker). The other two classes (class *ae-* and class *ao-*) are, from a diachronic perspective, combinations of the subject markers and the verbal markers *me-* and *mo-* (e.g. *ae-ala* 'I take' comes from earlier **a-me-ala*; see van den Berg 1991b:24–25 for details).
2. In the suffix *-Vmu*, *V* marks a vowel which is identical to the last vowel of the stem.

Table 7: Cia-Cia pronominal forms (van den Berg 1991c)

	Free forms	Subject realis	Subject irrealis	Possessive	Direct object	Indirect object
1SG	*('i)nda'u*	*o-*	*a-*	*-'u*	*-aa'u/-'u*	*-'isa'u*
2SG	*('i)so'o*	*mu-/mo-*	*cu-*	*-mu/-mo*	*-so*	*-'iso*
3SG	*'ia*	*no-*	*na-*	*-no*	*-'e/-e*	*-'isie*
1PL INCL	*'ingkita*	*to-*	*ta-*	*-nto*	*-kita*	*-'ikita*
1PL EXCL	*'isami*	*to-*	*ta-*	*-mami*	*-sami*	*-'isami*
2PL	*'isimiu*	*ka-*	*cu-ka-*	*-miu*	*-simiu*	*-'isimiu*
3PL	*mo'ia*	*no(-ka)-*	*na(-ka)-*	*-no (mo'ia)*	*-'e/-e*	*-'isie*

Notes to Table 7:

1. The 1SG direct object form *-'u* occurs after verb stems ending in *a*. The form *-aa'u* occurs elsewhere.
2. The 3SG direct object form *-e* occurs after verb stems endings in a high vowel (*i* or *u*). The form *-'e* occurs elsewhere.
3. There is considerable dialect variation in Cia-Cia and this chart represents the northern Sampolawa and Pasarwajo dialect. The Masiri/Mambulu dialect appears to be more conservative, having e.g. *ia'u* 'I' and a single form *a-* for both realis and irrealis (Daniël Vermonden, pers. comm.).
4. The indirect object forms display unusual dialectal variation (e.g. 1SG also *-sia'u*) which needs more research.

Table 8: Lasalimu pronominal forms (van den Berg, field notes)

	Free forms	Subject realis	Subject irrealis	Possessive	Direct object	Indirect object
1SG	*ia'u*	*a-*	*a-*	*-'u*	*-a'u*	*-'asuna'u*
2SG	*iso'o/ka'ancu*	*u-*	*si-*	*-mu*	*-so*	*-'asoso*
3SG	*i'ia*	*no-*	*na-*	*-no*	*-'e*	*-'ase'e*
1PL INCL	*ikita*	*to-*	*ta-*	*-nto*	*-kita*	*-'asokita*
1PL EXCL	*isami*	*to-*	*ta-*	*-mami*	*-sami*	*-'asosami*
2PL	*isimiu*	*i-*	*i-*	*-miu*	*-simiu*	*-'asosimiu*
3PL	*mo'ia*	*no-*	*na-*	*-no (mo'ia)*	*-'e (mo'ia)*	*-'aso mo'ia*

Notes to Table 8:

1. The 2PL form *i-* is also used as a singular honorific form.
2. The free pronoun *ka'ancu* is related to the demonstrative *ancu* 'there (near you)'.

Because of the very obvious similarities in form and function, I argue that the following system can be posited for Proto Muna-Buton, and that it constitutes a clear and unambiguous set of innovations defining the Muna-Buton group:

1. Realis and irrealis subject pronominal affixes.
2. A nominative–accusative system of pronominal marking.
3. The infix *-um-*, often in combination with irrealis affixes, to mark intentionality.
4. A set of indirect/dative object suffixes to mark beneficiaries, recipients, instruments and other non-core roles.

I will briefly discuss the first three points in more detail.

The use of realis and irrealis subject markers in the Muna-Buton languages is unique in Sulawesi (not even shared by Wolio) and is one of the strongest subgrouping arguments. Realis forms are used for the past and the present, whereas irrealis forms indicate future events. In Muna and Cia-Cia the irrealis forms are also used after negators, but this is not true for Tukang Besi, and hence presumably an innovation in Nuclear Muna-Buton. Notice that in Muna realis and irrealis subject prefixes are not distinguished for 1SG, 2SG/PL and 1PL EXCL, while in Tukang Besi the contrast is absent for 1SG only. Cia-Cia maintains the distinction throughout the entire paradigm, although not in all dialects. At this stage the extent of the realis-irrealis distinction in the Proto Muna-Buton paradigm is unclear, but it should be at least reconstructed for 3SG and 1PL INCL.

The pronominal system worked on a nominative–accusative basis, as illustrated in the following examples:

Tuk	*ku-rato*	*no-rato*	*no-sepa-'aku*	*ku-sepa-'e*
Cia	*o-rato*	*no-rato*	*no-sepa-'u*	*o-sepa-e*
Mun	*a-rato*	*no-rato*	*no-sepa-kanau*	*a-sepa-e*
	'I came'	's/he came'	's/he kicked me'	'I kicked him/her'

Notice that Wolio (Anceaux 1952) and Kulisusu (Mead 1998) have a nominative–accusative system as well, but most likely this is the result of areal influence from the Muna-Buton languages.

The irrealis prefixes are often used in combination with the infix *-um-*. In Muna this usage is obligatory for class *a-*, where irrealis prefixes and *-um-* must co-occur (e.g. realis *no-kala* 'he goes; he went'; irrealis *na-k[um]ala* 'he will go'; while **na-kala* and **no-k[um]ala* are ungrammatical). In Cia-Cia and Tukang Besi, however, this appears to be a tendency only, and irrealis forms without *-um-* do occur. In such cases, i.e. where the irrealis subject prefix is found without *-um-*, the focus appears to be on the intention of the action ('want, will'), rather than on the future itself.

All this means that a system of six pronominal forms can be reconstructed for PMB: one set of free pronouns and five sets of pronominal affixes (possessive, realis and irrealis subject, direct object and indirect/dative object). The exact reconstruction of the forms remains to be worked out, but there is clearly enough evidence to posit such a system. The allomorphy rules of *-um-* also remain to be worked out, but they are likely to be similar or identical to the ones we find in Tukang Besi, i.e. vowel-initial roots have *um-* (*m-* in Muna); *p-* and *b-*initial roots take *-um-* (but *p-* changes to *m-* in Muna and *b-* takes zero); roots with initial *m-* take the zero allomorph; while roots with initial *w-* show variation. Complications occur in the case of derived bases.

4.2 The infix *-um-* forming subject relative clauses

Apart from its function to indicate intentionality, the infix *-um-* is also used to create a specific verbal form which I have called an 'active participle'. Such forms are not inflected for person and are only employed in subject relative clauses. In Tukang Besi *-um-* is the only affix in active participles, but in Muna the infix occurs in combination with the suffix *-no*; this appears to be the case too in Cia-Cia and Lasalimu (although the evidence is limited).

(1) Tuk *na mia t[um]opa te La Udi*
NOM person slap-[A.PART] CORE La Udi
'the person who slapped La Udi'

(2) Mun *o mie t[um]ofa-no La Ali*
ART person slap-[A.PART] La Ali
'the person who slapped La Ali'

(3) Cia, Las *o mia t[um]opa-no La Ali*
ART person slap-[A.PART] La Ali
'the person who slapped La Ali'

The combination of *-um-* and *-no* (or *-na*) to create active participles is also found in Wolio (Anceaux 1952:25) and Kulisusu (Mead 1998:360–362). Again, this must be the result of areal diffusion.

4.3 Derivational morphology

A few other derivational morphemes appear to be innovations in PMB.

a. Prefix **pe-* 'make, build'

Mun	*fe-*	'make, build'	e.g. *ne-fe-ghato*	'build a roof'
Cia	*pi-*	'make'	e.g. *pi-ka'ana*	'build a house'
Tuk	*he-*	'produce, make'	e.g. *no-fe-ato*	'weave a thatch roof'

b. Prefix **pa-* 'occupation' (in Muna and Cia-Cia an unproductive nominal prefix, in Tukang Besi a verb-deriving prefix)

Mun	*galu*	'field'	*pa-galu*	'farmer'
	hulo	'hunt'	*pa-hulo*	'hunter'
Cia	*hamota*	'field'	*pa-pi-hamota*	'farmer'
	ase	'iron'	*pa-pi-rabu ase*	'blacksmith'
Tuk	*tutu*	'pound'	*no-pa-tutu*	'he is a blacksmith'
	langke	'sail'	*no-pa-langke*	'he is a sailor'

c. Requestive **pe-* (Muna *fe-*, Tukang Besi *hepe-*). In addition to two regular causative prefixes, both languages have a requestive prefix meaning 'ask X to do something'. Although there is only partial similarity in form, it is striking that in both languages the action takes place for the benefit of the causer (the subject) and that the causee is expressed in an oblique phrase (rather than as the object).

(4) Mun *ne-fe-gholi bhadhu ne ina-no*
3SG.R-REQ-buy shirt LOC mother-his
'he asked his mother to buy him a shirt'

(5) Tuk *no-hepe-'ita-'e na aroloji di ama-no*
3R-REQ-see-3OBJ NOM watch OBL father-his
'he is asking his father to show him the watch'

d. Iterative **para-*. Both Tukang Besi, Cia-Cia and Muna have a prefix *para-* which is used for iterative and/or habitual action.

Tuk	*aso*	'sell'	*para-aso*	'sell as a regular activity'
Mun	*aso*	'sell'	*para-aso*	'be a regular seller'

4.4 Demonstratives

The demonstrative systems of both Muna and Tukang Besi are fairly complex. For Muna see van den Berg (1989: Chapter 6) and (1997); for Tukang Besi, Donohue (1999: Chapter 6). There are many differences on a detailed level, but it is striking that both languages have a dual opposition of deictic elements with very similar meaning, crucially involving an opposition of the phonemes /t/ and /w/. Consider the following examples of the Tukang Besi 'topographic demonstratives':

Tuk	*ito*	'up, landwards, east, north, in'
	iwo	'down, seawards, west, south, out'

Compare these forms with the Muna prepositions *te* and *we* and with the anaphoric (or referential) demonstratives:

Mun	*te*	'locative preposition used for a position or direction which is higher, up, east (and sometimes north) of the point of orientation'
	we	'locative preposition used for a position or direction which is lower, down, west (and sometimes south) of the point of orientation'
	tatu	'that, there, yonder (higher, up, east, north)'
	watu	'that, there, yonder (lower, down, west, south)'

A clear etymological connection between the Muna forms *te* and *tatu* (and *we* and *watu*) has not been found yet, but it appears that in conjunction with the Tukang Besi data a demonstrative pair is reconstructible for Proto Muna-Buton, one containing the phoneme **t* for places 'up, higher, east and north', and one containing **w* for places 'down, lower, west and south'. It should be noticed that Wolio also has a dual pair *(nca)siate* 'up there', *(nca)siroo* 'down there' (Alberth 2000), but it lacks both the typical *t-w* opposition and the meaning component of the cardinal points.

4.5 Other shared morphology

Other shared affixes can be reconstructed for Proto Muna-Buton, but many of these are probably retentions from PMP. Some of these are illustrated in Table 9.

Table 9: Proto Muna-Buton morphology

	Proto Muna-Buton	Tukang Besi	Cia-Cia	Muna
passive in relative clauses	**ni-*	*i-*	*ni-*	*ni-/ne-*
applicative/indirect object	**-ako*	*-ako*	*-aso*	*-ghoo*
perfective	**-mo*	*-mo*	*-mo*	*-mo*
imcompletive/future	**-po*	*-ho*	*-po*	*-ho*
causative	**pa-*	*pa-*	*po-; pa-*	*fo-*
factitive	**p(a,o)ka-*	*hoko-*	*piko-; pika-*	*feka-*
reciprocal prefix	**po-*	*po-*	*po-*	*po-*
accidental passive	**ti-*	*te-*	*ci-*	*ti- ~ te-*
locative applicative	**-Ci*	*-Ci*	*-Ci*	*-Ci*
temporal 'when' (in combination with a possessive suffix)	**sa-*	*sa-*	*sa-*	*sa-*

5 Lexical innovations

In this section I present a number of lexical innovations in the Muna-Buton group. If an etymon is found in Tukang Besi and in Muna or Cia-Cia, but not in Wolio (or other Celebic languages), this clearly points to a lexical innovation and the etymon can be reconstructed for Proto Muna-Buton. But many words which look like lexical innovations in Muna-Buton have cognates in Wolio. If the hypothesis is correct that Wolio is not part of the Muna-Buton group but a relatively recent arrival on the island of Buton, then the presence of such a cognate could be accounted for in three ways:

(a) The word is a Muna-Buton etymon and has been borrowed into Wolio.

(b) The word was an original Wolio word and has been borrowed into the neighbouring Muna-Buton languages.

(c) The word is not an innovation at all, but a reflex of an older root (going back to Proto Celebic or PMP) which is directly inherited in both Wolio and Muna-Buton.

Distinguishing between these three options will be the main challenge for future comparative work in the Muna-Buton area.

In §5.1 I present lexical innovations in Muna-Buton without known cognates in Wolio, while in §5.2 I present possible lexical innovations with Wolio cognates. All Wolio lexical material is taken from Anceaux (1987). I offer this list somewhat hesitantly, as several etyma may turn out to have cognates in non-Muna-Buton languages and hence need to be removed. On the other hand, if the main thesis of this article is correct, these deletions will be hopefully be balanced by additions as new lexical material for these languages becomes available.

5.1 Muna-Buton lexical innovations (without known Wolio cognates)

***ambe** 'open, remove'
Cia *ambe* 'open, uncover', Tuk *ambe* 'change clothing, remove skin or husk'

***ɓai** 'friend, companion'
Mun Las *ɓai* 'friend, companion' (Tuk *wai* 'mosquito' and Cia *wai* 'gnat' are probably not cognate)

***ɓeka** 'cat'
Mun Cia Las Tuk *ɓeka* (Tolaki *beka* is probably a loan from Mun)

***ɓoru** 'k.o. palm tree'
Mun *ɓoru* 'palmyra tree', Cia Las *ɓoru* 'k.o. tree' (unidentified), Tuk *ɓoru* 'sago tree sp.'

***gande** 'give a lift to'
Mun Cia Tuk *gande*

***kaɓi** 'break, throw away'
Mun *kaɓi* 'break', Cia Las *kaɓi* 'throw away', Tuk *kaɓi* 'throw away, discard'

***kape** 'shoulder joint; broken (of arm or leg)'
Mun *kape* 'shoulder joint', Las *kape* 'paralysed (of arm)', Tuk *kape* 'wing'

***kapo** 'enough, full'
Mun *kapo* 'enough', Cia *kapo* 'full' (?), Las *kapo* 'cured of a habit' (Indonesian *jera*), Tuk *kapo* 'full (stomach)'

***kawea** 'wind'
Mun Kai Cia (Pasarwajo) TukKTB *kawea*, Bus *ʔawea*

***kenta** 'fish'
Mun Tuk *kenta*, Bus *ʔinta* (*i* irregular)

***kompa** 'eel'
Mun Cia Las Tuk *kompa*

***konta** 'hold'
Mun *konta* 'hold back, restrain', Cia *konta* 'efficient (of prayers)' (if cognate), Tuk *konta* 'hold, grasp'

***leŋke** 'sexually different'
Mun *leŋke* 'infertile, impotent', Cia Las *leŋke* 'effeminate male', Tuk *leŋke* 'homosexual'

***mena** 'hot, burn'
Mun *mena* 'catch fire, burn, on fire', Tuk *mena* 'hot'

***moapa** 'why?'
Mun *noafa* (with *-mo-* hidden as class prefix), Cia *moapa*, Las *moʔapa* (glottal unexplained), Tuk *nohaʔa* (with metathesis; *n* irregular) [cf. PMP **apa* 'what']

***poroqu** 'drink'
Mun *foroghu*, Bus *foɣoɣu* (with assimilation of second *ɣ*; expected ***foɣohu*), Kai *poʃohu*, Cia *poroku*, Tuk *moroʔu* (*m* irregular) (possibly cognate with Proto Sangiric **dou* 'thirst' Sneddon 1984:79; cf. Tuk *motindoʔu* 'thirsty')

***posolo** 'scabbard, sheath'
Mun *pusolo* 'penis sheath', Las Tuk *posolo* 'scabbard'

***potu** 'head'
Mun Bus *fotu* 'head', Kai *potu* 'head', Cia *pocu* 'head', Las *pocu* 'hair', Tuk *hotu* 'hair'

***puhoi** 'termite nest'
Mun Las *puhoi*, Tuk *puʔoi*

***qawa** 'get, receive'
Mun *ghawa*, Tuk *ʔawa*

***rimba** 'quick'
Mun *rimba*, Bus *ma/ɣimba*, Cia *ma/rimba*, Tuk *me/rimba*

***sula** 'burn weeds'
Mun Cia Tuk *sula*, Las *sule* (final *e* unexplained)

***toɗe** 'flee'
Cia *toɗe* 'run' Tuk *toɗe* 'flee'

***tonduri** 'object that sinks'
Mun Cia *tonduri* 'to bury in the sea, sink s.t.', Las Pnc (Kambowa) *tonduri* 'stone', Tuk *tonduri* 'hook for fishing'. Related to PMB ***tondu** 'sink, drown'.

***tuqo** 'fell, cut down'
Mun *tugho*, Tuk *tuʔo*, Las *cuko* 'stump of a felled tree'

***weŋka** 'split open fruit'
Mun *weŋka* 'split open; half (of coconut)', Las *weŋka* 'split open; betelnut', Tuk *weŋka* 'betelnut'

***woru** 'down, under'
Mun *woru* 'bent down, curved (of branches), Cia *woru* 'under', Tuk *woru* 'underneath'

***wuŋa-nu lima** 'finger' (lit: 'flower of hand')
Mun *wunano lima*, Bus Cia Las *wuŋano lima*, Tuk *wuŋanu lima* (but also Bungku *funga lima*, Wawonii *wunga lima*; possibly borrowed)

***wuŋa-nu qaqe** 'toe' (lit: 'flower of foot')
Mun *wunano ghaghe*, Bus *wuŋano hahe*, Cia Las *wuŋano kake*, Tuk *wuŋanu ae*

5.2 Possible Muna-Buton innovations, with Wolio cognates

***agori** 'immediately'
Mun Cia *agori* 'hurriedly, quickly', Tuk *agori* 'immediate' (also Wol *agori* 'do hastily, do speedily, speed up')

***anano (losu, (ka)tumbu(qa))** 'pestle'
Mun *anano katumbu*, Kai Cia *anano losu*, Tuk *anano tumbu?a* (also Wolio *anana nosu*)

***aso** 'sell'
Mun Cia Bus *aso*, Tuk *?aso* (glottal unexplained) (also Wol *aso*)

***ɓaguli** 'marble'
Mun Cia Tuk *ɓaguli* (also Wol)

***ɓake** 'heart, fruit'
Mun Kai Las *ɓake* 'fruit, heart', Cia *ɓake* 'heart', Bus *ɓa'e* 'heart', Tuk *ɓake* 'heart'; *ɓa?e* 'fruit' (glottal unexplained). (Also Wol *ɓake* 'fruit, heart'. Compare Pam (priestly language) *bake* 'fruit')

***ɓale** 'young leaf'
Mun Cia Tuk *ɓale* (also Wol)

***ɓalobu** 'object filled with water'
Mun Cia *ɓalobu* 'bowl with leg, cup (without handle)', Tuk *walobu* 'freshwater pool, sinkhole' (also Wol *ɓaloɓu* 'earthen or china jar for storing food')

***ɓosu** 'k.o. water container'
Mun Cia *ɓosu* 'earthenware water jar', Tuk *ɓosu* 'k.o. water container' (also Wol)

***ɓura** 'face powder'
Mun Cia Las *bura* Tuk *ɓura* (also Wol *ɓura*)

***ɓuso** 'smithy, bellows'
Mun *buso* 'smithy', Tuk *ɓuso/?a* 'bellows used when forging' (also Wol *ɓusoa* 'bellows, metal-casting house, funnel')

***ɗaoa** 'market'
Mun Cia Tuk *ɗaoa* (also Wol)

***ɗapi** 'twin'
Mun Cia *rapi*, Tuk *ɗapi* (also Wol *rapi*, Pam *rapi*)

***ɗawu** 'give, share'
Mun *ɗawu* 'divide, share', Cia Las *ɗawu* 'give', Tuk *ɗawu* 'portion' (also Wol *ɗawu* 'part; give, provide')

***ɗola** 'move along the surface'
Mun *ɗola* 'creep along the ground', Cia *ɗola* 'creep, surf on waves with a canoe', Tuk *ɗola* 'float' (cf. also Mun *sola* 'crawl on hands and knees') (also Wol *ɗola* 'writhe, wriggle, twist, wind')

***gai** 'pull in (a net?)'
Mun *gai* 'fish with a net with small meshes', Cia *gai* 'pull closer with the arm', Tuk *gai* 'pull in' (also Wol *gai* 'pull, draw')

***garaa** 'surprise particle'
Mun Tuk *garaa* (also Wol *garaaka*)

***gau** 'desire, wish'
Mun Cia Tuk *gau* (also Wol)

***giu** 'sort, kind'
Mun Cia *giu* 'sort, kind; matter, something', Tuk *giu* 'sort, kind' (also Wol)

***gua** 'pull, push, nudge'
Mun *gua* 'push with the body, nudge, elbow; rebel', Cia *gua* 'nudge, take a person's rights', Tuk *gua* 'pull back, withdraw' (also Wol *gua* 'not following, stubborn, stiff, reluctant, resistant, rebellious')

***hali** 'difficult, expensive'
Mun *hali*, Bus Tuk *mo/hali*, Cia *ka/hali* (also Wol *ma/ali*)

***ka(h)ipu** 'youngest child'
Mun *kahepu*, Bus *ʔaepu*, Kai *kaepu* (< Proto Munic **ka(h)epu*), Cia *kaʔopu*, Las *ka/kaʔopu* (*o* unexplained in Cia and Las), Tuk *kaipu* 'last born'(also Wol *kaepu*)

***kalambe** 'young girl'
Mun Tuk *kalambe* (also Wol)

***kamalo** 'paint'
Mun Cia Tuk *kamalo* (also Wol)

***kanda** 'blue'
Mun *(ka)kanda*, Cia Tuk *ka/kanda* (also Wol *ka/kanda*)

***kanu** 'get ready'
Mun *kanu* 'prepare', Tuk *ma/kanu* 'get ready' (also Wol *ma/kanu*)

***kapera** 'spit'
MunS *kapeha*, Kai *pe/kapeʔa*, Cia *pi/kapera*, TukBo *kapera* (also Wol *pe/kapera* 'spit out red betel-nut spittle')

***kaquabulu** 'coconut shell'
Mun *kaghabulu* (with reduction of medial vowel cluster), Kai *kahua* (with loss of *-bulu*), Cia *kabulu* (reduction of second syllable), TukW *kaʔawulu* 'coconut husk' (reduction as Mun), TukBo *ke'ua* (with loss of *-bulu*) (also Wol *kauwana bulu*)

***karia** 'initiation ritual'
Mun Cia *karia* 'puberty ritual for girls', Tuk *karia/ʔa* 'circumcision festival' (also Wol *kariaa* 'feast')

***kokombu** 'mast'
Mun Tuk *kokombu* (also Wol)

***kumbu** 'fist'
Mun *kumbu/no lima* 'fist', Tuk *kumbu* 'arrow-head fist' (also Wol *kumbu* 'fist')

***lagi** 'temporal adverb'
Mun *lagi* 'all the time, every time', Cia *lagi* 'permanent', Tuk *lagi* 'now' (Given the semantics, this is unlikely a loan from Malay *lagi* 'again') (also Wol 'continue; constantly, steadily')

***lalesa** 'wide, spacious'
Mun Cia Tuk *lalesa* (cf. Wol *lalese* 'wide, spacious')

***laŋu** 'intoxicated, drunk'
Mun *lo/lanu* 'slightly poisoned, intoxicated', MunS *lo/lanu* 'drunk', Cia *mo/langu* 'drunk', Tuk *mo/lango* 'drunk' (final *o* irregular) (also Wol *ma/lango*)

***lego** 'swinging arms'
Mun *lego*, Tuk *lego-lego* (also Wol *lego*)

***maka** 'and then'
Mun Tuk *maka* 'and then', Cia *maka* 'if' (also Wol *maka* 'but')

***marasai** 'difficult'
Mun *marasai* 'poor', Tuk *marasai* 'difficult' (also Wol *marasai* 'difficult, intricate, laborious, in trouble')

***mbaka** 'delicious'
Mun *mbaka*, Tuk *mo/mbaka* (also Wol *ma/mbaka*)

***mbali** 'can, become'
Mun *mbali* 'can, become', Cia *pi/mbali* 'can, become', Tuk *me/mbali* 'all right, OK, fine' (also Wol *me/mbali* 'become, come about')

***mbero** 'to wave, to fan'
Mun Cia *ka/mbero* 'fan', Cia *pi/mbero-mbero* 'to fly (of a flag)', Tuk *mbero* 'gesture to s.o. with hand', *kambero-mbero* 'butterfly' (also Wol *ka/mbero* 'fan')

***mente** 'surprised'
Mun Cia Tuk *mente* (also Wol)

***ndoke** 'monkey'
Mun Cia Tuk *ndoke*, Bus *ndoʔe* (also Wol *ndoke*)

***ŋgilo** 'clean, pure'
Mun *ŋgela* 'clean' (vowels unexplained), *ŋkilo* 'pure, holy' (loss of voicing irregular), Bus *mo/ŋgilo* 'holy', Cia *mo/ŋgilo* 'clean'. (Note: Las *mo/kilo* 'black' and Tuk *kili* 'clean' possibly cognate) (also Wol *ma/ngkilo* 'clean, pure, clear, holy')

***nturu** 'often'
Mun *ne/nturu* 'often', Tuk *me/nturu* 'normally' (also Wol *me/nturu* 'frequent, often')

***paiasa** 'mirror'
Mun *paeasa*, Cia Tuk *paiasa* (also Wol *paiasa*)

***pali** 'turn around'
Mun *pali* 'travel around, wander about', Cia *pali* 'turn aside, turn around', Tuk *pali* 'turn around' (also Wol *pali* 'turn, take a turn')

***pamuru** 'angry, furious'
Mun Cia Tuk *pamuru* (also Wol 'go berserk, run amuck')

***pandaŋa** 'k.o. spear'
Mun Cia Tuk *pandaŋa* (also Wol)

***pogau** 'say, word, language'
Mun Bus Kai Cia Las Tuk *pogau*. (Contains reciprocal prefix *po-*; cf. ***gau** 'make, do') (also Wol)

***pooli** 'able, finish, after'
Mun *pooli* 'can, be able; after', Cia *poʔoli* 'finish; able; after, then', Tuk *poʔoli* 'finish, complete' (also Wol)

***porai** 'fiancee'
Mun Tuk *porai* (also Wol *porae*)

***qopa** 'k.o. yam'
Mun *ghofa* 'yam', Tuk *opa* 'sweet potato' (also Wol *opa*)

***qoti** 'food'
Mun *ghoti* 'cooked rice, food', Tuk *hoti* 'meal given to the poor' (*h* irregular for **q*) (also Wol *hoti* 'food, nourishment')

***rambu** '(fibrous) string'
Mun *rambu* 'fibrous part in fruits', Cia *rambu* 'rope, string', Tuk *rambu* 'string' (also Wol *rambu* 'loose ends of thread along the edge of a piece of weaving')

***rampu** 'burn'
Mun Cia Las *rampu* 'burn', Tuk *rampu* 'burn (firewood), roast' (cf. Wol *rampu* 'scorched, blackened, pitch-black'

***randa-nu lima** 'palm of hand' (lit. 'chest of hand')
Mun Cia *randano lima*, Bus *ɣandano lima*, Tuk *randanu lima* (also Wol *randana lima*)

***randa-nu qaqe** 'sole of foot' (lit. 'chest of foot')
Mun *randano ghaghe*, Bus *ɣandano hahe*, Cia *randano kake*, Tuk *randanu ae* (also Wol *randana ae*)

***sagaa** 'sometimes'
Mun *sigaa ~ segaa*, Cia *agaʔa*, Tuk *sagaa ~ sagaʔa* (Compound of ***sa-** 'one' and ***gaa** 'part, separate') (also Wol *sagaa*)

***saŋka** 'complete'
Mun *saŋka* 'complete, ready, finished', Cia *saŋka* 'complete', Tuk *saŋka* 'pass, exceed' (also Wol *saŋka* 'complete, comprehensive, perfect')

***saori** 'very, too much'
Mun *soori*, MunS *saohi* 'serious; defeat', Cia *sauri* 'too much', Tuk *saori ~ sauri* 'very' (also Wol *saori* 'bad, serious (of illness), disobedient, misschievous')

***sepa** 'kick'
Mun Cia Tuk *sepa* (also Wol)

***sundu** 'command'
Mun *sundu* 'overwork s.o., make s.o. do slave work', Tuk *sundu* 'command' (also Wol *sundu* 'consider, think about, remember (harmfully, of spirits of the deceased)' — if cognate)

***taliku** 'behind, back'
Cia Tuk *taliku* (also Wol)

***tara** 'stay, endure'
Mun *tara* 'endure, hold out', Cia *tara* 'endure; *pa/tara* 'live, stay; stop;' Tuk *tara* 'depart' (The words are probably cognate, but the semantics is unclear) (also Wol 'make a firm stand, hold out, be stable; stand, endure')

***tiɗo** 'delouse s.o.'
Mun Cia Tuk *tiɗo* 'kill (lice) by crushing between thumb and finger' (also Wol *tiɗo/ki*)

***tonde** 'drinking vessel'
Mun Cia *tonde* 'drinking glass', Tuk *tonde* 'cup and saucer' (also Wol *tonde* 'glass')

***totumbu** 'house post'
MunS *totumbu*, Cia *cucumbu*, Tuk *totumbo* (final *o* unexplained) (also Wol *tutumbu*)

***tula-tula** 'story'
Mun Bus Tuk *tula-tula*, Cia *cula-cula* ~ *cucula* (also Wol *tula-tula*)

***tuwu** 'classifier for clothes'
Mun *tuwu*, Cia *cuwu*, Tuk *uwu* (loss of **t* unexplained) (also Wol *tuwu*)

***umba** 'appear, rise up'
Mun Cia *umba* ~ *omba*, Tuk *umba* (also Wol *umba* 'come, arrive')

***wulelu** 'eel'
Mun *wulelu* 'moray, river snake', Bus Cia Tuk *wulelu* 'eel' (also Wol *wulelu* 'eel')

6 Conclusion and remaining issues

Even though not all of the proposed lexical innovations may stand up to scrutiny, I believe there is very strong evidence to consider Tukang Besi to be part of the traditional Muna-Buton group. Shared sound changes, irregular phonological developments, the pronominal system, the demonstrative markers and a considerable number of lexical innovations constitute the firm basis for this classification. New grammatical and lexical evidence will likely emerge which will confirm this position. A position challenging the inclusion of Tukang Besi within Muna-Buton will somehow have to account for all these similarities.

However, Tukang Besi also has a large number of unique features not shared by any of the other Muna-Buton languages. The most conspicuous of these is the use of the articles *na* and *te* before noun phrases (see detailed discussion in Donohue 1999). Other such features include medial gemination of certain consonants and a considerable corpus of unique lexical material in daily vocabulary which accounts for the relatively low cognate percentages with the other Muna-Buton languages: around 31–37% with Muna and 40–48% with Cia-Cia (figures from Donohue 1993). These features are probably local innovations, although the articles may have retained information from an earlier protolanguage. Further comparative work is needed to determine which features are retentions, which ones are innovations and which have been borrowed. The same is true for Wolio, which has never been the subject of a thorough comparative investigation. Its 'expulsion' from the Muna-Buton group is almost exclusively based on phonological evidence, but a thorough historical phonology of Wolio has not yet been undertaken.

Two other issues need to be addressed. First, there is the question of what is the next higher macrogroup of Proto Muna-Buton. Elsewhere (van den Berg 1996) I have proposed that Muna-Buton, Bungku-Tolaki and Kaili-Pamona may be part of a Celebic supergroup.

Now that a substantial part of Proto Bungku-Tolaki has been reconstructed (Mead 1998), fresh material is available to test this hypothesis.[6]

The other issue concerns the homeland of the speakers of Proto Muna-Buton. Since little or no archaeological work has been done in South-east Sulawesi, conclusions can only be drawn on the basis of current language distribution and oral tradition. It seems clear that the Muna-Buton area was populated from the east and the south, rather than from the north through the Sulawesi mainland. Bhurhanuddin (1979:47) argues for southern Muna as the homeland for the Muna-Buton group, in which he includes only Muna, Pancana and Cia-Cia (he excludes Wolio and Kamaru, but also Lasalimu and Tukang Besi). Based on the linguistic evidence I put forward the hypothesis that the Proto Muna-Buton homeland was in eastern Buton, around the present-day area of Lasalimu and Kamaru. In Lasalimu there is an oral tradition that this area was an important political centre long before the Wolio-speaking people built the *kraton* in Baubau. This may well reflect historical reality. If so, a possible scenario could be that once this area was settled, one group crossed over to the Tukang Besi islands, while the majority stayed on Buton (the ancestors of Proto Nuclear Muna-Buton). They colonised south and central Buton and crossed over to southern Muna as well. Northern Buton had already been partly occupied by speakers of Bungku languages (Kulisusu and Taluki), but the remainder of Muna and Buton was probably empty. The west coast of northern Buton was colonised by speakers of Muna in relatively recent times through back-migration. Going even further back in time, I speculate that speakers of Proto Muna-Buton possibly originated around the Tolo bay area (the present day Mori homeland), where they left their Celebic kin behind and sailed southwards towards Buton. Putting a date to this event is even more speculative, but somewhere in the first millennium is probably not too far off the mark.

---

6 It has been argued by Hull (1998) in a lengthy article that the languages of Timor (both East and West) are in fact most closely linked with those of the Muna-Buton and Bungku-Tolaki group in a macrogroup which he calls 'Moributonic'. This putative macrogroup comprises all the Austronesian languages of Timor, plus the Muna-Buton and Bungku-Mori languages. Hull boldly asserts that the Austronesian colonisation of Timor took place via Southeast Sulawesi in two waves (illustrative maps accompany his article). The 'Austromunic' languages of Timor (Mambai, Kemak, Tokodede and Idalaka) are claimed to be the descendants of an 'Old Munic' subgroup. Their closest relatives are the Munic languages (Muna, Pancana, Busoa and Kaimbulawa). The other wave are the descendants of 'Old Fabronic' (from Latin *faber* 'blacksmith' = *tukang besi*), which led to modern-day 'Austrofabronic' languages such as Roti, Helong, Dawan, Tetum, Galoli, Wetar, Habu and Kawaimina. They are most closely linked with the Tukang Besi languages.

If Hull's thesis is correct, the Muna-Buton group should be expanded dramatically. However, quite apart from the sometimes outlandish terminology, there are considerable problems with his work. The evidential basis on which his bold statements are made is seriously flawed. The subgrouping hypothesis is solely made on the basis of apparent lexical similarity and little or no attempt is made to treat phonological correspondences or semantic change systematically. The three 'phonological considerations' for Austromunic (Hull 1998:149), for example, simply do not stand up to scrutiny. The chart of Proto Moributonic phonology also raises many questions. There is indeed a wealth of lexical data (but much of it is based on wordlists filled in by non-linguists and should therefore be used with considerable caution), but the methodology used is too weak to warrant Hull's far-reaching conclusions.

Language abbreviations and sources of data

Bus	Busoa	personal field notes
Cia	Cia-Cia	van den Berg (1991c), Daniël Vermonden (pers. comm.)
Kai	Kaimbulawa	personal field notes
Las	Lasalimu	personal field notes, Donohue unpublished field notes
Mun	Muna	van den Berg (1989, 1991a,b,) van den Berg and La Ode Sidu (1996)
MunS	Southern dialect	
Pam	Pamona	Adriani (1928)
PAn	Proto Austronesian	
PBT	Proto Bungku-Tolaki	Mead (1998)
PMB	Proto Muna-Buton	
PMP	Proto Malayo-Polynesian	
PNMB	Proto Nuclear Muna-Buton	
Pnc	Pancana	personal field notes
Tuk	Tukang Besi	Donohue (1999, 2000)
TukW	Wanci dialect	
TukK	Kaledupa dialect	
TukT	Tomea dialect	
TukBi	Binongko dialect	
TukBo	Bonerate dialect	
Wol	Wolio	Anceaux (1952, 1987); Alberth (2000)

References

Adriani, Nicolaus, 1928, *Bare'e-Nederlandsch woordenboek met Nederlandsch-Bare'e register.* Leiden: Brill.

Adriani, Nicolaus (with Albert C. Kruyt), 1914, *De Bare'e-sprekende Toradja's van Midden-Celebes. Derde deel: Taal- en letterkundige Schets der Bare'e-taal en Overzicht van het Taalgebied Celebes-Halmahera.* Batavia: Landsdrukkerij.

Alberth, S.Pd., 2000, A grammar of the Wolio language. Unpublished Master's thesis, University of Western Australia.

Anceaux, J.C., 1952 [1988], *The Wolio language: outline of grammatical description and texts.* 's Gravenhage: Martinus Nijhoff.

1978, The linguistic position of South-East Sulawesi: a preliminary outline. In S.A. Wurm and L. Carrington, eds *The Second International Conference on Austronesian Linguistics: proceedings*, 275–283. Canberra: Pacific Linguistics.

1987, *Wolio dictionary (Wolio–English–Indonesian). Kamus Bahasa Wolio (Wolio–Inggeris–Indonesia).* Dordrecht/Providence: Foris.

Berg, René van den, 1989, *A grammar of the Muna language.* Dordrecht/Providence: Foris.

1991a, Muna dialects and the Munic languages: towards a reconstruction. In Ray Harlow, ed. *VICAL 2. Western Austronesian and contact languages: papers from the Fifth*

International Conference on Austronesian Linguistics, 21–51. Auckland: Linguistic Society of New Zealand.

1991b, Muna historical phonology. In James N. Sneddon, ed. *Studies in Sulawesi linguistics.* part 2, 1–28. Jakarta: Badan Penyelenggara Seri Nusa, Universitas Katolik Indonesia Atma Jaya.

1991c, Preliminary notes on the Cia-Cia language (South Buton). In Harry Poeze and Pim Schoorl, eds *Excursies in Celebes. Een bundel bijdragen bij het afscheid van J. Noorduyn als directeur-secretaris van het KITLV*, 305–324. Leiden: KITLV Uitgeverij.

1996, The demise of focus and the spread of conjugated verbs in Sulawesi. In Hein Steinhauer, ed. *Papers in Austronesian Linguistics* No.3, 89–114. Canberra: Pacific Linguistics.

1997, Spatial deixis in Muna (Sulawesi). In Gunter Senft, ed. *Referring to space: studies in Austronesian and Papuan languages*, 197–220. Oxford: Clarendon Press.

Berg, René van den, and La Ode Sidu, 1996, *Muna–English dictionary.* Leiden: KITLV Press.

Bhurhanuddin, B.H., 1979, Bahasa-bahasa daerah di Sulawesi Tenggara. Unpublished typescript, Kendari.

Donohue, Mark, 1993, The pretenders to the Muna-Buton group. Unpublished typescript. (Revision forthcoming Pacific Linguistics).

1999, *A grammar of Tukang Besi.* Berlin/New York: Mouton de Gruyter.

2000, Tukang Besi dialectology. In Charles E. Grimes, ed. *Spices from the East: papers in languages of eastern Indonesia*, 55–72. Canberra: Pacific Linguistics.

Esser, S.J., 1938, Talen. Sheet 9b of *Atlas van Tropisch Nederland.* Batavia: Topografische Dienst in Nederlandsch-Indië.

Grimes, Charles E. and Barbara D. Grimes, 1987, *Languages of South Sulawesi.* Canberra: Pacific Linguistics.

Grimes, Barbara F., ed., 2000, *Ethnologue. Languages of the world,* 14th edition. Dallas: SIL.

Hull, Geoffrey, 1998, The basic lexical affinities of Timor's Austronesian languages: a preliminary investigation. In Geoffrey Hull and Lance Eccles, eds *Studies in languages and cultures of East Timor*, vol.1, 97–202. Sydney: Academy of East Timor Studies.

Mead, David, 1998, Proto-Bungku-Tolaki: reconstruction of its phonology and aspects of its morphosyntax. Unpublished PhD thesis, Rice University.

Mead, David and Tambunan, 1993, Tolaki phonology. Unpublished typescript.

Sneddon, James N., 1984, *Proto-Sangiric and the Sangiric languages.* Canberra: Pacific Linguistics.

Evidence for a Celebic supergroup

DAVID MEAD

1 Introduction

Sulawesi and its offshore islands are home to more than one hundred Austronesian languages. As these languages have become better known, researchers have parcelled them into from four to eleven subgroups. Following the most current scholarship I recognise the following ten subgroups (see map for their distribution).[1] An estimate of the number of languages in each group follows in parentheses.

Sangiric (5)
Minahasan (5)
Gorontalo-Mongondow (9)
Tomini-Tolitoli (11)
Kaili-Pamona (16)
Saluan-Banggai (5)
Bungku-Tolaki (15)
Muna-Buton (12)
Wotu-Wolio (5)
South Sulawesi (29)

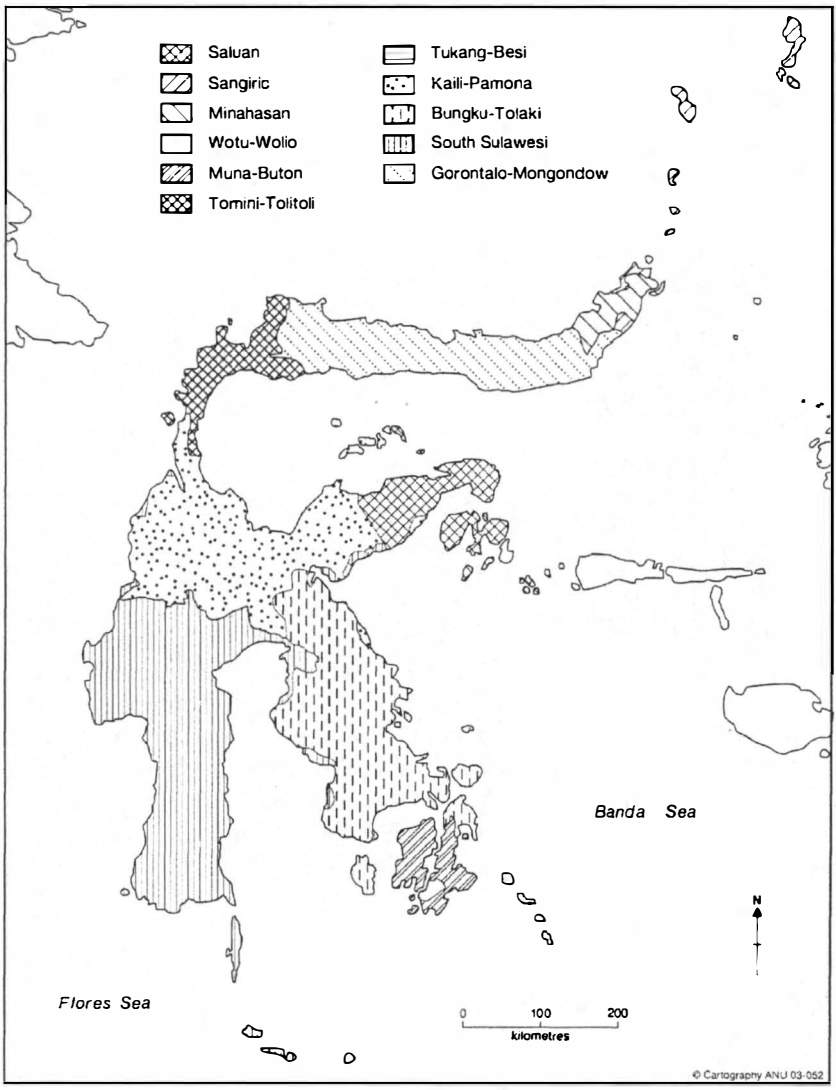
This paper is concerned with six of these subgroups, which collectively cut a broad swath across central and south-eastern Sulawesi — the Tomini-Tolitoli, Kaili-Pamona, Saluan-Banggai, Bungku-Tolaki, Muna-Buton and Wotu-Wolio subgroups. I propose that these groups are genetically related, composing what is here called the Celebic supergroup. My approach is bottom-up: I first show that the Bungku-Tolaki and Muna-Buton languages deserve to be united under a single node, the South-eastern Celebic macrogroup (§3). While it is clear that this macrogrouping includes the Tukang Besi languages, but the position of Tukang Besi within this

[1] Apart from the inclusion of Banggai with other Saluan languages (Mead this volume), and the splitting off of the Wotu-Wolio languages from Muna-Buton (Donohue in press), these are the same ten 'established microgroups' outlined in Sneddon (1993). For a summary of some other classification schemes of Sulawesi languages not discussed in this paper, see Mead (1999:179–180).

John Lynch, ed. *Issues in Austronesian historical phonology*, 115–141.
Canberra: Pacific Linguistics, 2003.

group requires further discussion (§4). In turn, the South-eastern Celebic languages group closely with the Saluan-Banggai languages under an Eastern Celebic node (§5). Finally, at the most inclusive level I adduce evidence for a Celebic supergroup comprising the newly proposed Eastern Celebic macrogroup along with the Tomini-Tolitoli, Kaili-Pamona and Wotu-Wolio subgroups (§6). The notion of a Celebic supergroup in turn refines our view of the Austronesian settlement of Sulawesi. I comment on this in the conclusions (§7), and list questions for further research.

Before proceeding to subgrouping arguments, I first review the historical and comparative literature concerning Sulawesi languages.



Map 1: The Celebic supergroup

2 Historical and comparative studies of Sulawesi languages

Rigorous historical and comparative studies of Sulawesi languages have been undertaken only in the past thirty years. Beginning with Mill's (1975) work on Proto South Sulawesi, reconstructions of protolanguages have also appeared for Minahasan (Sneddon 1978), Sangiric (Sneddon 1984), Gorontalo-Mongondow (Usup 1986), Kaili-Pamona (Martens 1989b) and Bungku-Tolaki (Mead 1998), while initial historical and comparative work has also been carried out in Muna-Buton (van den Berg 1991a, 1991b, and this volume), Saluan-Banggai (Mead this volume) and Wotu-Wolio (Donohue in press). All told, then, only one language group of Sulawesi has yet to come under the attention of historical and comparative linguists — the Tomini-Tolitoli languages spoken in northwestern Sulawesi.

These studies have brought to light a considerable amount of data on Sulawesi languages. They have refined our notions of subgroup boundaries and have done much to clarify our understanding of relationships within such lower-level groupings. Concurrent attempts to establish higher-level connections *between* subgroups, however, have largely been unsuccessful.

Starting from his work in South Sulawesi, Roger Mills (1975:517ff.) investigated but could find no reason for macrogrouping South Sulawesi languages with the Kaili-Pamona, Bungku-Tolaki or Muna-Buton groups, which he collectively referred to as 'Toraja' languages (the existence of a Wotu-Wolio group separate from Muna-Buton was unknown to Mills). Later Mills (1981) sketched out four major subgroups across the island of Sulawesi, namely the South Sulawesi languages, the 'Toraja' languages, the North Sulawesi languages, and finally the Saluan languages. His groupings, however, were mostly impressionistic. He speculated, for example, that the Saluan languages including Banggai were ultimately connected with Philippine languages, and might be relatively recent arrivals in Sulawesi. He left the Tomini-Tolitoli languages out of consideration owing to the small amount of material then available. Ülo Sirk (1981) reached the same conclusions as Mills concerning South Sulawesi languages. While Sirk identified several 'old' lexical items which South Sulawesi languages shared with their neighbours, particularly Kaili-Pamona languages, and structural similarities which they shared with languages of South-east Sulawesi, he concluded that such similarities merely pointed to a long period of contact.

Since those early days, other important contributions have been made to our understanding of macrogrouping of Sulawesi languages.

In 1989 James Sneddon published the results of his comparison of North Sulawesi languages from the perspective of both historical sound change and lexical innovations. In this careful study, he found no basis for grouping the Gorontalo-Mongondow, Minahasan or Sangiric languages with each other, nor did he find support for grouping any of these three with other language groups of Sulawesi. He concluded that 'the search for close affinities [of these three microgroups] must be directed northward, to the languages of the Philippines' (Sneddon 1989:103)

Work in that direction was already in progress. In 1991, Blust identified the Gorontalo-Mongondow languages (but not the Minahasan or Sangiric languages), as belonging to his newly proposed Greater Central Philippines macrogroup. In essence, the Gorontalo-Mongondow languages are relatively recent arrivals in Sulawesi and share a closer genetic affiliation with Tagalog and other Philippine languages than they do with any language group on Sulawesi.

Van den Berg (1996b) is notable as the only author to sketch out a possible basis for a Celebic macrogroup, corresponding to Mill's 'Toraja' group, while at the same time he left open the question whether this group might also include the Tomini-Tolitoli and Saluan languages with Banggai. Although the issue of subgrouping was tangential to the thrust of his paper, van den Berg was on track in regard to certain sound changes, namely the loss of consonant clusters, the monophthongisation of PMP final diphthongs, and the shift of PMP **e* (schwa) to a back rounded vowel. Final consonant loss, however — the change which van den Berg gave first in his list of shared phonological innovations — cannot be used for subgrouping. As I have argued elsewhere (Mead 1996), final consonant loss across central and south-eastern Sulawesi must be an areal feature, not a shared innovation, and this weakens van den Berg's overall argument. Below I refine his notion of a Celebic macrogroup, as well as adduce new evidence which allows the Tomini-Tolitoli and Saluan-Banggai languages to be brought into it.

Finally, in his forthcoming article 'The pretenders to Muna-Buton', Donohue demonstrates from historical sound change that five Muna-Buton languages belong in their own subgroup, which he labels the Wotu-Wolio group after the two most prominent of these five languages. Unlike the other language groups of Sulawesi which occupy geographically contiguous areas, the Wotu-Wolio languages are spoken in widely separated enclaves. Wolio and Kamaru are spoken on the island of Buton in South-east Sulawesi, Kalao and Laiyolo (including Barang-Barang) are spoken on and in the vicinity of Selayar Island off the southern coast of South Sulawesi, while the fifth, Wotu, is spoken at the northern tip of Bone Bay. At the same time, Donohue was unwilling to place the Tukang Besi languages either within his Wotu-Wolio group or with the remaining Muna-Buton languages, nor has he chosen to comment elsewhere on the classification of Tukang Besi.

3 South-eastern Celebic

On the heels of my work on the Bungku-Tolaki languages (Mead 1998), I suggested that these languages probably link closely with the Muna-Buton languages. Elsewhere I have referred to this grouping as 'South-eastern Celebic' (Mead 2001, 2002). I would now like to put this grouping on a firmer footing by citing the sound changes which are shared by all the indigenous languages of south-eastern Sulawesi, excluding Wolio and Kamaru.

This grouping includes Tukang Besi. The four principle Tukang Besi Islands stretch in a south-eastward direction off the southern coast of Buton Island, at the extreme tip of the south-eastern peninsula of Sulawesi. While there is some dialect chaining, following Donohue (2000) it is possible to recognise two Tukang Besi languages, Tukang Besi North spoken on the islands of Wanci and Kaledupa, and Tukang Besi South, spoken on the islands of Tomea and Binongko (and including Bonerate, spoken on islands to the south of Selayar Island in South Sulawesi).

The following twelve changes have been discussed at length by Mead (1998) for Bungku-Tolaki languages and by van den Berg (1991a, b) for Muna and its dialects. Van den Berg (this volume) also mentions most of these changes in regard to Tukang Besi. Where known to me, I include data from Kioko by way of exemplifying another Munic language, and from Cia-Cia for a Butonic language. Bungku-Tolaki data is from Mead (1998). Proto Muna forms are from van den Berg (1991a), otherwise Muna data is from van den Berg (1996a). Kioko data is from my own field notes. Cia-Cia data is from van den Berg (1991c and pers. comm.). Tukang Besi data

is from Donohue (1999, 2000). (PBT = Proto Bungku-Tolaki, PM = Proto Muna; TB = all Tukang Besi isolects, otherwise Tukang Besi isolects are referred to by island name. Leftmost forms are Proto Malayo-Polynesian (PMP) unless otherwise noted.)

1. Consonant cluster reduction $*C_1C_2 > *C_2$, provided the initial consonant C_1 was not a nasal

**sepsep*	'suck'	> PBT **sosoQ*, Muna *soso*, Cia-Cia *sosopi*
**dutdut*	'pluck'	> PBT **ruruQ*, Muna *ruru*
**tuktuk*	'knock'	> PBT **tutuk-i*, Wanci *tutu* 'pound, smith'
**qali/kali-petpet*	'firefly'	> PBT **olimpopoQ*, Muna, Kioko, Wanci *kalipopo*

In nasal clusters, the nasal assimilated to the point of articulation of the following consonant, compare PMP **demdem* dark' > PBT **rondoma*, Muna *rondo*, Wanci *morondo* 'night'.

2. Loss of PMP $*h > \emptyset$

**hapuy*	'fire'	> PBT **apuy*, Muna *ifi*, Cia-Cia *api*, Wanci, Kaledupa *ahu*
**hikan*	'fish'	> PBT **ikaN*, Wanci *ika*
**kahiw*	'wood'	> PBT **kayu*, Muna, Cia-Cia *sau*, TB *kau* (but Wanci *kau* ~ *kaʔu*)
**buhuk*	'head hair'	> PBT **wuuQ*, Muna, Kioko *wuu*
**dahun*	'leaf'	> PM **roo*, Cia-Cia, TB *roʔo*

Addition of glottal stop between like vowels is a regular feature in Tukang Besi. The Wanci form *kaʔu* 'wood' is problematic, but is insufficient to maintain that Tukang Besi has retained a non-zero reflex of PMP **h* (see further van den Berg in this volume).

3. Rhotacisation of PMP $*d > *r$

**depa*	'fathom'	> PBT **ropa*, Muna *rofa*
**duRi*	'thorn'	> PBT **rui*, PM **ka-rui*, Kioko *xii*, Cia-Cia *rui*, Wanci, Kaledupa, Binongko *ruhi*, Tomea *rihi*, Bonerate *rihu*
**daRaq*	'blood'	> PBT **raRaq*, PM **rea*, Kioko *xea*, Cia-Cia *rea*, TB *raha*
**daqan*	'branch'	> Tolaki *ra/a*, Muna *ragha*, Kioko *kaxaʔa*, Cia-Cia *raha*
**dalem*	'inside'	> PBT **laroN* (<met.), PM **lalo*, Kioko *lalo*, Wanci, Kaledupa, Bonerate *laro* (<met.), Tomea, Binongko *lalo*
PWMP **kiday*	'eyebrow'	> PBT **kire*, Muna *kire*, Kioko *kixe*, Wanci *kire*
**qadep*	'front'	> PBT **aroQ*, Wanci *aropa*, Kaledupa, Binongko *aro*
**tuduR*	'sleep'	> PBT **turuR*, Muna *tuturu*, TB *moturu*
**qudaŋ*	'shrimp'	> PBT **uraN*, Muna *ghura*

Kioko /x/ is a voiceless velar fricative (phonetically in free variation with [h]). Van den Berg (this volume) postulates a split of PMP $*d > *d, *r$, apparently on the basis of a few (clearly minority) cases where **d* remained /d/, for example PMP **duha* > PBT **rua*, Cia-Cia *rua*, but Muna *dua*, *rua-*, Kioko *xudua*, Wanci *dodua*; also PMP **deŋeR* > PBT **rongoR*, but Wanci, Kaledupa *rodongo*, Tomea, Binongko, Bonerate *dongo*.

4. Monophthongisation of PMP final diphthongs **-ay* and **-ey* > **e*

**qaZay*	'chin'	> PBT **ase*, Muna *ghase*, Kioko *ase*, Cia-Cia *hae*
**qatey*	'liver'	> PBT **ate*, PM **qate*, Kioko *ate*, Cia-Cia *hate*
**quey*	'rattan'	> PBT **ue*, PM **que*, Kioko *ue*, Wanci, Kaledupa *ʔue*, other TB *ue*
**m-atey*	'die, dead'	> PBT, PM **mate*, Kioko, TB *mate*

5. Monophthongisation of PMP final diphthongs **-aw* and **-ew* > **o*

**qalejaw*	'day, sun'	> PBT **oleo*, PM **qoleo*, Kioko *oleo*, Cia-Cia *holeo*, TB *ʔoloo*
**babaw*	'over'	> PBT **wawo*, Muna, Kioko *wawo*, TB *wawo*
**kasaw*	'rafter'	> PBT **kaho*, Muna *saho*, Kioko *saʔo*, Wanci *kaso* 'ridge pole'
**behew*	'odour'	> PBT **woo*

6. Backing of PMP **e* (schwa) > **o*

**telu*	'three'	> PBT, PM **tolu*, Kioko, Cia-Cia *tolu*
**qitem*	'black'	> PBT **itoN*, Muna *ghito*; Kioko *ito*, Cia-Cia *kito*
**qatep*	'roof'	> PBT **atoQ*, PM **qato*, Kioko *ato*, Cia-Cia *hato*

7. Lowering of PMP **i* > **e* preceding final **q*

**putiq*	'white'	> PBT **pute*, Muna, Kioko, Cia-Cia *pute*, TB *mopute*
**binehiq*	'seed rice'	> PBT **bine*, Muna, Kioko, Cia-Cia, Wanci *wine*
**uliq*	'return'	> Tolaki *pule* 'return home' TB *pule* 'repeat'
**piliq*	'choose'	> PBT **pile* (but Muna *pili*, from Wolio or Malay?)

The lowering of PMP **i* > **e* preceding final **q* is not recognised by van den Berg (1991a, 1991b), and is presented here for the first time as a regular sound change characterising Muna-Buton and Tukang Besi languages.

8. Raising of PMP pretonic **a* > **o*

**qalejaw*	'day, sun'	> PBT **oleo*, PM **qoleo*, Kioko *oleo*, Cia-Cia *holeo*, TB *ʔoloo*
**qalipan*	'centipede'	> PBT **olipaN*, Wanci *oliha*
**paniki*	'bat'	> PBT **poniki*, Muna, Kioko *ponisi*, TB *honiki*
**baqeRu*	'new'	> Tolaki *woʔohu*, PM **buqou*, Kioko *wuʔou*, Cia-Cia *wukou*, TB *woʔou*

9. Loss of PMP medial **-w-* > *Ø*

**sawa*	'snake, python'	> PBT **saa*, Muna *saa*, Cia-Cia, TB *saʔa*
**hawak*	'waist'	> PBT **aaQ*, Muna *aa*
**tawa*	'laugh'	> PBT **taa*, Muna *futaa*, Kioko *fotaa*
**kawit*	'hook'	> PBT **kaiQ*, Muna, Wanci *kai*

10. Split of PMP **s* > **s*, **h*

**lasuq*	'penis'	> PBT **lahuq*, Cia-Cia, Wanci *lau*
**salaq*	'mistake'	> PBT **halaq*, Muna *hala*, Cia-Cia, TB *sala*
**sabuR*	'scatter'	> PBT **hawuR*, Muna *hewi*
**sa-puluq*	'ten'	> PBT **hopuluq*, Muna, Kioko, Cia-Cia, TB *ompulu*
**tasik*	'sea'	> PBT **tahiQ*, Muna *tehi*, Kioko *te?i*, Cia-Cia, Tomea, Binongko, Bonerate *tai*
**tasak*	'ripe'	> PBT **tahaq*, Muna, Cia-Cia *taha*, Wanci *mota?a*
**qasiRa*	'salt'	> PBT **ohia*, Muna *ghohia*
**pusej*	'navel'	> PBT **puhoy*, Muna *puhe*, Cia-Cia *puse*
**isi*	'contents'	> PBT **ihi*, Muna *ihi*, Cia-Cia *isi*
**siku*	'elbow'	> PBT **hiku*, Muna, Cia-Cia, Kioko, Wanci *siku*
**siwa*	'nine'	> PBT **sio*, Muna, Kioko, Cia-Cia *siua*, TB *sia*
**pisaw*	'knife'	> PBT **piso*, Muna, Kioko *piso*
**beRsay*	'paddle'	> PBT **bose*, Muna, Kioko, Wanci *ɓose*
**esa*	'one'	> PBT **asa*, Muna, Cia-Cia *ise*, Kioko *seise*, Wanci *sa?asa*, Bonerate *asa*, other TB *assa*

The split of PMP **s* into both **s* and **h* is one of the most significant changes for subgrouping South-eastern Celebic languages together, yet it is not unproblematic. A notable feature about this change is that there is no ready explanation for the conditioning environment which caused the split, either in terms of word stress, position within the word, or surrounding phonemes. The two major patterns are either that PMP **s* became **h* in Proto Bungku-Tolaki, /h/ in Muna, and zero in Cia-Cia and Tukang Besi, or else it remained /s/ in all four. Nonetheless, sometimes a mixed pattern is observed. Table 1 is a rearranged presentation of some of the above data, particularly where Cia-Cia and/or Tukang Besi reflexes are known. Forms that exhibit weakening of PMP **s* are shown in bold.

Table 1: Weakening of PMP **s*

		PBT	Muna	Tukang Besi	Cia-Cia
**lasuq*	'penis'	****lahuq***	—	***lau***	***lau***
**sa-puluq*	'ten'	****ho-puluq***	***ompulu***	***ompulu***	***ompulu***
**tasik*	'sea'	****tahiQ***	***tehi***	***tai***	***tai***
**tasak*	'ripe'	****tahaQ***	***taha***	***ta/a***	***taha***
**pusej*	'navel'	****puhoy***	***puhe***	—	*puse*
**isi*	'contents'	****ihi***	***ihi***	—	*isi*
**salaq*	'mistake'	****halaq***	***hala***	*sala*	*sala*
**kasaw*	'rafter'	****kaho***	***saho***	*kaso*	*kaso*
**siku*	'elbow'	****hiku***	*siku*	*siku*	*siku*
**esa*	'one'	**asa*	*ise*	*asa*	*ise*
**siua*	'nine'	**sio*	*siua*	*sosia*	*siua*
**sawa*	'snake'	**saa*	*saa*	*sa?a*	*sa?a*
**beRsay*	'paddle	**bose*	*bose*	*bose*	—

One interpretation of this data is that while the weakening of **s* may have begun in Proto South-eastern Celebic, it continued to diffuse lexically and areally, reaching its fullest extent in Bungku-Tolaki languages, and its least extent in Tukang Besi and Cia-Cia. This account, however, runs into a conundrum. The weakening of **s*, to the extent that it did occur, seems to have largely been completed by Proto Bungku-Tolaki; daughter languages simply fail to exhibit evidence of any lexical or areal diffusion post–Proto Bungku-Tolaki.[2] On the other hand, if diffusion of this change is a recent phenomenon in the Muna-Buton area, then there is no way to account for the high degree of concordance between Muna and Bungku-Tolaki languages. It is also possible, however, that forms such as Cia-Cia *puse* and *isi*, and Cia-Cia and Tukang Besi *sala*, *kaso* and *siku* are later borrowings which have obscured an earlier, more regular state of affairs. In this case a larger proportion of *s*-forms would indicate greater influence from an outside language (for example Wolio) in which PMP **s* never weakened. Until such time as more lexical material becomes available, both from Tukang Besi and other Muna-Buton languages, it may be premature to decide between the two explanations.

11. Depalatalisation of PMP **Z* > **s*

**Zalan*	'road'	> PBT **salaN*, Muna, Wanci, Kaledupa *sala*
**qaZay*	'chin'	> PBT **ase*, Muna *ghase*, Kioko, Wanci *ase* (but Cia-Cia *hae*)
**quZan*	'rain'	> PBT **usaN*, Muna *ghuse*, Kioko *ise*, Wanci *uselau* 'k.o. storm' (but Cia-Cia *kia*)

The Cia-Cia forms, which exhibit PMP **Z* becoming zero, are problematic unless it can be shown that they went through an /s/ stage before being lost. Otherwise, it may be necessary to reconstruct **Z* for Proto Muna-Buton (a difficulty which van den Berg does not address), and hence also for Proto South-eastern Celebic.

12. Depalatalisation of PMP **ñ* > *n*

**peñu*	'turtle'	> PBT **ponu*, Muna, Kioko *ponu*
**wañi*	'bee'	> PBT **hoani*, Muna *ani*
**nia, ña*	'3S possessive'	> PBT **-no*, Muna, TB *-no*

The depalatalisation of PMP **ñ* is best viewed as dependent upon (or going along with) the depalatalisation of **Z*. A parallel case is found in Saluan-Banggai languages, where **ñ* was depalatalised to /n/ only in those languages where **Z* was also depalatalised to /d/ or /s/ (Mead this volume). Cross-linguistically, no language is known to have more nasal stops than oral stops (Ferguson 1963).

A thirteenth and fourteenth sound change could be marshalled in support of a South-eastern Celebic group. The change of PMP **j* > **y*, *Ø* is exhibited by all South-eastern Celebic languages, as well as other Sulawesi languages. This change is discussed at some length in §6.

---

[2] The only forms where present-day Bungku-Tolaki languages differ is where the PMP form contained two occurrences of **s*. In this case phonotactic constraints may have played a further role. Compare PMP **susu* 'breast' > Moronene, Wawonii, Bungku *susu*, but Tolaki *uhu*, Mori and Padoe *uo*.

A split of PMP **b* into /b/ and /w/ is also exhibited by all South-eastern Celebic languages. Even more than the weakening of **s*; however, the weakening of **b* has clearly continued to diffuse into many present-day languages, and it is unclear to what extent this change should be attributed to their common ancestor. See further Mead (1998:35–40), van den Berg (1991b:12 and this volume) and Donohue (in press).

4 Relationships of South-eastern Celebic languages to each other

In his important paper clarifying the boundary between Wotu-Wolio and Muna-Buton languages, Donohue (in press) declined to affiliate the Tukang Besi languages with either group, effectively leaving Tukang Besi 'orphaned' in any classification scheme for south-eastern Sulawesi languages. Van den Berg considered this to be in error. In his paper (this volume), he dismisses two supposed objections to including Tukang Besi with the other Muna-Buton languages, and adduces phonological changes as well as grammatical and lexical evidence for bringing the Tukang Besi languages back into the fold, so to speak, with the other Muna-Buton languages.

While one can argue, as I have above, that the Tukang Besi languages are South-eastern Celebic languages, the question remains whether they share a further, closer relationship with Muna-Buton languages Compare the two diagrams of Figure 1.

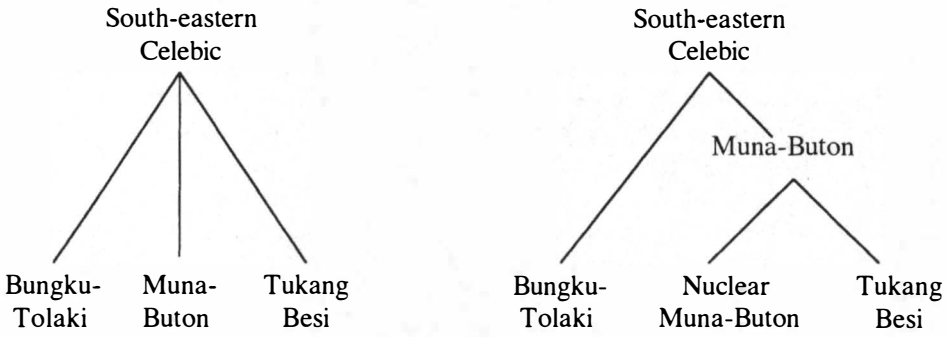


Figure 1: Two views of the classification of south-eastern Celebic languages

In that van den Berg does not intend his Muna-Buton node to also include the Bungku-Tolaki languages, he must support the configuration given on the right. However, in arguing for the position he gives to the Tukang Besi languages, van den Berg generally does *not consider their position also with respect to Bungku-Tolaki languages.* I devote the remainder of this section to a consideration of his claims in this light.

In regard to sound changes which could prove a close relationship, one could hazard that final consonants were lost in the ancestor to (nuclear) Muna-Buton and Tukang Besi. This would necessarily exclude the Bungku-Tolaki languages, which all lost final consonants in a drift-like tendency (reflexes of almost all PMP final consonants are reconstructible for *their* common ancestor; see Mead 1996). Apart from this, there is only one other sound change to my knowledge which supports van den Berg's grouping, namely the fate of PMP initial **w*, which

became zero in Tukang-Besi and the remaining (nuclear) Muna-Buton languages, but became /h/ in Bungku-Tolaki languages (recall from the preceding section that PMP **-w-* became zero in all South-eastern Celebic languages). As can be seen from Table 2, to date evidence that Tukang Besi participated in this change is limited, as far as I know, to two forms of the number 'eight'.

Table 2: Fate of PMP initial **w*

		PBT	Muna	Tukang Besi
**wa-walu*	'eight' (free)	**hoalu*	*oalu*	*oalu*
**walu*	'eight' (bound)	**halu*	*alu*	*alu*
**wakat*	'root'	**haka*	*aka*	—
**(wa)wañi*	'honeybee'	**hoani*	*ani*	—

The picture from sound change is not conclusive. Against this weakly attested change (PMP **w- > Ø*), there are two well-attested changes which could link nuclear Muna-Buton languages instead with Eastern Bungku-Tolaki (!) languages. These changes are the merger of PMP **R* with **y*, and the subsequent loss of **y* (from all sources), usually with fronting of the preceding vowel. See van den Berg (1991b:14–15, this volume) and Mead (1998:114) for details of these changes.

Turning to grammatical changes, van den Berg suggests that Proto Muna-Buton had six pronoun sets. Cognates for five of these pronoun sets are also attributable to Proto Bungku-Tolaki. Only irrealis subject pronouns are not reconstructible for Proto Bungku-Tolaki, though they did develop later as an areal feature in at least four Bungku-Tolaki languages (termed 'future pronouns' in Mead 1998). It would be of greater validity to show that Muna-Buton and Tukang Besi exclusively shared innovations in *reconstructed pronominal forms* — but no Proto Muna-Buton pronoun sets have yet been reconstructed.

Likewise the use of *-um-* to form subject relative clauses is characteristic of Bungku-Tolaki languages (Mead 1998). A requestive prefix *pepe-* with parallel semantics and morphosyntax is also found in Mori Bawah (Esser 1933:326), while in the Kaili-Pamona language Uma, this prefix has the form *pope-* (Martens 1988:184). Only the prefixes **pa-* 'occupation' and **para-* 'iterative' appear not to have cognates in Proto Bungku-Tolaki.[3]

A deictic opposition between *t*-forms meaning 'upwards' and *w*-forms meaning 'downwards' is also found in Tolaki; compare *ikita* 'up there' with *ikua* 'down there' (from earlier **i-ki-wa*). While the opposition has been lost in present-day Padoe, Esser (1927) recorded (among other forms) Padoe deictic adverbs *tehea* 'up there' and *wehea* 'down there' and deictic determiners *ta?a* 'that upwards' and *wa?a* 'that downwards'. Clearly the **t* versus **w* deictic distinction must be a retention from Proto South-eastern Celebic, not an innovation which exclusively links Muna with Tukang Besi.

Therefore while the grammatical evidence does not contradict grouping the Muna-Buton languages with Tukang Besi vis-a-vis the Bungku-Tolaki languages, it cannot be said to argue for such a grouping either. This leaves van den Berg's (rather impressive) list of possible lexical

---

3 Esser (1933:301) lists four Mori Bawah nouns formed with the prefix *pa-*, which he considered to be borrowings from Buginese. Van den Berg's Proto Muna-Buton prefix **para-* may have a cognate in the Uma diffuse prefix *mpara-* (Martens 1988:197).

innovations (including irregular phonological developments in specific lexical items) as the primary basis upon which to posit a close link between the Tukang Besi and other Muna-Buton languages. Like any other initial offering of supposed lexical innovations, van den Berg's list will undergo a process of scholarly refinement, as more scholars become involved. Even though lexical innovations by their nature constitute a weaker kind of evidence, I provisionally accept van den Berg's classification of the Tukang Besi languages. At this point a great deal depends on finding out more about the languages of central and south-eastern Buton, where unfortunately our best data is often still inadequate.

5 Eastern Celebic

If we now take the sound changes which South-eastern Celebic languages share in common, and compare them with surrounding language groups, it becomes apparent that South-eastern Celebic languages are most closely related to languages directly to the north. No less than eight of the fourteen sound changes outlined in §3 are also shared by the five Saluan-Banggai languages of eastern Sulawesi. The following data have been excerpted from Mead (this volume).

1. Consonant cluster reduction $*C_1C_2 > *C_2$, where C_1 was not a nasal

**tuktuk*	'forge'	> Banggai, Balantak *tutuk*
**gisgis*	'rub'	> Balantak, Andio, Saluan, Bobongko *geges*
**sepsep*	'suck'	> Banggai, Balantak, Saluan, Bobongko *sosop*

2. Loss of PMP $*h > \emptyset$

**hapuy*	'fire'	> Balantak, Andio, Bobongko *apu*, Saluan *apu* ~ *apuu*
**hasaŋ*	'fish gills'	> Balantak, Andio, Saluan, Bobongko *ansang*
**buhuk*	'head hair'	> Banggai *buuk*, Balantak *wuuk*

3. Rhotacisation of PMP $*d > *r$

**duha*	'two'	> Banggai *lua*, Balantak, Andio *rua*, Saluan *ohua?*
**duRi*	'thorn'	> Balantak *ruri?*, Andio *rii?*, Saluan *hii?*
**dahun*	'leaf'	> Banggai *loon*, Balantak, Andio *roon*, Saluan *hoon*, Bobongko *ron*
**daRaq*	'blood'	> Balantak *rara?*, Andio *raa?*
**qudaŋ*	'shrimp'	> Balantak *urang*, Saluan *uhang*
**qadep*	'front'	> Balantak *arop*, Andio *aropon*, Saluan *ahop*
**pawed*	'sew thatch'	> Banggai *paul*, Balantak *paur*
**tuhud*	'knee'	> Banggai *tuul*, Balantak *tuur*, Andio *utur* (<met.)

Note that **r* subsequently became /l/ in Banggai and /h/ in Saluan.

4. Monophthongisation of PMP final diphthongs **-ay* and **-ey* > **e*

**qaZay*	'chin'	> Banggai *ade*, Balantak *asi*, Andio *ade*, Saluan, Bobongko *aje*
**m-atey*	'die, dead'	> Banggai, Andio, Saluan, Bobongko *mate*
**qatey*	'liver'	> Banggai, Balantak, Andio, Saluan, Bobongko *ate*

5. Monophthongisation of PMP final diphthongs **-aw* and **-ew* > **o*

**linaw*	'clear (water)'	> Banggai, Balantak, Bobongko *molino*
**babaw*	'above'	> Banggai, Andio *babo*, Balantak *wawo*, Saluan *bawo*, Bobongko *bafo*
**kasaw*	'rafter'	> Banggai, Balantak *kaso*, Andio, Saluan, Bobongko *kaso?*
**behew*	'odour, stink'	> Banggai *boo*, Balantak *woo*

6. Backing of PMP **e* (schwa) > **o*

**qatep*	'roof'	> Balantak, Andio, Saluan, Bobongko *atop*
**utek*	'brain'	> Balantak, Andio, Saluan, Bobongko *utok*
PWMP **beŋel*	'deaf'	> Banggai, Balantak, Andio, Saluan, Bobongko *bongol*
**telu*	'three'	> Banggai *tolu*, Bolantak, Andio *tolu?*, Saluan *totolu?*, Bobongko *totolu*

7. Lowering of PMP **i* > **e* preceding final **q*

**putiq*	'white'	> Banggai *moute*, Saluan, Bobongko *mopute?*
**piliq*	'choose	> Banggai *ilei*, Andio, Saluan, Bobongko *pile?i*
**binehiq*	'seed rice'	> Balantak *wine?*, Andio, Saluan, Bobongko *bine?* 'seedling'

8. Raising of PMP pretonic **a* > **o*

**laqia*	'ginger'	> Balantak, Andio, Saluan *loiya?*
**paniki*	'bat'	> Banggai *uniki*, Balantak, Saluan *poniki?*
**qasawa*	'spouse'	> Banggai *osoaan* 'to marry', Saluan, Bobongko *osoa*

Taken together, these eight changes suggest a close relationship between Saluan-Banggai and South-eastern Celebic languages. I unite them in an Eastern Celebic macrogroup as shown in Figure 2.

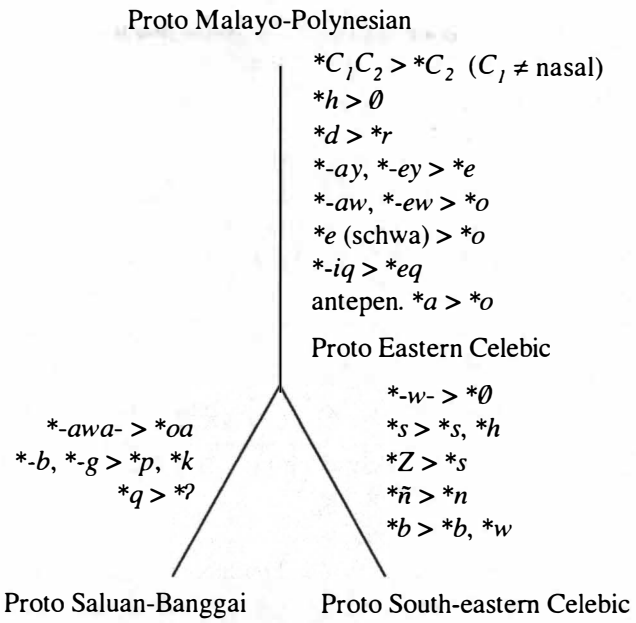


Figure 2: The eastern Celebic macrogroup

The Saluan-Banggai languages are distinguished from Eastern Celebic by three further changes, also detailed in Mead (this volume). Two of these changes are problematic for subgrouping. The change of PMP **q* to glottal stop is widespread in Austronesia. It is found, incidentally, in all Bungku-Tolaki languages, in Tukang Besi, and in many Muna-Buton languages.[4] The value of this change for subgrouping is close to nil. Second, devoicing of final consonants could perhaps even be attributed to Proto Eastern Celebic, but since Bungku-Tolaki and Muna-Buton languages lost final consonants, at this point we simply lack the evidence for attributing this change to a higher level.

Despite the meagre remaining evidence, namely the single change PMP **-awa- > *oa*, I maintain Saluan-Banggai as a separate subgroup. This change distinguishes the Saluan-Banggai languages from South-eastern Celebic languages where **w* was lost word-medially in Bungku-Tolaki, Muna-Buton and the Tukang Besi languages.

South-eastern Celebic languages in turn are distinguished from Eastern Celebic by possibly five further changes, though with caveats discussed in §4.

6 Evidence for a Celebic Supergroup

If we are correct in postulating an Eastern Celebic macrogroup, then it is clear that a 'Toraja' group as conjectured by Mills (1975, 1981) or a Celebic group as envisioned by van den Berg (1996b) encompassing the Kaili-Pamona, Bungku-Tolaki and Muna-Buton subgroups can not be supported. The Bungku-Tolaki and Muna-Buton languages (including Tukang Besi) are more

---

4 Notably PMP **q* remained a uvular fricative in Muna, while in Cia-Cia, Kumbewaha and Kaimbulawa it became /k/ (usually preceding high vowels) or /h/ (usually preceding mid to low vowels).

closely related to the Saluan-Banggai languages on the eastern peninsula of Sulawesi than they are to the Kaili-Pamona languages in the heart of central Sulawesi. Is it possible, though, that the Kaili-Pamona languages could share some relationship, albeit a more distant one, with Eastern Celebic languages?

Of the eight changes listed in §5 which constitute the basis for an Eastern Celebic macrogroup, five are shared by Kaili-Pamona, as well as the Tomini-Tolitoli and Wotu-Wolio language groups. Proto Kaili-Pamona (PKP) reconstructions are from Martens (1989b). Pamona data are from Adriani (1928). Wotu-Wolio data are from Donohue (in press), Laidig and Maingak (1999), Anceaux (1987), and an unpublished Wotu word list collected by Wyn Laidig. Tomini-Tolitoli (TT) data is from Himmelmann (2001). Capital *E* (as in Dondo and Tialo *i?anE* 'fish') represents a paragogic vowel.

1. Consonant cluster reduction $*C_1C_2 > *C_2$, where C_1 was not a nasal

**tuktuk*	'forge'	> Pamona, Wolio *tutu* 'pound, crush' Tialo, Dondo, Totoli, Boano *tutu* 'pound'
**gisgis*	'rub'	> Pamona *gegesi*, Wolio *gigisi*, Taje *gegesi*, other TT *geges*
**sepsep*	'suck'	> Wolio *sosopi,* Dampelas *monosop*, most other TT *sosop*
**qali/kali-petpet*	'firefly'	> PKP **kalipopo?*, Dampelas *kalipopo*, Taje, Boano *alipopo*, Tajio *alipopot*
**kitkit*	'bite'	> Pamona *kiki*, Kalao *kekiti* (but Laiyolo *kikki?* from South Sulawesi influence?)

2. Loss of PMP $*h > \emptyset$

**hapuy*	'fire'	> PKP **apu*, Wotu, Laiyolo, Kalao, Kamaru *apu*, Tialo, Dondo *apiy*, other TT *api*
**hikan*	'fish'	> Lauje *i?ang*, Tialo, Dondo *i?anE*, Boano *ikan*
**buhuk*	'head hair'	> Wolio *buu* 'nape of neck', Totoli, Boano *buok*
**kahiw*	'wood'	> PKP **kayu*, Laiyolo, Kalao *kaju*, Dampelas, Totoli, Boano *kayu*,Tajio *ayu*, other TT *?ayu*

3. Rhotacisation of PMP $*d > *r$

**duha*	'two'	> PKP **ro-*, Laiyolo, Kalao, Wolio *rua*, Taje *rorua*, Tajio *orua* (but PKP **dua*, Wotu *dua-*, *duango*, Balaesang *dorua*, Pendau *doruo*, Lauje *doluo*, Totoli, Boano *dóua*)
**duRi*	'thorn'	> PKP **rui*, Wotu, Laiyolo, Kalao, Wolio, Kamaru *rui*, Lauje, Totoli, Boano *lui*, Tialo, Dondo *lugitE*, other TT *rui*
**daRaq*	'blood'	> PKP **raa?*, Wotu, Wolio, Kamaru *raa*, Laiyolo, Kalao *ra?a*, Lauje, Totoli *laa*, Boano *la?a*, Tialo *laga*, Dondo *laganyo*, other TT *raa*
**deŋeR*	'hear'	> Laiyolo, Kalao, Wolio *rango*, Lauje, Tialo, Dondo *longo*, Boano *longa* (but Pamona *donge*)
**danaw*	'lake'	> PKP *rano, Wolio *rano*, most TT *rano* (but Lauje, Tialo, Boano *dano*)

**dalem*	'deep'	> PKP **rala*, Wotu, Laiyolo, Kalao *lara* (<met.), Taje, Tajio *raro*, Pendau, Lauje, Dondo *lalong*, Tialo *lalongE*, Totoli *lalom*, Boano *laom*
**dahun*	'leaf'	> Uma *rau*, Balaesang, Dampelas, Taje, Pendau, Tajio *roong*, Lauje, Tialo, Dondo *loongE*, Totoli *laeng* (but Boano *da?un*)
**qudaŋ*	'shrimp'	> Pamona *ura*, Balaesang, Tajio *urang*, Boano *ulang*
**kuden*	'cookpot'	> PKP **kura*, Laiyolo *kuro*, Kalao *kura*

Northern Tomini (Lauje, Tialo and Dondo) languages along with Totoli and Boano exhibit the further change of **r* > /l/.

4. Monophthongisation of PMP final diphthongs **-ay* and **-ey* > **e*

**qaZay*	'chin'	> PKP **aje*, Laiyolo, Kalao, Wolio, Kamaru *ade*, Balaesang, Totoli *ngade*, Boano *ngade?*, Pendau *nganje*, other TT *ngaje*
**anay*	'termite'	> PKP **ane*, Wolio *ane*, Balaesang, Tajio *ane* (also Taje, others *siane*)
PWMP **kiday*	'eyebrow'	> PKP **kire*, Muna *kire*, Kioko *kixe*, Wanci *kire*, Balaeasang *kire*
**m-atey*	'die'	> PKP **mate*, Laiyolo, Kalao, Wolio *mate*, Balaesang *maate*, Pendau *naate*, Lauje *mate*, all other TT *ate*
**qatey*	'liver'	> PKP **ate*, Wotu, Wolio, Kalao *ate*, Laiyolo *ati*, all TT *ate*

5. Monophthongisation of PMP final diphthongs **-aw* and **-ew* > **o*

**babaw*	'above'	> PKP **wawo*, Layolo *bafo*, Kalao *bavo*, Wolio *bawo*, Taje *wawo*, Lauje *babo?*, Totoli *babo*
**kasaw*	'rafter'	> PKP **kaso*, Laiyolo, Kalao *kaso*, Dampelas, Taje, Lauje, Tialo *?aso*, Dondo *aso*, Totoli, Boano *kaso*
**pisaw*	'knife'	> Wotu, Laiyolo, Kalao, Wolio, Lauje, Boano *piso*, Tialo, Dondo *pisoyE*
**behew*	'odour, stink'	> Pamona *boo*, Dampelas *noboomo*, Pendau, Lauje, Dondo *boo*, Tialo *memboo*

A sixth change could possibly be added to this list, namely PMP **e* (schwa) > **o*, which is regular in all these languages apart from a number of exceptions where **e* > *a* in Kaili-Pamona and Wotu-Wolio languages (Martens 1989b; Donohue in press). Even if we were to take all five (or even six) changes together, however, their value for defining a larger Celebic group remains low, since all five changes occur relatively frequently in the Austronesian world. There is, however, another change by which these languages may be grouped, and which, when added to the five changes above, makes a strong case for macrogrouping. It concerns the fate of PMP **j*.

PMP **j* has been reconstructed only in word-medial and -final position. Assigning a phonetic value to PMP **j* has been problematic. While Dahl (1981:92, 152) takes the position that **j* was a palatal stop or affricate, other scholars have favoured a velar interpretation. Wolff (1988) considered **j* (his **g*) to be a voiced velar stop, Blust (1990:234) considers it to be a palatalised velar stop [gʲ] which had no voiceless counterpart, while Ross (1992) has argued that it was

likely a voiced velar fricative [ɣ, ɣʲ]. Its reflexes in Tagalog are *-l-* (medially) and *-d* (finally), in Toba Batak *-g-* and *-k*, in Malay *-d-* and *-t*, in Javanese *r* in all positions and in Madurese *l*.

Consider now the realisation of **j* in Sulawesi languages. In the three microgroups of northern Sulawesi, PMP **j* became **d* in final position; compare Proto Gorontalo-Mongondow **pusod*, Proto Minahasan **pusəd* and Proto Sangiric **pusid* (< PMP **pusej* 'navel'). In medial position **j* merged with the reflexes of PMP **Z* in Proto Gorontalo-Mongondow (Usup 1986:277–279), and with reflexes of **Z*, **D* and **d* in Proto Sangiric (Sneddon 1984). Among Minahasan languages, medial PMP **j* and **D* both became Proto Minahasan **d* following **e* (schwa), otherwise PMP **-j-* has the distinct realisation pattern of /l/ in Tonsawang and /r/ in other Minahasan languages (for which correspondence Sneddon reconstructed Proto Minahasan **r*,) (Sneddon 1978; 1989:97–98).

Word-medially PMP **j* must have been maintained as a separate phoneme into Proto South Sulawesi, as it is realised as *-s-* in Buginese but as *-r-* in other languages (on the basis of which Mills reconstructed PSS **z*). In word-final position, however, PMP **-j* presumably became **d* and thence merged with **t* in the general final-stop devoicing which occurred prior to Proto South Sulawesi. From there it is reflected as glottal stop in most present-day South Sulawesi languages (Mills 1975:553, 556).

In contrast to these four language groups—three in the north and one in the south of Sulawesi — across the rest of Sulawesi **j* did not merge with **Z* or **d* in any position. In fact it has only two principal realisations, either as *y* (sometimes resegmented in final position to *i*) or as zero, often accompanied by fronting of the preceding vowel. Because this change is significant for postulating a Celebic macrogroup, I discuss it here in some detail. Table 3 gives reflexes of nine PMP etyma containing **j*, both in medial and in final position. Non-cognate forms (lexical replacements) are indicated by underlining, while a dash (—) indicates a lack of data. Forms in bold receive further discussion below.

On the basis of these cognate sets, I reconstruct Proto Celebic **qapəyo*, **pae* (from earlier **paye*), **qaləyo*, **ipian* (from earlier **ipiyan*), **ngayam*, **quləy*, **laləy*, **pusəy* and **palay*. PMP **j* became **y*, and further, it would appear, became a transition glide and was subsequently lost between /a/ and a front vowel.[5]

From Table 3, we can also note the occurrence of 'irregular' reflexes of **j* in border areas from the influence of surrounding languages. Among the Wotu-Wolio languages spoken in the area of Salayar Island off the southern tip of South Sulawesi, Kalao *asə* 'field rice' (with **j* reflected as /s/) must clearly be ascribed to Buginese influence. Similarly Kalao *pi?du* and Laiyolo *pidu* 'gall' are also likely borrowings from a South Sulawesi language, particularly as the Kalao form exhibits consonant doubling (compare PSS **pizzu*). In Laiyolo *pəllə?* 'palm, sole', final glottal stop is likewise the typical South Sulawesi reflex of PMP **-j*, and so this form must also be ascribed to South Sulawesi influence.

---

5 The simpler statement, that **j* was lost contiguous to a front vowel, is contraindicated by Padoe *penei* 'wing' < Proto Celebic **paniy* < PMP **panij* 'wing' (see Mead 1998:64). Further data — especially a full account of the fate of PMP **y* in medial position — may lead this statement to be refined. In languages which have been investigated to date, PMP **-j-* and PMP **-y-* shared the same fate in Kaili-Pamona (Martens 1989b), Bungku-Tolaki (Mead 1998:67) and Muna (van den Berg 1991b:14–15). Evidence also points in this direction in Saluan-Banggai languages, but (because of a lack of data) is somewhat inconclusive (Mead this volume).

Table 3: Etyma containing PMP **j* and reflexes in selected Sulawesi languages

	'gall, bile' **qapeju*	'field rice' **pajey*	'sun' **qalejaw*	'when'[1] **i-pija-n*	'name' **ŋajan*	'snake' **qulej*	'fly' **lalej*	'navel' **pusej*	'palm' **palaj*
TOMINI-TOLITOLI									
Boano	*poyu*	*pae*	***ondo***	***pilan***	***langan***	*ule?*	*pikot*	***pusol***	***paak***
Totoli	*peu*	*bini*	***ondo***	***pilan***	***ngalan***	*ule*	*lale*	***pisol***	***palak***
Dondo	*poyu*	*bo?ung*	*oloyo*	*sogaubengi*	*tope*	*ule*	*lale*	*puse*	*pale*
Lauje	*poyu*	*bo?ung*	*oleo*	*sogaumbéng*	*tope*	*ule*	*lale*	*puse*	*pale*
Balaesang	*peit*	*boas*	*sekat*	*mpiang*	*tope*	*ule*	*lale*	*puse*	*pale*
Pendau	*opoyu*	*pae*	*eleo*	*nasae*	*sango*	*ule*	*lale*	*puse*	*tanatang*
KAILI-PAMONA									
Da'a	*mpoyu*	*pae*	*eo*	*nepia*	*sanga*	*ule*	*lale*	*puse*	*pale*
Pamona	***poju***	*pae*	*eo*	*impia*	*sanga*	*ule*	*yale*	*puse*	*pale*
Uma	***poju***	*pae*	*eo*	*nto?uma*	*hanga?*	*ule*	***dali?***	*puhe*	*pale*
Napu	***puru***	***pare***	***alo***	***impira***	*hanga*	***ile***	***dale***	—	*palanta*
Bada	***puru***	***pare***	***alo***	***himpirɔ***	*hanga?*	***ile***	***dali?***	—	*palanta?*
Besoa	***puru***	***pare***	***alo***	***impira***	*hanga?*	*leloto*	***dali?***	***pohi?***	*palanta?*
WOTU-WOLIO									
Wotu	—	*bae*	*iyo*	*dipia*	*sanga*	*ulo*	*lale*	—	—
Wolio	*mapai*	*ɓae*	*eo*	*naipia*	*saro*	*ulo*	*lale*	*puse*	*randa*
Kalao	***pi?du***	***asə***	***ajo***	*laipia*	*sanga*	*ulo*	*lale*	—	*pele*
Laiyolo	***pidu***	*bae*	***ajo***	*ripia*	*sanga*	*ulo*	*lale*	*puse*	***pəllə?***

[1] PMP **pija* 'how many' has been lexically replaced in a number of Celebic languages, therefore I have chosen instead to cite reflexes of **i-pija-n* 'when' (which strictly may not be reconstructible to PMP; see Mead 2001:170). The pattern of realisation of **j* is the same, but more obvious from reflexes of the latter.

	'gall, bile' *qapeju	'field rice' *pajey	'sun' *qalejaw	'when' *i-pija-n	'name' *ŋajan	'snake' *qulej	'fly' *lalej	'navel' *pusej	'palm' *palaj
SALUAN-BANGGAI									
Banggai	*sopot*	*labue*	*oloyo*	*noian*	*sambu*	*uloy*	*poos*	*pusoy*	*palalap*
Balantak	*opoyu?*	*pae*	*ilio*	*ipi*	*ngaan*	*ule*	*laale*	*puse*	*palaa*
Andio	*poyu?*	*pae*	*sina*	*ipian*	*ngaan*	*ulo*	*laalo?*	*puse*	*pala*
Saluan	*pou?*	*pae*	*sina*	*hipian*	*sanggo*	*ulo, uloo*	*lalo, laloo*	*pusoo*	*palaa*
Bobongko	*opou?*	*pae*	*dolag*	*torikuka*	*sanggor*	*bintana?*	*lalo*	*puso*	*pala*
BUNGKU-TOLAKI									
Kulisusu	*upeu*	*pae*	*oleo*	*impia*	*ngee*	*ule*	*lale*	*puhe*	*pele*
Mori Bawah	*upeu*	*pae*	*oleo*	*te?ipia*	*ngee*	*ule*	*lale*	*puhe*	*pele*
Padoe	*upeu*	*pae*	*olo*	*te?epie*	*nee*	*ule*	*laloi*	*puhoi*	*palai*
Tolaki	***posu***	*pae*	*oleo*	*te?ipia*	*tamo*	*ule*	*lale*	*puhe*	*pele*
MUNA-BUTON									
Muna	*ghufei*	*pae*	*gholeo*	*naefie*	*nea*	*ghule*	*pepi*	*puhe*	*randa*
Kioko	*piu*	*ɓae*	*oleo*	*naifie*	*kona*	*ule*	*buhoto*	—	*handa*
Cia-Cia	*hopiu*	*ɓae*	*holeo*	—	*ngea*	*kule*	*ɓei*	*puse*	*randa*
Tukang Besi	*ho?ou*	*ɓae*	*?oloo*	*ehia*	*ngaa*	*sa?a*	*lalo*	—	—

Non-*y* reflexes of PMP **j* are also found among the Badaic languages of Central Sulawesi; compare Napu, Bada, Besoa *puru* 'gall', *pare* 'field rice', Napu, Besoa *impira*, Bada *himpirɔ* 'when?', Besoa *pohiʔ* 'navel'. As delineated above, /r/ (medially) and glottal stop (finally) are typical South Sulawesi reflexes of PMP **j*, compare Seko *puru* 'gall', Sa'dan Toraja *pare* 'field rice', Seko Tengah *napiranga* 'when?' and Seko Lemo *posiʔ* 'navel'. Likewise Uma, Bada and Besoa *daliʔ* and Napu *dale* 'fly' exhibit similarities with South Sulawesi languages. The irregular dissimilation of **l* > /d/ observed in these forms is also found across the border in Mamuju and Seko (Mills 1975:748), but since in these South Sulawesi languages PMP **-ej* > PSS **-it* > **-it* > /iʔ/ is *regular*, there can be no doubt about the direction of borrowing. Martens (1989a) considers such cases to reflect a long period of contact and borrowing between the Badaic languages and South Sulawesi languages, particularly Seko. He also considers it probable that Napu, Bada, and Besoa *alo* 'sun, day' was likewise borrowed from South Sulawesi — compare Seko *alo* 'day' (< PSS **ilzo*) — whereas other Kaili-Pamona languages consistently have *eo*.

This large number of supposed borrowings, however, leaves Napu and Besoa *ile* 'snake' as the only form known to me where it could be proposed that PMP **j* passed through a *y*-stage. This form, however, has unexplained **u* > /i/ in the initial syllable. Furthermore, this irregularity is also found in the northern South Sulawesi language area; compare Seko Lemo, Mamuju and Bambam *ile* (Mills 1975:877). While it is possible to account for the presence of non-*y* reflexes of **j* in Badaic languages through borrowing, the sheer number of exceptions suggests an alternative hypothesis: Badaic languages are genetically South Sulawesi languages.[6]

An inspection of Table 3 also reveals that in several cases the Boano and Totoli forms also exhibit unusual (non-*y*) reflexes of **j*. Both of these languages are located in the northern area where Celebic languages border the Gorontalic subgroup, and these forms may reflect influence from Gorontalo. However, until a study of historical sound change in the Tomini-Tolitoli languages is undertaken, the identification of loan words remains problematic. Indeed, the question remains open whether Boano and Totoli should even be subgrouped with the Tomini languages (see Himmelmann 2001:19–20).

---

6 Besides the fate of PMP **j*, other areas of difference to explore include PMP **R* (reflected as zero in other Kaili-Pamona languages but as /r/ in South Sulawesi languages), PMP **-uy* (reflected as /u/ in other Kaili-Pamona languages, but as /i/ in South Sulawesi languages), and PMP **-uq* and **-iq* (vowels were lowered in Proto South Sulawesi, but not in Proto Kaili-Pamona).

The claim that Badaic languages reflect PMP **q*, while this phoneme was lost in Proto South Sulawesi (Mills 1975:518), is a red herring. Unknown to Mills, Seko Padang reflects PMP **q* as length on the vowel, either when **q* occurred in final position or contiguous to a final vowel (in the latter case, presumably via intermediate metathesis/float to final position). Therefore reflexes of PMP **q* must be reconstructed for Proto South Sulawesi regardless. Compare the following data (Laskoswke 1995 and pers. comm.).

PMP		Seko Padang		
**puluq* 'ten'	>	*pulo:*		
**dilaq* 'tongue'	>	*lila:*		
**tuqah* 'old'	>	**tuaq*	>	*tua:*
**taqi* 'feces'	>	**taiq*	>	*tai:*

Apart from these exceptions, only a few other forms would appear to have non-*y* reflexes of PMP **j*. The Uma and Pamona form for 'gall', *poju* [podʒu], and the Tolaki form *posu*, however, are not borrowed, but rather result independently from fortition of **y*, which is regular in defined contexts in Uma and Pamona (Martens 1989b) and also occurred sporadically in Tolaki (Mead 1998:115). The Kalao and Laiyolo form for 'sun', *ajo* [adʒo], is unexplained in the present analysis. As far as I can tell it has not been borrowed from any present-day South Sulawesi language, and perhaps is an inherited form (from earlier **alyo*?).

Finally, while PMP **j* appears to have become zero or merged with PMP **y* in medial position, on the basis of counter-additive reasoning the change of PMP **-j* > Proto Celebic **y* must have occurred *after* the monophthongisation of PMP **-ay* and **-ey* to **e*. Among Celebic languages PMP **-ay* and **-ey* are universally reflected as /e/. And while in many cases PMP **-aj* and **-ej* are also reflected as /e/, enough languages reflect these otherwise to indicate that a merger of PMP **-ay*, **-ey*, **-aj* and **-ej* did not occur.

I would also like to make an initial proffering of two lexical innovations which support the newly proposed Celebic macrogroup. These forms are:

> PCel **panianan* 'parent-in-law' > Totoli *poneanan* 'parent-in-law' (also 'child-in-law'?), Boano *ponianan* 'parent-in-law' (also 'child-in-law'?), Uma *piniana* 'parent-in-law', Kulawi *paniana* 'parent-in-law', Kulisusu *poniana* 'parent-in-law', Moronene, Mori Bawah *poni* 'parent/child-in-law' (reciprocal term). Compare also Moronene, Muna *samponi* 'child's spouse's parent' (with prefix *sa-* 'one').
>
> PCel **manian* 'child-in-law' > Dampelas *maniang*, Lauje *meniang*, Tialo, Dondo *monianE* 'parent-in-law' (also 'child-in-law'?), Pendau *meniang* 'parent-in-law' (*not* reciprocal), Da'a, Kulawi, Lindu, Pamona (Ampana dialect) *mania* 'child-in-law', Uma *minia* 'child-in-law', standard Pamona (Adriani 1928:s.v.), Wolio *mania* 'child/parent-in-law' (reciprocal term), Balantak *monian* 'child/parent-in-law' (reciprocal term), Kulisusu *ana monia* 'child-in-law'.

Only three languages known to me — Kulawi and Uma in the Kaili-Pamona area, and Kulisusu in the Bungku-Tolaki area — have retained both forms (though further lexical research may uncover other languages where this is so). Nonetheless, in that these words constitute a derivationally related pair, their reconstruction is mutually supported. That a number of daughter languages reflect only one member of the pair must be accounted for by semantic shift to a reciprocal term, lexical replacement, or both (compare for example Pendau which now has *meniang* 'parent-in-law' and unrelated *tomodait* 'child-in-law'). For such reasons it is also untenable that present-day forms could have obtained their distribution through borrowing.

7 Conclusions

The principal results of this study are summarised in Figure 3. This figure shows the branches and nodes which have been argued for in the preceding section. As noted there and again below, some branches have a tentative status. At the highest level I propose a group which encompasses almost all the languages of central, eastern and south-eastern Sulawesi as well as a handful of languages located in South Sulawesi. Following van den Berg (1996b), I label this the Celebic supergroup. Notably, van den Berg is the only other author to have speculated about a

macrogroup comprising these same languages. However, he did not define his reasons for including the Saluan-Banggai or Tomini-Tolitoli languages in his macrogroup, and presented a different view of subgrouping within it.

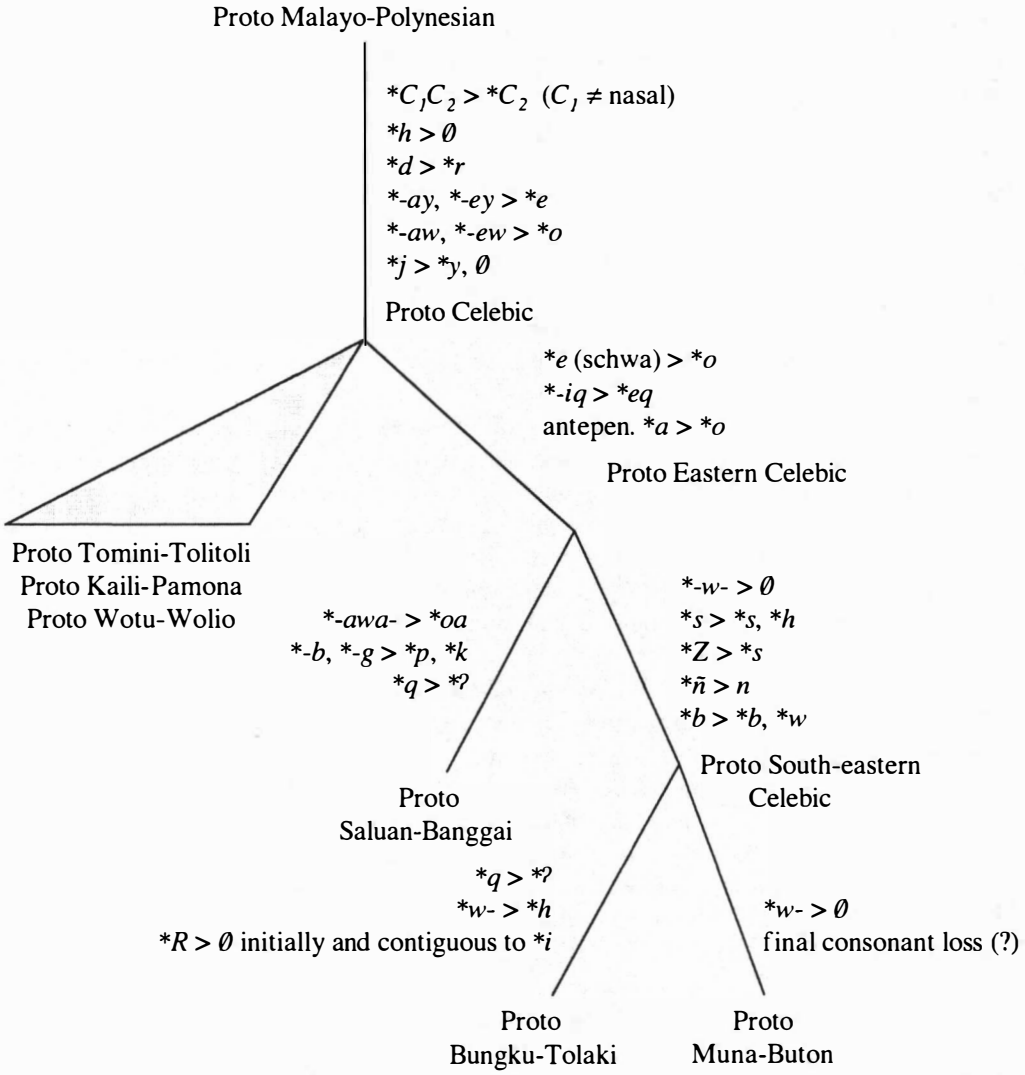


Figure 3: The Celebic supergroup

At present it is unclear how many primary branches should be posited under the Celebic node. The shaded triangle in Figure 3 indicates that while there is evidence for bringing the Tomini-Tolitoli, Kaili-Pamona and Wotu-Wolio subgroups into a Celebic Supergroup, I make no claim as to how these groups may be related to each other. On the other hand, there are three sound changes which would apparently allow us to distinguish Eastern Celebic as a separate branch.

The change of PMP **e* (schwa) > **o* is, however, also broadly characteristic of all Celebic languages. But because some exceptional cases where **e* > **a* are known from Kaili-Pamona and Wotu-Wolio languages, I have not assigned this change to a higher level. The raising of PMP antepenultimate **a* > **o* is also a change of a general nature. This leaves the lowering of PMP **i* > **e* before final **q* as the most significant change for subgrouping Eastern Celebic languages together. To be sure, this same lowering has been noted as an independent change elsewhere, even in other parts of Sulawesi including Proto Sangiric (Sneddon 1989:90), Proto South Sulawesi (Mills 1975:545), and, perhaps not surprisingly, in the Badaic languages (Martens, cited in Sneddon 1989:90). In these other cases, however, the lowering of **i* > **e* before **-q* was accompanied by a parallel lowering of **u* > **o* in the same environment. Only in Eastern Celebic do we find lowering of **i*, *unmatched* by a parallel lowering of **u*.[7]

Despite the encompassing nature of the present study, important questions remain. Here I mention areas for further research in the area of Sulawesi historical linguistics. Several of these issues have been mentioned in the preceding sections, but it seems beneficial to bring them together here. Many of the following comments can be considered additional commentary on Figure 3.

What is the position of Boano and Totoli? Totoli is spoken by 25,000 speakers in the border area where Celebic languages meet Gorontalic languages, and Boano in the same area by about a tenth that many speakers. These two languages are not Gorontalo-Mongondow languages, but if reflexes of PMP **j* are an indication (Table 3 above), they may not fit comfortably into a Celebic group as defined here either. Himmelmann, who has provided a wealth of new data on Tomini-Tolitoli languages, simultaneously cautions that the genetic unity of this group has never been established. In particular he singles out Totoli and Boano for their divergent phonology, lexicon and grammar (Himmelmann 2001:20).

What is the relationship between the Kaili-Pamona and Wotu-Wolio languages? A number of tantalising leads have been proposed, but none have been followed up. The Dutch linguist J.C. Anceaux recognised striking structural similarities between Wolio and Pamona, but never published on this topic (René van den Berg pers. comm.). Donohue (in press) suggests the change of PMP **e* (schwa) > **a* in a number of lexical items (otherwise regularly PMP **e* > /o/) could be used to link the Kaili-Pamona and Wotu-Wolio languages. Even a consideration of geographical location suggests that the Wotu-Wolio languages are more likely to share a closer relationship to Kaili-Pamona than to Tomini-Tolitoli languages. Unfortunately, only Wolio has been well documented, and other Wotu-Wolio languages have been influenced by South Sulawesi languages. A careful study will be needed to pull out the traces of historical connection. To date, not even an internal classification of the five Wotu-Wolio languages has emerged.

What is the position of the Badaic languages? Culturally, speakers of Napu, Bada and Besoa (collectively referred to as the Badaic languages) identify with their Kaili-Pamona neighbours, yet their word stock bears affinities to South Sulawesi languages. Martens (1989) investigated this situation, and concluded that Badaic languages were genetically Kaili-Pamona languages that had borrowed lexically from South Sulawesi languages, particularly Seko. In that Badaic languages overwhelmingly reflect **j* as /r/ (medially) or glottal stop (finally) — the prototypical South Sulawesi reflexes — this study suggests the opposite. That is to say, the Badaic languages

---

7 The lowering of **u* occurred as a further change in two sub-branches of Eastern Celebic, namely Eastern Saluan-Banggai (Mead this volume) and Western Bungku-Tolaki (Mead 1996:80ff.).

may genetically be South Sulawesi languages which have been influenced lexically by Kaili-Pamona languages.

What is the internal classification of South Sulawesi languages? If Badaic languages are genetically South Sulawesi languages, then clearly Mills' (1975:490ff.) internal classification of South Sulawesi languages needs to be reworked. In fact such a re-evaluation has been needed on other grounds. In particular, Mills' internal groupings were based heavily (but not exclusively) on what happened to consonants in word-final position. Experience elsewhere in Sulawesi, however, has shown that processes of final consonant weakening, merger and loss are likely to exhibit areal diffusion, and thus not to be valid indicators of genetic affiliation (Sneddon 1993; Mead 1996). Since Adelaar (1994), it has been clear that the Tamanic languages of Borneo are genetically South Sulawesi languages, most closely related to Buginese. These languages, which have been more conservative with regard to final consonants, are likely to help in sorting out what changes can (or cannot) be attributed to higher genetic levels within South Sulawesi.

Are Muna-Buton and Bungku-Tolaki valid subgroups? Evidence from sound change alone is — and will probably remain — an insufficient basis for establishing a Muna-Buton subgroup. As discussed in §3, a Muna-Buton grouping is provisionally accepted based on the probable lexical innovations set forth in van den Berg (this volume). I have not touched on the genetic unity of the Bungku-Tolaki languages. Although the loss of PMP **R* initially and contiguous to **i* can be attributed to Proto Bungku-Tolaki (not mentioned previously in this paper, but detailed in Mead 1998:58–60), this change will need to taken in hand with the fate of PMP **R* in nuclear Muna-Buton languages. The remaining changes attributed in Figure 3 to Proto Bungku-Tolaki (PMP **w-* > **h*, PMP **q* > glottal stop) are less consequential for subgrouping. Lexical innovations in support of a Bungku-Tolaki subgroup can be found in Mead (1998:86–87). Within the framework provided by this study, can additional evidence be adduced in support (or refutation) of these two subgroups?

Despite these unknowns, by stages we are improving our understanding of historical relationships among Sulawesi languages. Instead of the ten subgroups listed at the beginning of this paper, following the results of this study we need recognise only five genetic groupings across the island of Sulawesi (though the position of Totoli and Boano, tentatively included as Celebic languages, remains suspect):

Sangiric
Minahasan
Gorontalo-Mongondow
Celebic
South Sulawesi

Of these five groups, Gorontalo-Mongondow could be considered exo-Sulawesian in that these languages originated at a later date from the Philippines, Proto Gorontalo-Mongondow speakers supposedly arriving in northern Sulawesi around 500BC (Blust 1991:103–104). Whether the other four groups are truly indigenous languages of Sulawesi — directly descended from their Proto Malayo-Polynesian ancestor — or whether they too share higher-level genetic relationships to languages outside of Sulawesi, remains to be seen. Sangiric and Minahasan languages have long been considered 'Philippine' languages on typological grounds, but proving a genetic connection to particular (or all) Philippine languages has proved more elusive. Initial evidence for their inclusion in a larger Philippines group has been accumulated by Zorc (1986),

but consists of lexical innovations only. Among his ninety-eight Proto Philippine lexical innovations, sixteen have reflexes in Proto Minahasan or a Minahasan language, while fifteen have reflexes in Proto Sangiric or a Sangiric language. This evidence has not been critically evaluated.

A link between South Sulawesi languages and languages of central and south-eastern Sulawesi has been disparaged. Adriani commented, 'Van het Boegineesch onderscheiden zich de Toradjasche talen in vele opzichten' (The Torajan [Kaili-Pamona] languages distinguish themselves from Buginese in many respects), then proceeded to marshal two pages worth of evidence (Adriani & Kruyt 1914:91–93). In similar fashion Mills noted that South Sulawesi languages differed from languages of central and south-eastern Sulawesi in respect to both historical sound change and verb morphology, and was inclined to note instead, with few specifics, affinities between South Sulawesi languages, Malay, and Madurese (Mills 1975:499, 517–519, 1981:60). Ross, however, has suggested that South Sulawesi may not be as distinct from Celebic languages as is often supposed, at least typologically. He partially reconstructs a 'Proto Sulawesi' system of verb morphology, from which South Sulawesi systems could also be derived (Ross 2002:462–464).

Finally, it is interesting to note how much the view of macrogrouping of Sulawesi languages presented above comes back around to a view presented ninety years ago by the pioneer of Sulawesi linguistic studies, the missionary Nicolas Adriani.

> Door zijne verwantschap met het Bobongkosch (op de Togian-eilanden) en het Gorontaleesch, wijst het Loinansch op eene strooming der bevolking van de N. helft van het Noordelijk schiereiland naar het Z. toe, die voorbij het gebeid van het Gorontaleesch zich heeft verdeeld, waarbij de Loinansche tak over de Togian-eilanden naar den vasten wal ten O. van Tandj. Api is gegaan, om zich daarop naar het O. (Balantaksch) en naar het Z. (Boengkoesch) te verbrieden, terwijl een andere tak naar het W., daarop naar het Z. is gegaan, en ten Z. van den Evenaar weder een tak in O.lijke richting heeft afgescheiden. Zoo mag men dus aannemen dat het Bare'e, als meest O.lijke uitlooper van dezen laatstgenoemden zijstroom, bij Tandj. Api weder op het Loinansch is gesluit.
>
> [By its relationship with Bobongko (on the Togian Islands) and Gorontalo, the Loinan language points to a southward migration of the inhabitants of the north half of the northern peninsula, which divided near present-day Gorontalo: the Loinan [Saluan] branch proceeded via the Togian islands to the further shore east of Tanjung Api, spreading therefrom to the east (Balantak) and to the south (Bungku), while another branch proceded to the west and then to the south, and then south of the equator a further branch separated back in an easterly direction. So one may consider that Bare'e [Pamona], as the most eastern extension of this last-named side flow, was arrested up against Loinan.] (Adriani & Kruyt 1914:89)

Given that Adriani developed his picture of migration at a time when typological concerns played an equal role with sound change in determining language relationships, it is remarkable how little needs to be changed. A clear amendment to Adriani's hypothesis is that we can no longer maintain that Celebic languages are descended from Gorontalic languages. Adriani's Loinan branch, however, has clear parallels to the Eastern Celebic branch proposed in this study, though here we must add that this 'southward flow' did not stop at present-day Bungku, but eventually encompassed all of peninsular south-eastern Sulawesi, including Muna, Buton, and the Tukang Besi Islands. Whether the Tomini-Tolitoli and Kaili-Pamona (and the now recognised Wotu-Wolio) languages together constitute a second, separate branch of Celebic, as

Adriani also supposed, remains to be seen. Once that issue has been settled, we can begin (again) to theorise about a Proto Celebic homeland.

Considering that Adriani had far less — and less reliable — data than were available to this author, the present work stands as a tribute to his genius and prescience.

References

Adelaar, K. Alexander, 1994, The classification of the Tamanic languages. In Tom Dutton and Darrell T. Tryon, eds *Language contact and change in the Austronesian world* 1–42. Trends in Linguistics, Studies and Monographs 77. Berlin: Mouton de Gruyter.

Adriani, N., 1928, *Bare'e-Nederlandsch woordenboek met Nederlandsch-Bare'e register.* Leiden: E.J. Brill.

Adriani, N. and A.C. Kruyt, 1914, *De Bare'e-sprekende Toradja's van Midden Celebes*, vol. 3: *Taal- en letterkundige schets der Bare'e-taal en overzicht van het taalgebied Celebes–Zuid-Halmahera.* Batavia: Landsdrukkerij.

Anceaux, J.C., 1987, *Wolio dictionary (Wolio–English–Indonesian).* Dordrecht: Foris.

Berg, René van den, 1991a, Muna dialects and Munic languages: towards a reconstruction. In Ray Harlow, ed. *VICAL* 2: *Western Austronesian and contact languages: papers from the Fifth International Conference on Austronesian Linguistics*, 21–51. Auckland: Linguistic Society of New Zealand.

1991b, Muna historical phonology. In James N. Sneddon, ed. *Studies in Sulawesi linguistics,* part 2, 1–28. NUSA Linguistic Studies of Indonesian and Other Languages in Indonesia, vol. 33. Jakarta: Universitas Katolik Indonesia Atma Jaya.

1991c, Preliminary notes on the Cia-Cia language (South Buton). In Harry A. Poeze and Pim Schoorl, eds *Excursies in Celebes: een bundel bijdragen bij het afscheid van J. Noorduyn als directeur-secretaris van het Koninklijk Instituut voor Taal-, Land- en Volkenkunde* Verhandelingen van het KITLV 147, 305–324. Leiden: KITLV Press.

1996a, *Muna–English dictionary.* Leiden: KITLV Press.

1996b, The demise of focus and the spread of conjugated verbs in Sulawesi. In H. Steinhauer, ed. *Papers in Austronesian linguistics,* No. 3, 89–114. Canberra: Pacific Linguistics.

Blust, Robert, 1990, Summary report: linguistic change and reconstruction methodology in the Austronesian language family. In Philip Baldi, ed. *Linguistic change and reconstruction methodology*, 133–153. Trends in Linguistics, Studies and Monographs 45. Berlin: Mouton de Gruyter.

1991, The Greater Central Philippines hypothesis. *Oceanic Linguistics* 30:73–129.

Dahl, Otto Christian, 1981, *Early phonetic and phonemic changes in Austronesian.* Oslo: Universitetsforlaget.

Donohue, Mark, 1999, *A grammar of Tukang Besi.* Mouton Grammar Library 20. Berlin: Mouton de Gruyter.

2000, Tukang Besi dialectology. In Charles E. Grimes, ed. *Spices from the east: papers in languages of eastern Indonesia*, 55–72. Canberra: Pacific Linguistics.

in press, The pretenders to Muna Buton. In John Bowden and Nikolaus Himmelmann, eds *Papers in Austronesian dialectology and subgrouping.* Canberra: Pacific Linguistics.

Esser, S.J., 1927, *Klank- en vormleer van het Morisch*, part 1. Verhandelingen van het Koninklijk Bataviaasch Genootschap van Kunsten en Wetenschappen, vol. 67/3. Leiden: Vros.

1933, *Klank- en vormleer van het Morisch*, part 2. Verhandelingen van het Koninklijk Bataviaasch Genootschap van Kunsten en Wetenschappen, vol. 67/4. Bandoeng: Nix.

Ferguson, Charles A., 1963, Some assumptions about nasals: a sample study in phonological universals. In Joseph H. Greenberg, ed. *Universals of language*, 2nd edn, 53–60. Cambridge: MIT Press.

Himmelmann, Nikolaus, 2001, *Sourcebook on Tomini-Tolitoli languages: general information and word lists.* Canberra: Pacific Linguistics.

Laidig, Wyn D. and Sahabu Dg. Maingak, 1999, Barang-barang phonology: a preliminary description. In Wyn D. Laidig, ed. *Studies in Sulawesi linguistics*, part 7, vol. 46, 46–83. NUSA Linguistic Studies of Indonesian and Other Languages in Indonesia. Jakarta: Universitas Katolik Indonesia Atma Jaya.

Laskowske, Thomas, 1995, A grammatical sketch of Seko Padang. Unpublished typescript.

Martens, Michael, 1988, Notes on Uma verbs. In H. Steinhauer, ed. *Papers in Western Austronesian linguistics*, No.4, 167–237. Canberra: Pacific Linguistics.

1989a, The Badaic languages. In Sneddon, ed. 1989:19–53.

1989b, Proto Kaili-Pamona: reconstruction of the protolanguage of a language subgroup in Sulawesi. Unpublished typescript, Hasanuddin University and SIL.

Mead, David, 1996, The evidence for final consonants in Proto–Bungku-Tolaki. *Oceanic Linguistics* 35:180–194.

1998, Proto–Bungku-Tolaki: reconstruction of its phonology and aspects of its morphosyntax. PhD dissertation, Rice University.

1999, *The Bungku-Tolaki languages of south-eastern Sulawesi, Indonesia*. Canberra: Pacific Linguistics.

2001, The numeral confix **i- -(e)n*. *Oceanic Linguistics* 40:167–176.

2002, Proto Celebic focus revisited. In Wouk and Ross, eds 2002:143–177.

Mills, Roger F., 1975, Proto South Sulawesi and Proto Austronesian Phonology. PhD dissertation, University of Michigan.

1981, Additional addenda. In Robert A. Blust, ed. *Historical linguistics in Indonesia*, part 1, vol. 10, 59–81. NUSA Linguistic Studies in Indonesian and Languages in Indonesia. Jakarta: Universitas Atma Jaya.

Ross, Malcolm D., 1992, The sound of Proto-Austronesian: an outsider's view of the Formosan evidence. *Oceanic Linguistics* 31:23–64.

2002, Final words: research themes in the history and typology of western Austronesian languages. In Wouk and Ross, eds 2002:451–474.

Sirk, U.K., 1981, The South Sulawesi group and neighboring languages. *Indonesia Circle* 25:29–36.

Sneddon, James N., 1978, *Proto-Minahasan: phonology, morphology and wordlist.* Canberra: Pacific Linguistics.

1984, *Proto-Sangiric and the Sangiric languages.* Canberra: Pacific Linguistics.

1989, The North Sulawesi microgroups: in search of higher level connections. In Sneddon, ed. 1989:83–107.

1993, The drift towards final open syllables in Sulawesi languages. *Oceanic Linguistics* 32:1–44.

Sneddon, James N., ed., 1989, *Studies in Sulawesi linguistics, part 1* NUSA Linguistic Studies of Indonesian and Other Languages in Indonesia, vol. 31. Jakarta: Universitas Katolik Indonesia Atma Jaya.

Usup, Hunggu Tadjuddin, 1986, Rekonstruksi proto-bahasa Gorontalo-Mongondow. PhD dissertation, Universitas Indonesia.

Wouk, Fay and Malcolm Ross, eds, 2002, *The history and typology of Western Austronesian voice systems.* Canberra: Pacific Linguistics.

Wolff, John U., 1988, The PAN consonant system. In Richard McGinn, ed. *Studies in Austronesian linguistics*, 125–147. Ohio University Monographs in International Studies, Southeast Asia Series 76. Athens, OH: Center for International Studies.

Zorc, R. David, 1986, The genetic relationships of Philippine languages. In Paul Geraghty, Lois Carrington and S.A. Wurm, eds *FOCAL II: papers from the Fourth International Conference on Austronesian Linguistics*, 147–173. Canberra: Pacific Linguistics.

Vowelless words in Selau

ROBERT BLUST

1 Vowelless languages versus vowelless words[1]

As Roman Jakobson once noted, there are no vowelless languages. It is now generally agreed that all known languages have at least two phonemic vowels, and nearly all have three or more (Maddieson 1984:126; Ladefoged & Maddieson 1996:286ff.).[2] Despite this limitation, a very small number of the world's roughly 6000 languages permit *vowelless words* — that is, free morphemes in which every underlying segment is a consonant. Such a structural feature has been reported for Kalam, a Papuan language spoken by about 15,000 persons in the Kaironk Valley on the northern edge of the Central Highlands of New Guinea. In Kalam, according to Pawley (1993:91), 'Many words contain no phonemic vowels. Consonant phonemes standing alone or before another consonant phoneme in a word are released with a predictable epenthetic or transitional vowel'. Among the examples Pawley gives are *wsn* [wusɨn] 'sleep and *ytk* [yirɨk] 'forest'. Languages like Kalam appear to be extremely rare, and none have been reported among the more than 1000 members of the Austronesian language family.[3] It thus comes as a surprise to discover that Selau, an

1 I am indebted to Peter C. Lincoln and John Lynch for comments which led to improvements in an earlier version of this paper. They bear no responsibility for my interpretations or conclusions.

2 Halle (1970:65) credits Jakobson with the observation that no vowelless language has ever been reliably reported, but I have not been able to locate the original statement. Kuipers (1960) claimed that the twelve short and five long vowels that occur phonetically in the North-west Caucasian language Kabardian can be reduced to two vowel phonemes, *a* and *E*. He then proceded to eliminate *a* and *E* by shifting the burden of contrast from these vowels to features of stress and juncture. As Halle demonstrated, Kuipers' analysis of Kabardian as a vowelless language appears to have been motivated more by the desire to discover a counteruniversal, even if only by notational contrivance, than by a well-supported analysis of the data. Halle proposed two underlying vowels *a* and *E* for Kabardian. Following up Allen's (1965) analysis of Abaza, another North-west Caucasian language, Anderson (1978) proposed a single vowel compromise for Kabardian, but two or more vowels have been recognised by most subsequent scholars, including Colarusso (1992), who admits two, and by Ladefoged and Maddieson (1996:286), who admit three, but curiously attribute Kuipers' position on this issue to Halle, and vice versa.

3 James A. Matisoff (pers. comm.) informs me that many Tibeto-Burman languages permit free morphemes that contain only consonants, and similar analyses have been proposed for Mandarin Chinese. John Lynch (pers. comm.) has also suggested that a few other Oceanic languages, including Big Nambas and Lenakel, may permit vowelless words. He notes in particular that Lenakel, spoken on the island of Tanna in southern Vanuatu, has a number of CVCVC words in which the vowels are a possibly predictable schwa.

John Lynch, ed. *Issues in Austronesian historical phonology*, 143–152.
Canberra: Pacific Linguistics, 2003.

Austronesian language spoken in western Melanesia, permits vowelless words. As an added benefit, comparative linguistic data show in a number of cases how these canonically bizarre morphemes have arisen from an antecedent stage which can be characterised as 'vowel-rich.'

2 Vowel-rich languages: Mon-Khmer versus Oceanic

Languages can be said to be 'vowel-rich' in either of two senses: (1) the proportion of vowels in the *phoneme inventory* is high, or (2) the proportion of vowels in *morphemes* is high. Many of the Mon-Khmer languages of mainland South-east Asia are vowel-rich in the first sense. In a relatively conservative analysis Huffman (1970:6ff.) describes Standard Cambodian as having 13 vowel phonemes, and 18 consonant phonemes. The vowels thus constitute some 42% of the total phoneme inventory. Other analyses of Standard Cambodian recognise a larger number of underlying vowels, and other Mon-Khmer languages, such as Bru, are said to carry this tendency for vowels to constitute an exceptionally large percentage of the phoneme inventory to even greater extremes.

The second sense in which languages can be called 'vowel-rich' is seen in many languages belonging to the Oceanic branch of the Austronesian language family. Hawaiian, for example, permits a number of non-reduplicated trisyllables that contain only vowels, as with *aea* 'rise up, raise the head', *aia* 'there, there it is', or *aua* 'look, observe', as well as quadrisyllables which are reduplicated (at least historically), as with *eaea* 'air, breath', or *uaua* 'tough, sinewy'. The five-syllable word *aoaoa* 'small shrub, *Wikstroemia* sp.', which contains only vowels, may or may not be a historical reduplication.

Since examples such as the foregoing are selective it is necessary to examine comparable data from languages belonging to different language families in order to determine cross-linguistic differences in the vowel : consonant ratio within morphemes. Table 1 presents the vowel : consonant ratio expressed as a percentage of vowels per morphemes from five languages: English, Malay, Roviana, Trukese and Samoan. To obtain these figures a variant of the Swadesh 200-item lexicostatistical test-list was completed for each language. The total number of vowels and consonants was then counted for each of the first 50 words on this list, and these totals were added for all 50 entries and divided by 50 to yield mean vowel percentages per morpheme. Diphthongs were counted as single vowels in all languages, English syllabic *r* was counted as a vowel, and long vowels and consonants in Trukese and Samoan were counted as single segments.

Table 1: Percentage of vowels per morpheme in English and four Austronesian languages

Language	No. of vowels	No. of phonemes	Percentage of vowels
English	56	171	33
Malay	121	289	42
Roviana	155	286	54
Trukese	78	182	43
Samoan	138	239	58

To date, however, he has proposed no such analysis, nor given any indication as to how it might be justified.

Although Malay, Roviana, Trukese and Samoan are all Austronesian languages, the last three belong to the Oceanic (Oc) subgroup of Austronesian, while Malay does not. Roviana is typical of many Austronesian languages in the western Solomons in having added echo vowels to Proto Oceanic (POc) CVCVC stems. Trukese is typical of many of the Oceanic languages of Micronesia and some parts of Melanesia, in having lost the last vowel and consonant of POc CVCVC stems, and Samoan is typical of most other Oc languages in having lost only the final consonant of POc CVCVC stems. As can be seen, the percentage of vowels per morpheme in all four Austronesian languages is significantly higher than in English, reflecting the fact that initial and final consonant clusters are common in English, but rare or absent in An languages. The percentage of vowels per morpheme in Trukese, which has lost POc final -VC sequences, is similar to that in Malay, which preserves most Proto Malayo-Polynesian final consonants, while that of Roviana, which has added echo vowels, and Samoan, which has lost final consonants, is significantly higher. Most of the languages of the western Solomons chain, like Roviana, have added echo vowels to POc CVCVC stems, and so typically have morphemes in which approximately half of the segments are vowels. In this sense, in comparison with languages such as English, they can be called 'vowel-rich' languages. Whatever the analysis of Selau shows us today, then, it appears likely that pre-Selau was a 'vowel-rich' language in terms of the balance of vowels and consonants within a morpheme.

3 Selau phonology: a first approximation

Selau is spoken at the northern tip of Bougainville island in the Solomons chain, within the political domain of Papua New Guinea. It is one of a number of dialects of the Halia language, most of which are spoken on the adjacent smaller island of Buka. Allen and Hurd (1965) list four varieties of Halia (also known as Hanahan, Tulon and Tasi). These are Halia proper, with 3833 speakers in Bougainville District at that time (but larger numbers on Buka), Haku, with 2951 speakers, Hangan, with 1562 speakers, and Selau, with 1540 speakers. The first three varieties are described as 'dialects' of a single language, but Selau is distinguished as a 'sub-language'. Lincoln (1976b) lists 'Halia (Selau variety)' as one of the languages spoken in Bougainville Province of Papua New Guinea. Allen and Allen (1987) state that Halia is 'spoken or understood by some 16,000 people on the north and east coasts of Buka island, the Selau peninsula on northern Bougainville, and the Carteret (Tuloun) atoll east of Buka, all areas included within the North Solomons Province of Papua New Guinea'. They list five dialects: Hanahan (east Buka), Tuloun, Hangan (south Buka), Hakö (north Buka), and Selau. Selau is again classified as a 'sub-language', meaning that it borders on being a separate language from Halia. In the most complete and detailed classification to data Ross (1988:217) assigns Halia to the Nehan/North Bougainville Network within his North-West Solomonic chain, with a further breakdown as follows:

Proto-Buka
Petats
Halia (including Haku, Hanahan, Kilinailau, Selau)[4]

---

[4] The names Tuloun, Kilinailau and the Carteret islands evidently apply to one and the same atoll. Allen and Allen (1987) use the alternative names Carteret and Tuloun, while Lincoln (1976b) identifies the Carterets with Kilinailau.

In one of the earliest descriptions of the area the anthropologist Beatrice Blackwood (1935:15ff.) classified Petats, spoken on the coral islets of Pororan and Hitau off the west coast of Buka, Saposa, spoken on a cluster of small islands off the north-west corner of Bougainville, and the language spoken on the east and west coasts of Buka as dialects of a single language, stating that 'All these dialects are mutually intelligible'. She noted that the 'dialect' of Kurtatchi, spoken in the villages of Kurtatchi, Baniu and Timputs on the north coast of Bougainville, was lexically quite different from that spoken on Buka, but that her Petats servant was able to pick it up 'in the course of a short time.' She never mentions the names Halia or Selau, and on the whole appears to greatly underestimate the degree of linguistic difference which actually distinguishes some of these communities.

Selau has 13 (possibly 14) surface consonants and six surface vowels, as shown in Table 2:[5]

Table 2: Surface consonants and vowels of Selau

Consonants				Vowels		
p	*t*	*c*	*k*	*i*		*u*
b	*(d)*		*g*	*e*	*ə*	*o*
m	*n*		*ŋ*		*a*	
	s					
	l, r					
w						

The phonemes *p, t, k* are unaspirated bilabial, alveolar and velar stops. *c* is an alveolar affricate. Although *c* derives historically from the affricatisation of **t* before a high front vowel, it now contrasts with *t*:

təna	'taro sucker'	*cəna*	'mother'
tor	'belly'	*con*	'man; husband'
tunus	'urine'	*cunu*	'bone'
ptaa	'earth'	*pcəpcə*	'large growth stage of squid'
wat	'stone'	*wac*	'four'

The phonemes *b, d* (the latter is rare in my data) and *g* are simple (non-prenasalised) voiced stops, and *r* is a 3–4 tap trill. The contrast of *l* and *r* can be illustrated with:

lei	'far, distant'	*rei*	'tree'
loŋo	'housefly'	*roŋono*	'to hear'
mola	'fat, grease'	*kori*	'possum'
tual	'eight'	*wiar*	'two'

The phoneme *y* was recorded in *yar* 'a shore tree: *Casuarina equisetifolia*' (Tok Pisin: *yar*), but this appears to be confined to loan words.

---

5 Selau material was collected in Canberra, Australia during the second half of 1975 in three sessions totalling approximately six hours. My primary source of data was Luke Pawen, with further input from Margaret Pinil during the last meeting. Both speakers were students temporarily resident in Australia at the time we worked together, and both hailed from the village of Torte on the northern tip of Bougainville.

Both vowels and consonants occur long. To avoid a premature introduction of the analysis which will follow these are cited in phonetic transcription:

[wakac]	'think, mull over'	[wa:kac]	'lightning'
[ara:]	'all'	[ara]	'fishing with line and hook'
[koto]	'bite'	[ko:to]	'kind of oval nut'
[siya]	'nine'	[s:ə]	'breast'
[anus]	'charcoal'	[un:awa]	'turtle shell'

Unlike most Oceanic languages, Selau permits consonant clusters in both initial and intervocalic positions in phonetic forms: [cta] 'putty nut', [ptən] 'coconut husk', [balsə] 'dove', [ləmtə] 'moss, algae'. Final clusters never appear on the surface but, as will be seen, do appear in underlying forms. Words may end either in an underlying vowel or an underlying consonant, and stress generally falls on the penult, although both final stress in disyllables and initial stress in trisyllables was recorded in some forms. Attempts to determine minimal or subminimal pairs, however, were invariably met with frustration, suggesting that stress is non-phonemic, but movable under still unstatable conditions.

The most interesting issues in Selau phonology appear in connection with the vowels. Phonetically Selau has straightforward reflexes of POc **a*, **e*, **i*, **o* and **u*. In addition, a number of words contain a mid-central or high-central vowel that was most frequently transcribed as schwa. All vowels can optionally be followed by a non-contrastive glottal stop in word-final position:

/mia/	[miya] ~ [miya?]	'tongue'
/sopene/	[sopene] ~ [sopene?]	'saucepan'
/tebeli/	[tebeli] ~ [tebeli?]	'clay pot'
/malto/	[malto] ~ [malto?]	'ten'
/cku/	[cku] ~ [cku?]	'sew sago leaves'

Allen and Allen (1987) list three front vowels, two central vowels, five back vowels and three glides for Halia. However, they provide no evidence of contrast, and it is almost certain that the number of underlying vowels is smaller than they indicate. What is most relevant to the problem at hand is that Halia lacks a mid-central or high-central vowel comparable to Selau /ə/. Rather, where Selau has a schwa the corresponding segment in Halia usually is a high vowel, or less commonly the low vowel /a/. I know of no instances in which Selau schwa corresponds to Halia /e/ or /o/.

4 The phonemic status of Selau schwa

Both synchronically and diachronically the most problematic segment in Selau is the mid-central vowel. Two types of synchronic evidence suggest that the schwa in many morphemes is non-phonemic, and this conclusion is further supported by historical considerations. One type of synchronic evidence will be examined with the diachronic evidence here, while the second type of synchronic evidence will be reserved for discussion in the following section.

The first indication that Selau schwa may not always be phonemic comes from the movable position of this segment in some transcriptions. The word for 'hand/arm' (POc **lima*) was recorded both as [ləma] and as [əlma], and the word for 'nit, egg of a louse' (POc

**lisa*) was recorded both as [ləsa] and as [əlsa]. Such free variation suggests that the schwa in these forms is little more than an automatic facilitating vowel which enables speakers of the language to pronounce the underlying consonant clusters in *lma* and *lsa* respectively.

Although the evidence of variant pronunciations is confined in the data recorded to these two forms, comparative-historical evidence points in the same general direction. As seen in Table 3, Selau has generally lost POc high vowels in final position.

Table 3: Loss of earlier high vowels in word-final position in Selau

POc	Selau	English
**boŋi*	*boŋ*	'night'
**keju*	*er*	'back of the head'
**-gu*	*-g*	'1SG genitive'
**-mu*	*-m*	'2SG genitive'
**maqati*	*mac*	'low tide, dry reef'
**mali*	*mal*	'bitter'
**mami*	*mam*	'our (excl.)'
**muri*	*mər*	'back, behind'
**nasu*	*nas*	'cook, boil'
**kaŋaRi*	*ŋar*	'canarium nut'
**rani*	*ran*	'day'
**tuRu*	*tur*	'housepost'
**pati*	*wac*	'four'
**patu*	*wat*	'stone'
**pitu*	*wit(i)*	'seven'
**poli*	*wol*	'buy'
**kutu*	*wut*	'louse'

Vowels generally did not drop in final position if they were not high:

**tina*	>	*cəna*	'mother'
**lima*	>	*ləma*	'five'
**tupa*	>	*tua*	'derris root'
**muno-muno*	>	*mənməno*	'caterpillar'

Although some high vowels remained word-finally, the general pattern of apocope for final high vowels is clear. Like many other languages in the western Solomons chain Selau added echo vowels to POc CVCVC forms. These vowels have been preserved if they are non-high, as with:

**ikan*	>	*iana*	'fish'
**limas*	>	*nəmsa*	'canoe bailer'
**onom*	>	*nomo*	'six'

but have disappeared if they were high, as with:

**qatun*	>	*atən*	'tuna, bonito'
**ranum*	>	*ramən* (metathesis)	'fresh water'
**saRum*	>	*saləm*	'needle'

As each of these forms shows, the high vowel that preceded a final consonant is reflected phonetically as a schwa, not as zero. Schwa and zero thus appear to be complementarily distributed reflexes of earlier high vowels, and in this sense can be regarded as structurally equivalent. Stated differently, POc high vowels underwent a conditioned sound change, disappearing in final position in CVCV forms and in certain other environments, but becoming schwa between consonants (or following derived consonant clusters). Etymologies such as **ikan* > *iana* 'fish', **tupa* > *tua* 'derris root fish poison', or **kutu* > *wut* 'louse' suggest that the centralisation and loss of high vowels between consonants took place only after the loss of **p* and **k*. Reflexes which contain a consonant cluster further support the analysis of schwa as structurally equivalent to zero. POc **putun* > *ptən* 'coconut husk', for example, shows a zero reflex of **u* in the first syllable, but a schwa in the second.

One might argue that high vowels underwent a conditioned sound change, disappearing in the penult but weakening to a phonemic schwa in the last syllable before a POc final consonant. However, this analysis is contradicted by etymologies such as **tina* > *cəna* 'mother', **lima* > *ləma* 'five', or **lumut* > *ləmtə* 'moss, algae', where a POc high vowel weakened to schwa in the penult, and (in the case of *ləmtə*) disappeared before a POc final consonant. The last example provides still another piece of evidence that schwa and zero are historically equivalent in Selau. Etymologies such as **lumut* > *ləmtə* 'moss, algae', **baluj* > *balsə* 'dove' and **laŋit* 'sky' > *laŋcə* 'rain' contain a final schwa. Should this schwa be regarded as a reflex of the echo vowel which these forms earlier contained? If so, there is no explanation as to why a similar weakened reflex does not appear in **qatun* > *atən* 'tuna, bonito', **putun* > *ptən* 'coconut husk', **saRum* > *saləm* 'needle' and the like. Rather, the schwa in *ləmtə*, *balsə* and *laŋcə* appears to be little more than a transitional vowel permitting speakers to pronounce what would otherwise be a disallowed final consonant cluster.[6]

5 Simple imperative and ambulatory imperative

This brings us to the third type of evidence that Selau schwa is sometimes, if not always, added by rule to underlying representations which lack it. Selau has two imperative constructions. The first of these, which I will call the 'simple imperative', has the general form verb-*i* (object), and can be translated as 'VERB (it)'. The second, which I will call the 'ambulatory imperative', has the general form *na* verb-*ia* (object), and can be translated as 'go and VERB (it)'. Examples of each are given in Table 4:

---

6 Detailed conditioning of the phonetic form of reflexes remains to be worked out (e.g. given **baluj* > /balsə/ 'dove', why not a parallel development **qatun* > **/atnə/ 'tuna, bonito'?). Although it is beyond the scope of this paper to investigate this conditioning, it is likely that zero vs. schwa as the reflex of an earlier high vowel in Selau has been determined by universal constraints on the form of consonant clusters. In the great majority of recorded cases high vowels were lost between medial and final consonants if the first consonant was equal to or greater in sonority than the second; otherwise the high vowel weakened to schwa.

Table 4: Simple and ambulatory imperative, formed with the suffixes *-i* and *-ia*

Base	Simple imperative	Ambulatory imperative	Gloss
ase	*ase-i*	*na ase-ia*	'count'
atəŋ	*atəŋ-i*	*na atəŋ-ia*	'fight'
cku	*cku-i*	*na cku-ia*	'sew sago leaves'
nas	*nas-i*	*na nas-ia*	'cook, boil'
nu	*nu-i*	*na nu-ia*	'eat'
pəla	*pəla-i*	*na pəla-ia*	'gun; shoot'
piu	*piu-i*	—	'blow on'
sani	*sani-i*	*na sani-ia*	'dance'
sowo	—	*na sowo*	'sleep'
suru	*suru-i*	*na suru-ia*	'sip'
tapala	*tapala-i*	*na tapala-ia*	'slap'
wa	*wa-i*	*na wa-ia*	'drink'
wana	*wana-i*	*na wana-ia*	'shoot with bow'
wol	*wol-i*	*na wol-ia*	'buy'

The full expression of the ambulatory imperative involves some redundancy, since it requires both the verb *na* 'go (and do something)', and the suffix *-ia*, either of which is sufficient to distinguish the two imperative constructions by itself. Although the available evidence is fragmentary, it appears that either the verb 'to go' or the imperative suffix (but not both) may be omitted without ambiguity. It is possible that *na* VERB is used only with intransitive verbs or with transitive verbs that take non-expressed indefinite objects, and that *na* VERB-*ia* or VERB-*ia* is used with verbs that take a definite object, but more data would be needed to settle the matter. Example sentences which illustrate this variation include the following:

(1)	*ase-ia moni*	'go and count the money!' (suffix only)
(2)	*atəŋ-i*	'kill it!'
	(na) atəŋ-ia pum	'go and kill the pig!' (verb optional)
(3)	*nas-i*	'cook it!'
	na nas	'go and cook (something)!' (verb only)
	na nas-ia iana	'go and cook the fish!' (both verb and suffix)
(4)	*na sowo*	'go to sleep!' (verb only).

The examples in Table 4 show that verbs which end in *a, e, i* or *u* retain this vowel before the suffixes *-i* and *-ia*. By contrast, verbs which end in schwa show no stem-final vowel before a suffix, as seen in Table 5:

Table 5: Verbs that end in phonetic schwa and their suffixed imperative forms

Base form		Simple imperative	Ambulatory imperative
garsə	'wash (hands)'	*gars-i*	*na gars-ia*
kcə	'to tie'	*kc-i*	*na kc-ia*
ssə	'breast'	*ss-i*	*na ss-ia*
warsə	'drop, throw down'	*wars-i*	*na wars-ia*

Example sentences include:

(1)	*(na) gars-ia ləma-mli* go wash-IMP hand-your	'go and wash your hands!'
(2)	*kc-i*	'tie it!'
(3)	*ss-i*	'nurse it! (e.g. a crying baby)'
	na ss-ia aksə go-nurse-IMP child	'go and nurse the child!'
(4)	*wars-i*	'drop it! throw it down!'

Only limited data was collected for Selau, and it is therefore not surprising that the number of relevant examples is small. Nonetheless they are sufficient to show that the morpheme for 'tie' must be a vowelless stem *kc*, and the morpheme for 'breast' must be a vowelless stem *ss*. Given this analysis it follows that [ənnə] 'earthquake', from POc **nunu*, is also a vowelless stem *nn*. If the central vowels were regarded as underlying we would be forced to conclude that **u* was irregularly reflected as schwa in an open final syllable, and that that **u* had both centralised and metathesised in the first syllable. Similar problems occur in other forms, as Selau [ləmtə] 'moss, algae' (POc **lumut*), which must be analysed as underlying *lmt*. The question remains open as to whether *any* instances of schwa in Selau are phonemic, but regardless of the answer to this question it is clear that a number of vowelless free morphemes must be admitted as underlying forms.

Before concluding it is worth stressing again that the present analysis *is not* based on the mere fact that some free morphemes in Selau contain no phonetic vowels other than schwa. In any phonemic analysis it is possible to represent one of the segments by zero. The consonant that is most commonly subject to zero-representation is glottal stop, and the vowel that most frequently receives this treatment is schwa. In such an analysis it would be possible to posit vowelless words as an analytical artifice, but the results would be subject to the same types of criticisms that Halle raised against the analysis of Kabardian as a vowelless language. Whether Big Nambas and Lenakel or any other An language actually contains vowelless free morphemes, or simply free morphemes that contain no phonetic vowels other than schwa, remains an open question. The evidence for vowelless free morphemes in Selau goes beyond such artifices in at least three ways. First, the position of schwa is variable in some recorded morphemes, as *ləma/əlma* 'hand' or *ləsa/əlsa* 'nit, egg of a louse'. Second, the imperative forms of bases that end with a vowel other than schwa show no vowel deletion or vocalic contraction when suffixed with *-i* or *-ia*, but bases which end with a phonetic schwa show no final base vowel under suffixation, and in this respect they behave no differently from bases which end in a consonant. Third, if we recognise the schwa as phonemic the rules of diachronic correspondence become exceedingly complex and implausible in a number of etymologies, for which **nunu* > *ənnə* 'earthquake' can serve to illustrate. POc high vowels normally disappeared in final position, and there is no obvious source for the initial vowel unless we arbitrarily introduce a hypothesis of metathesis. If schwa is allowed as a reflex of POc **u* in these cases it thus becomes necessary to recognise two phonological irregularities in this form. Under the same type of analysis, **susu* > *ssə* 'breast' would show a parallel irregularity in retaining the final high vowel as schwa, and would also show a morphophonemic anomaly in deleting the final vowel under suffixation. Observations such as this make it abundantly clear that a number of generalisations about Selau phonology — both synchronic and diachronic — would be obscured if we were to avoid the recognition of vowelless free morphemes in the language.

Occasionally low vowels appear to have deleted and were replaced with non-phonemic transitional schwas as well, as with [ənta] = *nt* 'egg' (Halia *nata*). Generally, however, Selau schwa is a historically secondary phonetic development which followed the deletion of earlier high vowels. As noted earlier, the greatest remaining challenge is to state the phonetic rules which govern the insertion of schwa. More challenging still is to find a reason why Selau, apparently alone among the more than 1000 Austronesian languages, has evolved a canonical shape which permits vowelless words, a development which is all the more remarkable considering that its immediate antecedent was a language which contained an exceptionally large percentage of vowels per morpheme.

References

Allen, Jerry and Conrad Hurd, 1965, *Languages of the Bougainville District.* Port Moresby: Department of Information and Extension Services.

Allen, Jerry and Janice Allen, 1987, *Halia grammar.* Data papers on Papua New Guinea languages, vol. 32. Ukarumpa, Papua New Guinea: Summer Institute of Linguistics.

Allen, W.S., 1965, On one-vowel systems. *Lingua* 13:111–124.

Anderson, Stephen R., 1978, Syllables, segments, and the Northwest Caucasian languages. In Alan Bell and Joan B. Hooper, eds *Syllables and segments*, 47–58. New York: North Holland Publishing Co.

Blackwood, Beatrice, 1935, *Both sides of Buka Passage: an ethnographic study of social, sexual, and economic questions in the North-Western Solomon Islands.* Oxford: Clarendon Press.

Colarusso, John, 1992, *A grammar of the Kabardian language.* University of Calgary Press.

Halle, Morris, 1970, Is Kabardian a vowel-less language? *Foundations of Language* 6:95–103.

Huffman, Franklin E., 1970, *Cambodian system of writing and beginning reader, with drills and glossary.* Ithaca, New York: Cornell University, Southeast Asia Program.

Kuipers, Aert, 1960, *Phoneme and morpheme in Kabardian.* Janua Linguarum, Series Minor 8. Mouton: The Hague.

Ladefoged, Peter, and Ian Maddieson, 1996, *The sounds of the world's languages.* Oxford: Blackwell.

Lincoln, Peter C., 1976a, History of research in Austronesian languages: Bougainville Province. In S.A. Wurm, ed. 1976:197–222.

1976b, Austronesian languages: Bougainville Province. In S.A. Wurm, ed. 1976:419–439.

Maddieson, Ian, 1984, *Patterns of sounds.* Cambridge studies in speech science and communication. Cambridge, New York: Cambridge University Press.

Pawley, Andrew, 1993, A language which defies description by ordinary means. In William A. Foley, ed. *The role of theory in language description*, 87–129. Berlin and New York: Mouton de Gruyter.

Ross, Malcolm D., 1988, *Proto Oceanic and the Austronesian languages of western Melanesia.* Canberra: Pacific Linguistics.

Wurm, S.A., ed., 1976, *New Guinea area languages and language study,* vol. 2: *Austronesian languages.* Canberra: Pacific Linguistics.

Blust, R. "Vowelless words in Selau". In Lynch, J. editor, *Issues in Austronesian historical phonology.*
PL-550:143-152. Pacific Linguistics, The Australian National University, 2003. DOI:*10.15144/PL-550.143*

The bilabials in Proto Loyalties

JOHN LYNCH

1 Introduction

The Loyalties subgroup of the New Caledonian family consists of three languages: Iaai, spoken on Ouvéa; Drehu, spoken on Lifou; and Nengone, spoken on Maré (see map). The other subgroup consists of the twenty-five or so languages of the New Caledonian mainland. Nengone is highly unusual — though not unique (see §4 below) — among Oceanic languages in showing apical (in this case alveolar) reflexes of Proto Oceanic (POc) labials in some environments (Ozanne-Rivierre 1986:50). All three languages show fronting of various vowels following certain bilabial consonants. These facts can be explained by assuming that the relevant bilabials were palatalised, or had palatalised allophones, in Proto Loyalties, and that this palatalisation has had the effects briefly described above.[1] In later sections of this paper, I also look at the development of POc bilabials as apicolabials or dentals/alveolars in some languages of Vanuatu, and at the behaviour of bilabials in some Micronesian languages, and try to draw some comparisons.

2 Proto Oceanic vowels in Loyalties languages

The POc five-vowel system has undergone some interesting developments in the three Loyalties languages. Iaai has developed a ten-vowel system, Drehu has seven vowels, and only Nengone has a five-vowel system:

[1] Lexical data are from Ozanne-Rivierre (1984) for Iaai, Sam (1995) for Drehu, and Tryon and Dubois (1969, 1971) for Nengone; Ozanne-Rivierre (1986) provides relevant data on all three Loyalties languages. Françoise Ozanne-Rivierre, who is currently working on a detailed phonological history of the Loyalties languages, has been particularly helpful in sharing data and commenting on various hypotheses. I am also grateful to Henning Andersen, Byron Bender, Ross Clark, Terry Crowley, Robert Early, Ian Maddieson, Claire Moyse-Faurie, Hans Schmidt, Jan Tent and Hannah Vari-Bogiri for comments on an earlier draft of this paper. I am solely responsible for errors or misinterpretations.

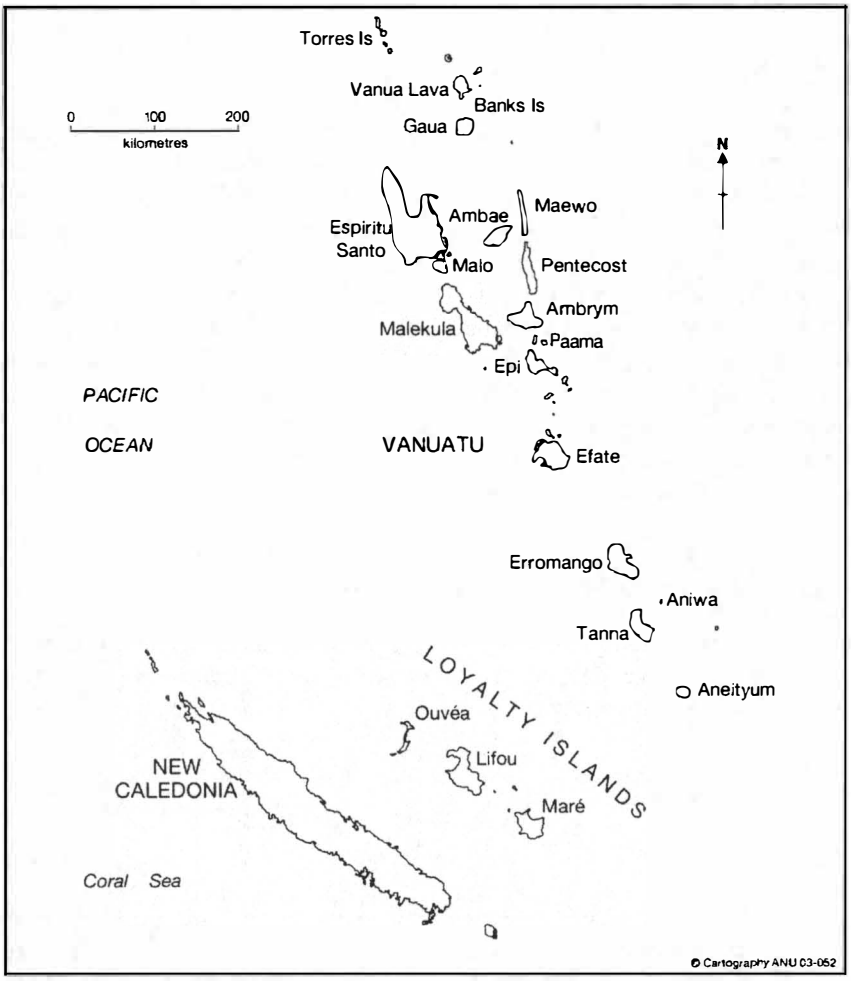
John Lynch, ed. *Issues in Austronesian historical phonology*. 153–173.
Canberra: Pacific Linguistics, 2003.

(1)

Iaai				**Drehu**			**Nengone**	
i	*ü*		*u*	*i*		*u*	*i*	*u*
e	*ø*	*ə*	*o*	*e*	*ə*	*o*	*e*	*o*
æ			*ɔ*	*ɛ*				
		a			*a*			*a*

In Iaai, all ten vowels occur after bilabials; all except *æ* and *ø* occur after non-labials; but only *i*, *e* and *ə* occur after the labiovelars (Ozanne-Rivierre 1976:81).[2]

In this section, I will concentrate on the development of POc **a*, **u* and **o*, especially after bilabials.



Map: Southern Oceania (Vanuatu and New Caledonia)

2 In this paper, I use the term *labiovelar* to refer to consonants like [pʷ] and [mʷ] except in direct quotes. These have sometimes been referred to as labialised bilabials, sometimes as velarised bilabials. For the Micronesian languages at least, Byron Bender (pers. comm.) is of the view that what is distinctive about the labiovelars is their velarisation. The term labiovelar avoids making a choice between these interpretations, and allows for cross-language comparisons.

2.1 PO **a*

The reflexes of POc **a* are difficult to characterise, since there appear to be a number of idiosyncratic changes in individual words. However, it seems clear that, while the default reflex was *a* in all three languages, the presence of a high vowel in the following syllable had different effects on **a* in Drehu and Nengone on the one hand and in Iaai on the other. In addition, **a* was fronted in all three languages when preceded by certain bilabials.

2.1.1 Drehu and Nengone

I begin with reflexes of **a* when it was preceded by a non-labial consonant. The Drehu and Nengone reflexes are generally identical, though occasional exceptions or variations occur (and are enclosed within square brackets in all examples in this paper). POc **a* seems to have become Drehu, Nengone *e* under two conditions: (a) when the following syllable contained a high vowel, and (b) when immediately followed by **ŋ* (and possibly also **g*, which was phonetically [ᵑg] in POc).

(2)		**POc**		**Drehu**	**Nengone**
	a.	**kani*	'eat'	*xen, əni*	*[kaan]*
		**kasupe*	'rat'	*[aði]*	*ɣeli*
		**kati*	'bite'	*heð*	
		**kawil*	'fishhook'	*ge* (?)	
		**natu-*	'child'	*nekə/n*	*tene* (metathesis?)
		**taŋis*	'cry'	*ʈeiθ*	(see also (b) below)
		**taci-*	'younger same-sex sibling'	*ʈeðin*	*cel*
		**tali*	'rope'	*wa/cen*	
		**tanum*	'bury'	*kelem*	
		**tapi*	'hit'	*xe(e)*	*cʰeie*
		**tasik*	'sea'	*keðe*	*cele*
	b.	**laŋo*	'a fly'	*neŋ*	*neŋo*
		**taŋa*	'basket'	*ʈeŋ*	*wa/ceŋ*
		**taŋis*	'cry'	*ʈeiθ*	(see also (a) above)
		**waga*	'canoe'	*weŋ*	*weg*

In other environments, POc **a* became *a* in both languages:

(3)	**POc**		**Drehu**	**Nengone**
	**draRaq*	'blood'	*ma/ɖa*	*ɖa*
	**kapak*	'wing, shoulder'	*ap*	*ade*
	**qata*	'mark'	*haʈe-*	*kʰace-*
	**qata*	'person'	*aʈ*	*si/ac*
	**qate-*	'liver'		*gɯ/at*
	**sapa*	'what?'		*laa*
	**tamʷaqane*	'male'	*ʈahmañ*	*cahman*
	**taqon*	'roast'	*san*	*cʰani*
	**wakaR*	'root'	*waa-ne*	

The following, however, show unexpected reflexes in both languages: *a* for expected *e* in (4a), and *e* for expected *a* in (4b):

(4)		**POc**		**Drehu**	**Nengone**
	a.	**qasu*	'smoke'	*hað*	*kʰali*
		**rani*	'day'	*lai*	*ran*
		**taliŋa-*	'ear'		*-canae*
	b.	**kape*	'k.o. crab'	*i/xe*	
		**(q)aca(n,ŋ)*	'name'	*eð*	*yele-*
		**salan*	'path'	*go/ðen*	*len*
		**tama-*	'father'	*kem*	*cen*

POc **a* was also raised and fronted to *e* in both of these languages when preceded by a *simple* bilabial (i.e. one which was neither aspirated nor a labiovelar), irrespective of the nature of the vowel in the following syllable. In Nengone, labials become apicals in this environment, as will be discussed in more detail in §3.

(5)	**POc**		**Drehu**	**Nengone**
	**madraR*	'ripe'	*hmeɖ*	*neɖe*
	**mamasa*	'dry'	*hmeð*	*hnede*
	**manaŋ*	'power'	*mene*	*nene*
	**maputa*	'sleep'	*meek*	*neic*
	**mata*	'eye'	*mek*	*wa-nec*
	**mata-*	'point'	*meke-*	*nece-*
	**mate*	'die'	*meci*	*neti*
	**pati*	'four'	*eke-*	*ece*
	**patu*	'stone'	*ete*	*ete*

This fronting and raising did not, however, occur after the POc labiovelars **bʷ* or **mʷ*, nor did it occur after Proto Loyalties (PLoy) **pʰ*, which derives from POc **pVp* sequences — see below.

(6)	**POc**		**Drehu**	**Nengone**
	**bʷatu*	'head'		*ba/nac* 'dried-up skull'
	**tamʷaqane*	'man'	*ʈahmaɲ*	*cahman*
	**Rumʷaq*	'house'	*uma*	*mma*
	**papa* (>PLoy **pʰa*)	'mouth'	*[hwɛ]*	*pʰa* 'opening'
	**papa* (>PLoy **pʰa*)	'carry'	*hwa* 'burden'	*pʰa/n* 'burden'

2.1.2 Iaai

The situation in Iaai is more complex. POc **a* is reflected in at least a handful of etyma as *a*, *ə*, *e*, *o* and *ɔ*.[3] Here I can only note tendencies, not generalisations.

---

3 The symbol *ə* represents a vowel slightly back of central, approximating [ɤ], while *a* is slightly more front after non-labials and more central after labials.

First, there is a tendency for **a* > *e* when the vowel in the following syllable was **i*; however, there are exceptions, listed on the right below.

(7)

POc *a/_Ci		**Iaai *e***	**POc *a/_Ci**		**Iaai *a***
**taŋis*	'cry'	*teŋe*	**kani*	'eat'	*an, han*
**tali*	'rope'	*te/keñ*	**ñami*	'taste'	*hnamen*
**taci-*	'younger same-sex sibling'	*kei/ñ*	**rani*	'day'	*laañ*
			**tapi*	'hit'	*xaii*
					Iaai *ə*
			**taliŋa-*	'ear'	*-kəñe*
			**tasik*	'sea'	*kəiə*

In two cases, the adjacency of **s* or **c*, whose non-initial reflex is *i*,[4] may cause the change of **a* to *e*: **kanase* 'mullet' > *enei* and **(q)aca(n,ŋ)* 'name' > *ie-*. (See also the reflex of **taci-* in (7), but contrast this with the reflex of **tasik*.) It is possible that this might also account for the development of **salan* 'path' as *den*, though this shows the regular initial (non-palatal) reflex *d*.

Similarly, there is a tendency for **a* > *ɔ* or *o* when the vowel in the following syllable was **u*, or when **a* was followed by **k* or **q* (both of which are lost in Iaai), but once again there are exceptions:

(8)

POc *a/_Cu		**Iaai *ɔ***	**POc *a/_Cu**		**Iaai *a***
**saRu*	'needle, sew'	*dɔ*	**qasu*	'smoke'	*hai*
**tanum*	'bury'	*kɔnɔm*	**tabu*	'taboo'	*kap*
		Iaai *o*			**Iaai *ə***
**natu-*	'child'	*nokV-*	**drapu*	'ashes'	*ɖəu*
**tapu(n)i*	'placenta'	*koii*			
POc *a / _k, _q		**Iaai *ɔ***			
**taqon*	'roast'	*θɔɔn*			
**wakaR*	'root'	*wɔɔ-*			
**kapak*	'wing, shoulder'	*abɔ* [second **a*]			
**tanaq*	'land'	*kɔnɔ* [second **a*]			

In other environments, the most regular reflex of **a* is Iaai *a*, but there are others which I cannot explain:

(9)

POc *a		**Iaai *a***	**POc *a**		**Iaai *ɔ***
**draRaq*	'blood'	*ɖa*	**tanaq*	'land'	*kɔnɔ* [first **a*]
**kapak*	'wing, shoulder'	*abɔ* [first **a*]			**Iaai *ə***
**qata*	'mark'	*hate*	**laŋo*	'a fly'	*nəŋ*
**qata*	'person'	*at*	**(ŋ)awaŋ*	'open'	*x/əəŋ*
**qate-*	'liver'	*aki-*	**sake*	'(go) up'	*dəə*
**taŋa*	'basket'	*taŋ*			**Iaai *o***
			**kape*	'k.o. crab'	*xop*
			**qatop*	'thatch'	*ot*
			**waga*	'canoe'	*ok*

---

4 For example **tasik* 'sea' > *kəiə*, **qasu* 'smoke' > *hai*.

POc **a* was regularly fronted when preceded by a simple bilabial. It appears that this fronting normally involved a change of **a* > *æ*, which apparently never occurs as a Iaai reflex of **a* except in this historical environment. (However, *æ* contrasts with both *e* and *a* in modern Iaai.)

(10)	**POc**		**Iaai**	
	**baga*	'banyan'	*bæk*	
	**bani*	'bait'	*o/bæ*	
	**baRa*	'wall'	*bææ*	'build a wall'
	**madraR*	'ripe'	*mæʈ*	
	**manaŋ*	'power'	*mæn*	
	**mañawa*	'breathe'	*mænɔ*	
	**mata*	'eye'	*ec-mæka-*	
	**paqal*	'thigh'	*je/ßææ-*	
	**pati*	'four'	*ßæk*	

In a couple of cases, we find *e* instead of *æ* as the reflex of *a here:

(11)	**POc**		**Iaai**
	**mamasa*	'dry'	*hme*
	**paRi*	'stingray'	*ße*
	**patu*	'stone'	*ßeto*

The **s* in **mamasa* changed regularly to *i*, and the **R* in **paRi* was lost very early;[5] it is probable that the resulting *æi* sequence resolved itself as the single vowel *e*. I have no explanation for the *e* reflex of **a* in **patu*; however, since *o* is not the regular reflex of **u* in this environment, there may be some other factor involved.

As in Drehu and Nengone, this fronting did not apparently occur after the labiovelars **b*^{w} or **m*^{w} or after PLoy **p*^{h}:

(12)	**POc**		**Iaai**	
	**b*^{w}*atu*	'head'	*ba-*	
	**Rum*^{w}*aq*	'house'	*uma*	
	**papa* (>PLoy **p*^{h}*a*)	'mouth'	*hwa/n-uma*	'door'

2.1.3 Discussion

The raising and fronting of **a* when the following syllable contained a high front vowel may have occurred in Proto Loyalties (with POc **a* > PLoy **e* in this environment), since all three languages show this development. However, the raising of *a when the following syllable contained a back vowel probably did not occur in Proto Loyalties, but must have been a later development, since **a* in this environment changed to *e* in Drehu and Nengone but to *o* or *ɔ* in Iaai.

In neither of these cases, however, is the Iaai reflex *æ*; this occurs *only* when the consonant preceding **a* was a simple bilabial, and the nature of the vowel in the following syllable is irrelevant. The data presented above suggest that POc **a* became PLoy **e* in this environment. It is possible that **e* had an allophone [æ] after bilabials, and that later

---

5 POc **R* appears to be lost in all etyma in all New Caledonian languages.

phonological changes in Iaai saw a split in **e*, such that /æ/ developed separate phonemic status. Further investigation of vowel correspondences is necessary to determine the developments in the PLoy vowel system, which in any case is of only marginal interest to the topic under discussion.

Now in modern Iaai, 'it is possible that all the "plain" labials in front or fronted vowel environments should be considered to be palatalized' (Maddieson & Anderson 1994:176), although the same is apparently not true of the bilabials in Drehu and Nengone (Claire Moyse-Faurie pers. comm.) where, however, no acoustic studies have been done. Nevertheless, if we assume that Proto Loyalties bilabials were so palatalised, then this provides an explanation for the fronting of **a* following a bilabial. If PLoy **ba*, for example, was phonetically [b^{y}a], then there is clear phonetic motivation for this to change to [b^{y}æ] or [b^{y}e], phonemically **be*. Presumably, **b^{w}* was not palatalised, and the vowel remained *a* after **b^{w}*. Subsequently, labiovelars lost their velarisation, with **b^{w}* becoming *b*; although this is now phonetically [b^{y}] in Iaai, no further fronting has taken place.[6] Thus the ordering of the relevant rules (with *B and *B^{w} symbolising the relevant consonant classes) is:

1. **Ba* > **Be*
2. **B^{w}a* > **Ba*

Note further, however, that PLoy **p^{h}* was probably *not* palatalised. While **pa* sequences became PLoy **pe*, **p^{h}a* sequences remained PLoy **p^{h}a*, as shown in (6) and (12) above. There are two possible explanations for this:

(a) Jan Tent (pers. comm.) suggests that, since aspiration involves a lot of air being expired, there must be a fairly free passage through the oral cavity; for that to happen, the tongue must make way for that airstream, and must therefore move away from the palate. Aspirated **p^{h}*, then, would not have been phonetically palatalised, or at least not to the same degree that unaspirated **p* was before **a*, though it must have been palatalised to at least some extent before front vowels — see (24)–(26).

(b) Another possible explanation is this. An examination of the data in (24)–(26) below shows that PLoy **p^{h}* has a labialised reflex *hw* in Iaai and Drehu when it occurs before **a* (as opposed to *h* before back vowels and *f* before front vowels). It is possible that, before **a*, PLoy **p^{h}* developed secondary labialisation (phonetically [p^{wh}]) which protected the following **a* from fronting; this secondary labialisation was retained in Iaai and Drehu, where the stop weakened, but not in Nengone.

2.2 POc **u*

As was the case with **a*, the reflexes of POc **u* are not simple to describe. I begin again with reflexes following non-labials.

The best general statement I can make is the following:

---

6 I am grateful to Ian Maddieson for suggesting palatalisation as a possible explanation for the labial-to-apical shift in Nengone — see §3 and §4 below. The behaviour of **a* – and, as we shall see, **u* — tends to confirm the view that the bilabials were palatalised in Proto Loyalties.

(13)	POc *u	Iaai	Drehu	Nengone
a.	after **k*	*u*	*ə*	*o*
b.	after **q*	*o (u?)*	*u? e? ɛ?*	*e*
c.	after other non-labials	*u*	*u*	*u*

Examples of each of these reflexes are given below:[7]

(14)	POc		Iaai	Drehu	Nengone
a.	**kuluR*	'breadfruit'	*o/un*	*ən*	*wa-on*
	**kulit*	'skin'	*un*		
	**kumi*	'squeeze'	*hum, [hom]*		
	**kurat*	'*Morinda* sp.'	*hulak*		
	**kuron*	'pot'	*[ət]*	*əl*	*ore*
	**kutu*	'cut'	*[xiiii]*	*xəð*	
	**kutu*	'louse'	*uto*	*ətɛ*	
b.	**quraŋ*	'lobster'	*oʈ*		
	**qunapi*	'scale'	*uneü*	*enienin*	*ehnie*
	**qusan*	'rain'			*el*
	**quta*	'burden'	*hook*		
	**qutan*	'inland'	*hoot*		*ete*
	**qutin*	'penis'		*ku*	*[u]*
	**qutok*	'brain'	*[haec]*	*xɛʈ*	
	**tuqur*	'stand'	*tooʈ*		
c.	**niuR*	'coconut'	*nu*	*nu*	*wa-nu*
	**Rum^{w}aq*	'house'	*uma*	*uma*	*[mma]*
	**suRi*	'bone'		*ðun*	*dun*
	**tuqaka-*	'older same-sex sibling'	*tuha*	*[tixe]*	
	**upi*	'blow'		*ufĩ*	*ut^{h}i*

There are a few cases of **u* > *i* after non-labials, but I will leave these for the moment.

After a voiced bilabial, POc *u is fronted as *i* (with one case of *e* in Iaai):

(15)	POc		Iaai	Drehu	Nengone
	**bune*	'pigeon'	*biñ*	*piñ*	*gu/din*
	**butoŋ*	'navel'	*bi/bikV-*	*pit*	*wa/didi*
	**mutaq*	'vomit'	*hmita*	*ɦmiʈa*	*hnija*
	**tubu-*	'grandparent'	*kibe*		
	**tubuq*	'grow'	*xibi*	*[cipa?]*	*[teda?]*

Fronting also occurs after **p*, which was subsequently lost in this environment. However, in this case the Iaai reflex is *ü*, while in Drehu and Nengone both *i* and *e* are found.[8] (The

---

7 While ungeminated **k* and **q* are regularly lost, the geminate equivalents are retained: these merge as Iaai *h* and Nengone *k^{h}* but are kept distinct in Drehu, where **kk* > *h* but **qq* > *x*. (There are other reflexes as a result of various morphophonemic processes.)

8 Two apparent exceptions are **pudi* 'banana' > Iaai *o/vic*, (Nengone *ura*?), and **drapu* 'ashes' > Iaai *ɖəu*. POc **pua* sequences have become PNC **pwa*; in the Loyalties languages, the **p* was lost regularly but the **w* was retained: e.g. **puaq-* 'fruit' > Iaai, Nengone *wa-*, Drehu *we/n*.

Drehu and Nengone reflex is probably *i*, with *e* developing later as a result of other factors.) Underlined forms in the list below suggest that the form reflects **b* rather than **p*, possibly as a result of initial prenasalisation (see footnote 12).

(16)	**POc**		**Iaai**	**Drehu**	**Nengone**
	**maputa*	'sleep'	*møøk* (via *meük*)	*meek*	*neic, hneic*
	**puŋao*	'parent-in-law'	*üŋo*		
	**pukot*	'net'	*üʈ-*[9]*, [eet]*	*eəʈ*	*eoc*
	**pulu*	'body hair'	*le/ün*	*pen*	*din*
	**pulu*	'rub, wash'	*üña*		
	**puni*	'hide'			*dene*
	**punuq*	'finished'	*o/biñ*	*[a/pɛn]*	*[bun]*
	**puqun*	'base'	*ü-*[10]		
	**puru*	'run'	*üʈ*		
	**topu*	'sugarcane'	*-küii*		

Once again, we might explain these reflexes by the palatalisation hypothesis, although we find variation within the reflexes. Although *modern* Iaai labials have palatalised allophones before unrounded vowels, this does not occur before rounded vowels. Nevertheless, I assume the following:

(a) **bu*, **mu* were phonetically [bʸu], [mʸu]. The palatalisation affected the following vowel, changing it to PLoy **i* (i.e. *i* in all three languages).

(b) **pu* was probably phonetically [βʸu] rather than [pʸu], as the data in §3 would suggest. It is not clear to me why **u* became Iaai *ü* rather than *i* in this environment. It is possible that the lips were more rounded and protruded in the production of the fricative than the stop or nasal, causing **i* to become front rounded.

In addition to these cases, there is a number of examples in which a **u preceding* a labial has been fronted, where in other environments one would expect the reflex *u* in all three languages:

(17)	**POc**		**Iaai**	**Drehu**	**Nengone**
	**kasupe*	'rat'		*aði*	*ɣeli*
	**tubu-*	'grandparent'	*kibe*		
	**tubuq*	'grow'	*xibi*	*[cipa?]*	*[teda?]*

It is possible that the phonetic palatalisation of the labial may also account for these changes.

2.3 POc **o*

As with other vowels, the reflexes of POc **o* show some variation, but the following general tendencies can be observed. POc **o* generally retains its rounding after **b*:

---

9 Possessive classifier used with nets.

10 Classificatory prefix used with trees.

(18)	**POc**		**Iaai**	**Drehu**	**Nengone**
	**boŋi*	'night'	*boŋ* 'day'		*bun* 'day'
	**boni*	'smell'	*bon*	*pun, pui*	*bo, bun*
	PEOc **boRe*	'dream'	*boo*		

In other environments, including after **p* (and **m*?), its reflexes are usually front or central:[11]

(19)	**POc**		**Iaai**	**Drehu**	**Nengone**
	**posi*	'squeeze'		*əð*	
	**monoŋ*	'stay'			*meneŋ*
	**qoda*	'raw'	*hat*	*haʈ*	*kʰaiʈe ?*
	**kona*	'bitter'	*hiñ*		
	**roŋoR*	'hear'	*leŋ, liŋ*	*ɖeŋ*	*ɖeɖeŋ*
	**kuron*	'pot'	*{əʈ}*	*{əl}*	*ore*

However, there are some cases where a rounded reflex occurs in at least one language (these are underlined in the data below):

(20)	**POc**		**Iaai**	**Drehu**	**Nengone**
	**lipon*	'tooth'	*ñu*	*ñə*	
	**mapo*	'cure, heal'		*meu*	*meu*
	**poñu*	'turtle'	*uñ*	*se/wen*	*ce/wen*
	**pukot*	'net'	*eet*	*eəʈ*	*eoc*
	**qatoluR*	'egg'	*wa/akuñ*		
	**qone*	'sand'	*ən*	*ŋ/əni*	*g/uni/n, k/uni/n*
	**tolu*	'three'	*kun*	*kəni-*	*ten*
	**tokon*	'crutch'		*tə*	*tu*
	**laŋo*	'a fly'	*{nəŋ}*	*{neŋ}*	*neŋo*

2.4 Summary

I summarise the above discussion in Table 1.

Table 1: Summary of vowel reflexes

	Vowel reflexes after the following PLoy consonants											
	b			***m***			***p***			***pʰ***	***bʷ***	***mʷ***
POc	**a*	**u*	**o*	**a*	**u*	**o*	**a*	**u*	**o*	**a*	**a*	**a*
PLoy	**e*	**i*	**(o, u)*	**e*	**i*	*?*	**e*	**i*	*?*	**a*	**a*	**a*
Iaai	*æ*	*i*	*o*	*æ*	*i*	*?*	*æ*	*ü*	*u*	*a*	*a*	*a*
Drehu	*e*	*i*	*u*	*e*	*i*	*?*	*e*	*e*	*ə, e*	*a*	*a*	*a*
Nengone	*e*	*i*	*o, u*	*e*	*i*	*e*	*e*	*e*	*e*	*a*	*a*	*a*

[11] Braces enclose cognate forms in which **o* has been lost. The form meaning 'stay' in (19) is usually reconstructed as **mono*. However, final **ŋ* is attested in the Nengone reflex *meneŋ*, in Bugotu *mono* 'to abide, stay, dwell, be', *monoŋi* 'to abide in', and possibly in Tolai *mono* 'to remain and take care of the home, boat etc.', *monoŋ* 'to afflict severely' (Meredith Osmond pers. comm.). Note also in this connection Proto Western Malayo-Polynesian **geneŋ* 'dwell, reside' (Blust in prep.) which resembles POc **monoŋ* except for the initial consonant.

3 The labial-to-apical shift in Nengone

I alluded above to the fact that some bilabials become apicals in Nengone in certain environments. I discuss that change in this section, and follow it in the next section with a comparison of a similar change in some languages of Santo and Malakula in Vanuatu.

Labiovelars consistently remain bilabials in Nengone, though they have lost their labialisation/velarisation; this loss presumably occurred after the labial-to-apical change.

(21)	**POc**		**Nengone**
	**b^{w}atu-*	'head'	*ba/nac* 'dried-up skull'
	**p^{w}anaq*	'bow, shoot'	*pehna*
	**Rum^{w}aq*	'house'	*mma*
	**tam^{w}aqane*	'man'	*cahman*

Before **o*, voiced bilabials remain bilabials:

(22)	**POc**		**Nengone**
	**boni*	'smell'	*bo(n), boni*
	**boŋi*	'night'	*bun* 'day'
	**monoŋ*	'stay'	*meneŋ*

In other environments, POc voiced bilabials are reflected as alveolars in Nengone. In the list below, I give a very rough Pre-Nengone intermediate form, incorporating the vowel changes discussed in §2. Note that in some cases the root appears to have been reduplicated in Pre-Nengone; this accounts for the aspirated stops (e.g. **kVkV- > k^{h}V-*) and also the voiceless nasals (**mVmV- > hm-, hn-*) — see below.[12]

(23)	**POc**		**Pre-Nengone**	**Nengone**
	**bune*	'pigeon'	***bine*	*gu/din*
	**butoŋ*	'navel'	***bi-bito*	*wa/didi*
	**kumi*	'squeeze'	***ku-kumi > **k^{h}umi*	*k^{h}on, k^{h}uni*
	**lima*	'hand'	***nime*	*nin*
	**maputa*	'sleep'	***me(p)ita*	*neic*
	**marama*	'light, shine'	***merame*	*neren*
	**mata*	'eye'	***meta*	*wa/nec*
	**mimiR*	'urinate'	***mimi > **hmi*	*hna* [**i > a* unexplained]
	**mutaq*	'vomit'	***mi-mita > **hmita*	*hnija*
	**ñamuk*	'mosquito'	***mina*	*nin* [metathesis]
	**pulu > *bulu*	'hair'	***binu*	*din*
	**puni > *buni*	'hide'	***bini*	*dene*
	**tama-*	'father'	***tame-*	*cen*
	**tubuq*	'grow'	***t(u,i)bi*	*teda* [**u > a* unexplained]

---

12 There are a couple of cases where **m* does not become *n* before **a*. With **mapo* 'heal(ed)' > *meu*, the Nengone form may be a loan from Drehu, which also has *meu*. With **maqurip* 'live' > *hmor*, the **aqu* sequence may have become *o* before the **ma > me* change began to operate; Iaai *m^{w}əəʈ* also suggests this (though Drehu *mel* does not).

In several words in Nengone (and also in Drehu), **p* is reflected as if it was **b* — see **pulu* and **puni* in (23) — but this was probably due to prenasalisation as a result of accretion of the article **na*, a widespread phenomenon in New Caledonian languages generally (see, for example Ozanne-Rivierre 1995).

It will be seen from the data above that **b* > Nengone *d* and **m* > *n* when the following vowel was a front vowel *in Proto Loyalties*, whether that front vowel continued unchanged a POc front vowel or whether it derived from POc **a* or **u* by the palatalisation process described in §2.

With the voiceless stop there are a number of complications. First, Proto New Caledonian (PNC) **p* is lost in all environments in Drehu and Nengone, and in most in Iaai. Second, Proto Loyalties inherited from PNC a simple/geminate distinction in voiceless stops (as well as in nasals): sequences of C_iVC_i, whether within a root (like POc **papa* 'mouth') or as a result of reduplication (see (23) above), produced a geminate stop, and these geminates in turn became aspirates in PLoy (see Haudricourt 1968; Rivierre 1978; Ozanne-Rivierre 1986, 1995). This did not occur with voiced stops.

I begin with reflexes of geminated POc **p*. Although there are not many POc roots reflected with geminate **p*, there are other comparisons which show the same correspondences. As one might expect from what has gone before, we have two sets of correspondences, depending on the nature of the following vowel. PNC **pp*, PLoy **p^h* was reflected in some etyma as Iaai and Drehu *hw* (*h* before *o* or *u*) and Nengone *p^h*:

(24)	**POc**		**Iaai**	**Drehu**	**Nengone**
	**papa*	'mouth'	*hwa/n-uma* 'door'	*hwɛ*	*p^ha* 'opening'
	**papa*	'carry on back'		*hwa* 'burden'	*p^ha/n* 'burden'
	**potak*	'crack open'		*huʈa, huʈi* 'break'	*p^huci* 'break'
		'injure'		*hwahwa*	*p^hap^{h}a*
		'clear bush'		*hweu*	*p^heu*
		'above'	*hon*	*hun, hon*	*p^hon*
		'press, push'		*huð, huði*	*p^huze, p^huli*
		'Lifou'		*ɖehu*	*ɖip^{h}u*

and in other etyma as Iaai *hv* (not exemplified here) or *f*, Drehu *f* and Nengone *t^h*:

(25)	**POc**		**Iaai**	**Drehu**	**Nengone**
	**upi*	'blow'		*uf, ufi*	*ut^{h}i*
		'also'		*fe*	*t^he*
		'to string'	*fiŋec*	*finiθ, finii*	*t^hini*
		'to stick to'		*feɖ*	*t^heɖe*
		'below'		*fen*	*t^hen*
		'dig up'		*feð, feke*	*t^hize, t^hice*

I have already noted that POc **a* does *not* become a front vowel after PLoy **p^h*. Thus PLoy **p^h* remained *p^h* in Nengone before central and back vowels, but underwent the shift **p^h* > *t^h* before front vowels — certainly before **i*, and probably also before **e*. Thus:

(26)	**PLoy**		**Iaai**	**Drehu**	**Nengone**
	**p^h*	/ __ *a*	*hw*	*hw*	*p^h*
	**p^h*	/ __ *o, u*	*h*	*h*	*p^h*
	**p^h*	/ __ *i, e*	*f? hv?*	*f*	*t^h*

Ungeminated **p* is lost in all environments in Drehu and Nengone, but conditioned some vowel changes before being lost, as we have already seen. In Iaai, **p* is lost before POc **u* (which became PLoy **i*, Iaai *ü* in this environment) and, in non-initial syllables, before POc **i*, as in (27).

(27)	**POc**		**Iaai**	**Drehu**	**Nengone**
	**maputa*	'sleep'	*møøk* (via *meük*)	*meek*	*neic, hneic*
	**puŋao*	'parent-in-law'	*üŋo*		
	**pukot*	'net'	*üʈ-*[13]*, eet*	*eəʈ*	*eoc*
	**pulu*	'body hair'	*le/ün*	*(pen)*	*(din)*
	**qunap-i*	'scale'	*uneü, eneü, hünüü*	*enienin*	*ehniè, enenien*
	**tapi*	'hit'	*xaü, xəü*	*xe, xee*	*cʰeie*
	**topu*	'sugarcane'	*-küü*		

POc **p* seems also to have been lost before **o* in all three languages, and in this case, fronting occurs in (some reflexes in) Drehu and Nengone but not in Iaai, where the reflex is *u*. (Note, however, that Drehu and Nengone seem to show *w* < **p* in reflexes of **poñu* 'turtle'.)

(28)	**POc**		**Iaai**	**Drehu**	**Nengone**
	**lipon*	'tooth'	*ñu*	*ñə*	
	**mapo*	'cure, heal'		*meu*	*meu*
	**poñu*	'turtle'	*uñ*	*se/wen*	*ce/we*
	**posi*	'squeeze'		*əð*	

Elsewhere, **p* is lost in Drehu and Nengone but is retained as non-final *ß*, final *p* in Iaai:

(29)	**POc**		**Iaai**	**Drehu**	**Nengone**
	**kape*	'k.o. crab'	*xop*	*ixoe, ixe*	
	**kasupe*	'rat'		*aði*	*ɣeli*
	paŋu* (<puŋa*)	'flower'	*ßəŋɔ/n*	*eŋe/n*	*eŋ*
	**paqus-i*	'tie together'		*oθ*	
	**paRi*	'stingray'	*ße*	*e*	
	**paRu*	'*H. tiliaceus*'	*ßææu*	*elu*	*eru*
	**pat*	'four'	*ßæk*	*eke/ʈe*	*ece*
	**patu*	'stone'	*ßeto*	*etɛ*	*ete*
	**patuR-i*	'weave'	*ßææk*	*ekɛ*	*et*
	**penako*	'steal'	*ßenɔu*	*ɛnə*	*eno*
	**pican*	'how many?'	*ße*	*iðe/ʈe*	*ha/el*

There is thus no apical reflex of ungeminated **p* in Nengone — indeed, no reflex of any kind if one excepts the *w* in the reflex of **poñu*. Following the pattern of the other bilabials, we would have expected **p* > Nengone *t* (or maybe *θ* or *s*) before PLoy front vowels. I can only assume that **p* was lost in this environment at a very early stage before the labial-to-apical shift began to take place, possibly because **p* was probably a fricative *ß* in PLoy, and thus more subject to lenition.[14]

The developments in Nengone are summarised in Table 2.

---

13 Possessive classifier used with nets.

14 It may just possibly be that PLoy **p* became some apical phoneme which was later lost, but this is far less likely.

Table 2: Development of Bilabials in Nengone

PLoy	*b^w	*p^w	*m^w	*b		*m		*p^h		*p
				/_*o	/_*i,(*e)	/_*o	/_*i,*e	/_*a,*o,*u	/_*i,(*e)	
Nengone	b	p	m	b	d	m	n	p^h	t^h	$\emptyset$

4 Comparisons

Languages in two other Oceanic-speaking areas show similar though not identical developments to those which have taken place in the Loyalties, and I briefly discuss here the labial-to-apical shift in some languages of Santo and Malakula in Vanuatu, and the palatal or palatalised nature of bilabials in some Micronesian languages.

4.1 Santo–Malakula

Tryon (1976:52–53), Clark (1985:205–206) and, in more detail, Maddieson (1989) have described the development of linguolabials in a few languages in a restricted area of northern Vanuatu.[15] Linguolabials are extremely rare in the world's languages: Maddieson (1989:350) says that they occur as regular linguistic units *only* in the Santo–Malakula area, although subphonemic or stylistically restricted linguolabials are found in two Brazilian languages and in one language in Tanzania.

In Santo–Malakula, linguolabials are found as phonemes in Mafea (spoken on a small island east of Santo), the Tangoa and moribund Araki dialects of South-west Santo, the now extinct Aore language spoken on Aore island off the south coast of Santo, Vao (off the coast of Malakula), and the Mpotovoro and V'ënen Taut (Big Nambas) languages of north Malakula. In a number of other languages in the immediate vicinity, POc bilabials are reflected as dentals/alveolars (probably via earlier linguolabials): Tolomako in north Santo, Shark Bay in north-east Santo, Mores (or Roria) in central Santo, the Tambotalo and Butmas-Tur dialects of South-east Santo, and the Vovo and Dirak (or Mae) languages in the north of Malakula.

Maddieson's conclusions are briefly summarised in (30) below, with Vao illustrating the changes. As far as the languages with linguolabials are concerned,

(30) a. simple bilabials before non-rounded vowels became linguolabials (e.g. POc **tama* 'father' > *taṃa-*, **bebe* 'butterfly' > *na-β̪e^{n}b̪*);

b. subsequently, labiovelars became simple bilabials (e.g. **(b^w, p^w)atu(k)* 'head' > *^{m}batu-*, **m^wata* 'snake' > *na-mat*);

c. simple bilabials before rounded vowels *generally* remained bilabials (**kabu* 'fire(wood)' > *na-ŋa^{m}b*, **damu* 'yam' > *na-ŋam*).

---

[15] Linguolabials — also referred to as apicolabials — are consonants 'produced by an articulation involving the tongue and the upper lip' (Maddieson 1989:349). In this paper (as in Maddieson 1989), they are written as labials with the dental diacritic – thus *m̪*, *b̪*, etc. Vanuatu language names and locations follow Lynch and Crowley (2001).

In some other adjacent languages, bilabials became apicals before non-rounded vowels (**tama* 'father' > Tolomako *tana-*, **bebe* 'butterfly' > Tolomako *tete*), probably via an intermediate linguolabial stage.

In discussing the labial to linguolabial to dental shift in Santo–Malakula, Clark (1985:205) says that

> this shift is phonetically so unusual, and the languages manifesting it so clearly concentrated in a single area, that we can hardly imagine it not to have had a single origin. Yet it cuts across six different local [genetic] groups, without including any of them entirely, and thus seems to contradict even our most plausible working assumptions about subgrouping. This is less disturbing, however, if we consider the nature of the change more closely.
>
> The apico-labial shift presumably begins with a shift from labial to apico-labial. This change has two important properties. First, it is reversible (no mergers are involved). Second, it creates a highly-marked type of articulation, very rare in human languages. This means that there will be a high likelihood of subsequent elimination of this series of consonants by further sound change. One possibility is to merge the apico-labials with other apicals (i.e. dentals). A second is to return them to ordinary labials, thus erasing all evidence of the shift. It is quite possible, then, that all the languages of [the six local genetic groups] originally underwent the apico-labial shift. A majority of languages subsequently reversed the shift and now show no evidence of it (at least from our limited data). The minority listed above either merged the apico-labials with the dentals or preserved the apico-labial articulation.

Clark (1985:205–206) also suggests that, since linguolabials 'are a highly salient but very superficial feature of language', one might expect that the change could be easily borrowed, and this might also account for their occurrence across genetic groupings.[16]

In discussing this same development, Maddieson (1989:360–361) says that, although the linguolabials 'are the result of a well-behaved historical change, ... we have no explanation to offer as to *why* the bilabial to linguolabial shift took place. A possible explanation for the exemption of labialised bilabials from participation can be proposed: the narrowing of the lip aperture associated with rounding may have inhibited the forward protrusion of the tongue'. A similar explanation, of course, might account for the fact that simple bilabials before back rounded vowels also did not undergo this change. In a recent personal communication, however, Maddieson has the following to say:

> I now think that the 'cause' of the bilabial to linguo-labial shift should be traced to a prior development of a fairly marked palatalization accompanying plain (non-labialized) bilabials before unrounded vowels in a fairly large subset of Oceanic (seen, for example, in Micronesian, especially Marshallese, and fairly apparent in Iaai among Loyalty languages). Palatalization entails a high and forward position of the tongue blade which if somewhat over-vigorously articulated can lead to the tip coming out between the lips and making contact with the upper lip as in a linguo-labial ...
>
> Palatalized bilabial stops are also prone to be perceived as coronals, since the palatalization gives them a high second formant locus, which is typical of dentals and alveolars and unlike the low second formant found with plain bilabials (and the even lower one found with labialized bilabials). These perceptual considerations suggest the

---

16 Clark bases this assumption in part upon the mistaken information (Camden 1979:113) that in Tangoa linguolabials are a feature of men's speech only — see Maddieson (1989:360) for a refutation of Camden's argument. Nevertheless, this does not affect Clark's view on their 'borrowability'. The linguolabials are actually being lost in Tangoa, though they are still retained in neighbouring and closely related Araki (Hannah Vari-Bogiri, pers. comm.)

> possibility of direct passage from (palatalized) bilabial to dental or alveolar without passing through a linguo-labial stage. However linguo-labials could also be adopted as the solution to a set of ambiguous place cues which seems to display both features of labiality (visual) and coronality (auditory) at the same time. (Ian Maddieson, pers. comm. 31/1/2001)

There is not a great deal of information available on the phonetics of these languages. In V'ënen Taut (Fox 1979:2), the bilabial phonemes *p* and *m* have *rounded* allophones before *front* vowels and plain (unpalatalised) allophones elsewhere. I am not sure, however, whether this is relevant to the issue under discussion.

However, there is an interesting situation a little further south. In (at least) two languages spoken on the island of Epi in central Vanuatu, bilabials (but not labiovelars) cause /a/ to have fronted allophones. In Lamen, /a/ has a fronted allophone [æ] following bilabial consonants (Early 2002:671). In Lewo (Early 1994:56), the same situation occurs, but in addition /a/ has the allophone [æ] preceding /v/; thus /mare/ [mære] 'die', /eveklavi/ [eβeklæβi] 'afternoon'.

4.2 Micronesia

Rehg (1991:397) draws attention to a phenomenon in a number of Micronesian languages where there are at least two sets of consonants, front or 'light' and back or 'heavy', which have different effects on vowels. In Ponapean, for example, front consonants are those which are either forward of dental or which involve palatalisation, while other consonants are back. The front consonants include /p m d l n s/, while the back consonants include /pʷ mʷ t r ŋ k/. In the environment of front consonants, back vowels are somewhat fronted (i.e. they have central allophones) and the central vowel *a* has a front allophone [æ] (Rehg 1981:43ff.).[17]

Marshallese is analysed as having the four vowel phonemes *i e ɛ a*. The allophones of these vowels 'are front next to unmarked (phonetically palatalised) consonants ... back (but unrounded) next to pharyngealised or velarised consonants ... and rounded next to rounded consonants' (Bender 1984:463). Among the bilabials, b^w and m^w are velarised and condition back unrounded allophones, while *p* and *m* are palatalised [pʸ] and [mʸ] and condition front allophones. For example, in words such as *yaj* 'weave' or *maniy* 'thin', where both consonants are front or light, the vowel /a/ is quite front — [yæt̪ʸ], [mʸæni]. On the other hand, words like $b^w ab^w b^w ib^w$ 'butterfly', with heavy consonants, have a back allophone — [bʷɒbʷ:ɯbʷ] (see Bender 1994 for further details).

It appears that these modern bilabials have the following origins (Bender & Wang 1985:68-69):

(31)	**POc**	**Marshallese**
	**b*	*p*
	**b*, **b^w*	b^w
	**m*	*m*
	**m*, **m^w*	m^w

17 The situation is more complex than this in Ponapean (and Marshallese), in that vowels have different (often diphthongised) allophones when the consonants on either side differ in lightness. This is not relevant to this discussion.

Bender and Wang (1985:68) also note that POc **p*, which became Proto Micronesian (PMc) **f*, has the reflex *y* in Marshallese before **a* and **i* (and probably also **e*, though there does not appear to be any evidence of this); before back rounded vowels, **p* > *w*:[18]

(32)	**POc *p**		**Marshallese *y***
	**patuR-i*	'weave'	*yaj*
	**pituqun*	'star'	*yijiw*
	**manipi*	'thin'	*maniy*
	**panua*	'land'	*yaney*
	**qapaRa*	'shoulder'	*hayɛr*

(33)	**POc *p**		**Marshallese *w***	
	**poñu*	'turtle'	*wen*	
	**topu*	'sugarcane'	*taw*	
	**mapo*	'heal'	*mɛw*	
	**pukot*	'net'	*wek*	
	**puaq*	'fruit'	*wiwah*	'bear many fruit/flowers'
	**pupu*	'fish trap'	*wiw*	

While **p* > *w* is a quite common sound change, **p* > *y* is not. Once again, though, this suggests that Proto Micronesian **f* (< POc **p*) was palatal, or palatalised, at least before unrounded vowels, and that **[fʸ] lenited to the corresponding glide *y*.

4.3 The Teták dialects of Czech

Andersen (1973) describes a similar change in the Teták dialects of Czech. Old Czech had a contrast between 'sharped' (i.e. palatalised) bilabials /pʸ bʸ mʸ/ and regular bilabials /p b m/.[19] In most Czech dialects, these have remained bilabials; some dialects preserve a sharp vs. plain distinction, others do not. In the Teták dialects, however, the sharped bilabials became dentals before /i ě ř/. For example (Andersen 1973:765):

(34)	**Standard Czech**	**Teták dialects**	
	koupiti	*koutit*	'buy'
	pěkně	*tekňe*	'nicely'
	bílý	*di:lej*	'white'
	běžeti	*dežet*	'run'
	míti	*ni:t*	'have'
	město	*nesto*	'town'

Andersen describes a two-way tonality distinction in Old Czech: (i) high or low basic tonality (i.e. [–grave] vs. [+grave]), and heightened vs. non-heightened tonality (i.e. [+sharped] vs. [–sharped]). This provides the following oppositions:

---

18 Marshallese lexical data are from Abo et al. (1976).

19 Andersen writes the sharped consonants with a following comma — thus *p,* — but this can be orthographically confusing, and I use a superscript *y* instead here.

(35)

	high tonality	low tonality
heightened ([+sharped])	t^y	p^y
non-heightened ([–sharped])	t	p

In most Czech dialects, the heightened tonality syntagm was reduced, heightened becoming non-heightened: this occurred twice — first in the high tonality consonants and then in the low). Thus *t* and t^y merged as *t*, then later *p* and p^y merged as *p*. The Teták dialects underwent the first change, but then merged p^y with *t*. Andersen (1973:771) refers to 'the ambiguous character of the acoustic manifestation of these [tonality] syntagms', with 'the manifestations of heightened tonality fall[ing] in an intermediate range' between [+grave] and [–grave]. While the majority of Czech dialects interpreted this as a manifestation of simple low tonality, the Teták dialects interpreted it as a manifestation of simple high tonality.

5 Discussion

A comparison of the Loyalties, Santo–Malakula and Micronesian situations shows some interesting similarities as well as some differences. In each case, we are dealing with bilabials which are, or were, palatal or palatalised, but the effects of this palatalisation vary somewhat.

1. In all three situations, labiovelars do not exhibit palatalisation, do not cause fronting of vowels, and do not undergo any labial-to-apical change.
2. In the Loyalties, the palatalisation of simple bilabials caused a phonemic change in at least two vowels, with a following POc **a* > PLoy **e* and POc **u* > PLoy **i*. In Nengone, simple bilabials then became dentals before front vowels, either directly (as in Teták Czech) or via an apicolabial stage (as in Santo–Malakula): both scenarios are equally possible, and there is no evidence to support one rather than the other. (The aspiration of PLoy *p^h prevented fronting of **a*, but did not prevent the labial-to-apical change before **i* and **e*.)
3. In Santo–Malakula, one assumes that simple bilabials must also have been palatalised. While no vowel fronting occurs here, the palatalisation has caused a labial-to-apical shift before unrounded vowels — probably bilabial > apicolabial, with the apicolabials then changing to dentals in some languages (and, as Clark suggests, possibly changing back to bilabials in others).
4. In Micronesia, vowels have front allophones in the environment of modern bilabials. Byron Bender (pers. comm.) feels that the light consonants (which include the simple bilabials) are the least marked, and the vowel quality in this context is the default quality: 'From this point of view, speakers aren't palatal*izing* so much as speaking palatally'. Historically, PMc **f* (< POc **p*) was probably also palatal or palatalised, and it has become *y* in Marshallese before unrounded vowels (though no similar development appears to have taken place in other Micronesian languages, at least that I am aware of).

In addition, the Czech situation provides a phonetic explanation for a direct change from palatalised bilabial to apical.

What connections are there between these developments, and what implications (if any) do they have for the reconstruction of Proto Oceanic and for the determination of genetic relationships within Oceanic?

The developments in the Loyalties and especially Nengone are almost certainly unconnected historically with those in north-central Vanuatu. No North-Central Vanuatu language appears to show fronting of **a* and **u* after bilabials, which occurs in all three Loyalties languages. No other New Caledonian language — including especially Nengone's closest relatives, Iaai and Drehu — show the labial-to-apical shift, which appears to have occurred *only* in Nengone. (It is possible that Iaai and Drehu underwent this shift and then reverted to bilabials; but in that case, one would expect **p* > *t* or *θ* in Nengone, which would not then be lost, since Iaai at least retains **p* in some environments.)

Nengone is no more closely connected genetically to the languages of Vanuatu than are the other languages of New Caledonia. Nor is there any evidence of the change being contact-induced: I am not aware of any specific contact between Santo–Malakula and the Loyalty Islands, and a glance at a map will show that there are numerous intervening languages which do not show any evidence of contact of this kind. If the Nengone labial-to-dental shift has a different origin from Santo–Malakula linguolabialisation, then we have the difficult task of trying to explain why a highly unusual change like the labial-to-apical shift occurred, quite independently, in just a few languages in the same general geographical area.

On the face of it, there seems to be even less chance of a close historical connection or of contact between the Loyalties and Micronesian languages. Most theories of Micronesian origins have them migrating from further north — the South-east Solomons and Northern Vanuatu are two areas which have been suggested (see Jackson 1986) — though in no case is the evidence particularly compelling. It is, I suppose, not impossible that the Micronesian languages originate from the Loyalties. Both groups of languages show loss of POc **p* before round vowels and unconditioned loss of **y* and (ungeminated) **q*.[20] They both also show consonant gemination, a plethora of possessive classifiers, and some other common features. If there *were* a close genetic connection between the Loyalties and Micronesia, then the palatalisation or palatal nature of the bilabials in both groups would have a single origin and a unified explanation. However, at this stage of research I am not suggesting that this was the case. I would suggest, however, that this is something that could well be further investigated, even if only to confirm that Micronesian languages did *not* originate in the Loyalties.

If neither common origin nor contact is the explanation, we are left with an interesting puzzle. Blust (1990, 1996, inter alia) has commented on the fact that a number of unusual phonological changes occur sporadically within the Austronesian family in languages or subgroups which have no specifically close genetic connection: low vowel dissimilation (e.g. **aCa* > *eCa*), velar stop voicing crossover, **t* > *k*, fronting of **u*, etc. It may be that palatalisation of bilabials (with subsequent apicalisation) represents one more of these. On the other hand, it is possible that the POc bilabials themselves were palatalised, or had palatalised allophones, with this feature subsequently being lost in most daughter languages. This possibility requires further research.

---

20 Micronesian innovations are from Jackson (1986:204). The Loyalties languages, however, retain PNC **qq*, which derives from POc **qVq*, with the reflexes Iaai, Drehu *h*, Nengone *k^h*.

References

Abo, Takaji, Byron W. Bender, Alfred Capelle and Tony DeBrum, 1976, *Marshallese-English dictionary*. Honolulu: University of Hawai'i Press.

Andersen, Henning, 1973, Abductive and deductive change. *Language* 49/4:765–793.

Bender, Byron W., 1984, Object marking in Marshallese. In Byron W. Bender, ed. *Studies in Micronesian linguistics*, 443–465. Canberra: Pacific Linguistics.

1994, Language reform in a Micronesian context. In István Fodor and Claude Hagège, eds *Language reform: history and future*, 85–99. Hamburg: Helmut Buske Verlag.

Bender, Byron W. and Judith W. Wang, 1985, The status of Proto-Micronesian. In Pawley and Carrington, eds 1985:53–92.

Blust, Robert, 1990, Patterns of sound change in the Austronesian languages. In Philip Baldi, ed. *Linguistic change and reconstruction methodology*, 231–267. Trends in Linguistics — Studies and Monographs 45. Berlin: Mouton de Gruyter.

1996, The Neogrammarian hypothesis and pandemic irregularity. In Mark Durie and Malcolm Ross, eds *The comparative method reviewed: regularity and irregularity in language change*, 135–156. New York and Oxford: Oxford University Press.

in prep. *An Austronesian comparative dictionary*. Electronic manuscript.

Camden, William G., 1979, Parallels in structure of lexicon and syntax between New Hebrides Bislama and the South Santo language spoken at Tangoa. In P. Mühlhäusler et al., ed. *Papers in pidgin and creole linguistics* No. 2, 51–117. Canberra: Pacific Linguistics.

Clark, Ross, 1985, Languages of north and central Vanuatu: groups, chains, clusters and waves. In Pawley and Carrington, eds 1985:199–236.

Early, Robert, 1994, A grammar of Lewo, Vanuatu. Unpublished PhD thesis, The Australian National University.

2002, Lamen. In John Lynch, Malcolm Ross and Terry Crowley, eds *The Oceanic languages*, 671-680. London: Curzon Press.

Fox, G.J., 1979, *Big Nambas grammar*. Canberra: Pacific Linguistics.

Haudricourt, André-Georges, 1968, La langue de Gomen et la langue de Touho en Nouvelle-Calédonie. *Bulletin de la Société de Linguistique de Paris* 63:218–235.

Jackson, Frederick H., 1986, On determining the external relationships of the Micronesian languages. In Paul Geraghty, Lois Carrington and S.A. Wurm, eds *FOCAL II: Papers from the Fourth International Conference on Austronesian Linguistics*, 201–238. Canberra: Pacific Linguistics.

Lynch, John and Terry Crowley, 2001, *Languages of Vanuatu: a new survey and bibliography*. Canberra: Pacific Linguistics.

Maddieson, Ian, 1989, Linguo-labials. In Ray Harlow and Robin Hooper, eds *VICAL I (Oceanic languages) – Papers from the Fifth International Conference on Austronesian Linguistics*, 349–375. Auckland: Linguistic Society of New Zealand.

Maddieson, Ian and Victoria B. Anderson, 1994, Phonetic structures of Iaai. *UCLA Working Papers in Phonetics* 87:163–182.

Ozanne-Rivierre, Françoise, 1976, *Le iaai, langue mélanésienne d'Ouvéa (Nouvelle-Calédonie): phonologie, morphologie, esquisse syntaxique*. Paris: SELAF.

1984, *Dictionnaire iaai-français (Ouvéa, Nouvelle-Calédonie)*. Paris: SELAF.

1986, Redoublement expressif et dédoublement des séries consonantiques dans les langues des Iles Loyauté (Nouvelle-Calédonie). *Te Reo* 29:25–53.

1995, Structural changes in the languages of northern New Caledonia. *Oceanic Linguistics* 34/1:45–72.

Pawley, Andrew and Lois Carrington, eds, 1985, *Austronesian linguistics at the 15th Pacific Science Congress.* Canberra: Pacific Linguistics.

Rehg, Kenneth L., 1981, *Ponapean reference grammar.* Honolulu: University Press of Hawai'i. [PALI Language Texts: Micronesia.]

1991, Final vowel lenition in Micronesian languages: an exploration of the dynamics of drift. In Robert Blust, ed. *Currents in Pacific linguistics: papers on Austronesian languages and ethnolinguistics in honour of George W. Grace*, 383–401. Canberra: Pacific Linguistics.

Rivierre, Jean-Claude, 1978, Accents, tons et inversion tonale en Nouvelle-Calédonie. *Bulletin de la Société de Linguistique de Paris* 73:415–443.

Sam, Léonard Drilë, 1995, *Dictionnaire drehu-français.* Nouméa: CTRDP et CPRDP de Iles.

Tryon, D.T., 1976, *New Hebrides languages: an internal classification.* Canberra: Pacific Linguistics.

Tryon, D.T. and M.-J. Dubois, 1969, *Nengone dictionary*, part 1: *Nengone–English.* Canberra: Pacific Linguistics.

1971, *Nengone dictionary*, part 2: *English–Nengone.* Canberra: Pacific Linguistics.

Lynch, J. "The bilabials in Proto Loyalties". In Lynch, J. editor, *Issues in Austronesian historical phonology*.
PL-550:153-174. Pacific Linguistics, The Australian National University, 2003. DOI:*10.15144/PL-550.153*

Tematheses in Rotuman

HANS SCHMIDT

1 Introduction

Rotuma is a small island in the South Pacific, located roughly at the crossroads between Polynesia, Melanesia and Micronesia. Politically, the island forms part of the Republic of Fiji, though the closest Fijian island, Cikobia, is about 465 km distant (Woodhall 1987:1). Nowadays the island is accessible from Suva, the capital of Fiji, in a two and a half day boat trip or in two and a half hours by plane. In contrast to its northern neighbour Tuvalu, Rotuma is not a coral atoll but a so-called 'high' island of volcanic origin (Pleistocene); its surface area is 46 km^2 and its soil is very fertile.

Rotuma has a population of approximately 2700 inhabitants (59 per km^2) living in 20 villages scattered along the coast, which gives it the highest population density of all Fijian islands (Walsh 1982:20). However, three times as many Rotumans have left their home island for the urban areas of Fiji or live overseas. Many of these Fiji-Rotumans have never been on Rotuma or have been there just for a brief Christmas holiday.

Rotuman, Fijian, and the Polynesian languages form the Central Pacific subgroup of Oceanic. In contrast to its small number of speakers, Rotuman has featured frequently in works of general and comparative linguistics. What makes Rotuman so interesting in the eyes of linguists is its two 'phases', metathesis, and the unusually complex vowel phonology — unusual at least for Oceanic or Polynesian languages. 'This language has provoked Oceanic linguists into doing some of their best work. Its wonderfully intricate morphophonology has teased phonological theorists and the challenge of trying to work out Rotuman's historical position and development has had some important spin-offs for Oceanic historical linguistics. Rotuman has been the *agent provocateur* in two of the foundation studies of the modern period of Oceanic comparative linguistics, those of Grace (1959) and Biggs (1965)' (Pawley 1996:86).

2 What is metathesis in Rotuman like?

Let us look at some examples which might be familiar to many readers:[1]

[1] Language abbreviations used are PCP = Proto Central Pacific, PEO = Proto Eastern Oceanic, PPn = Proto Polynesian, and SF = Standard Fijian.

John Lynch, ed. *Issues in Austronesian historical phonology*, 175–207.
Canberra: Pacific Linguistics, 2003.

Table 1: Related words from Central Pacific languages

gloss	Fijian	Samoan	Tongan	metathesised Rotuman form	non-metathesised Rotuman form
'orange'	*moli*	*moli*	*moli*	*mör*	*mori*
'strand of hair'	*tobe*	*sope*	*tope*	*söp*	*sope*
'five'	*lima*	*lima*	*nima*	*liam*	*lima*
'figure'	*vika*	*fika*	*fika*	*fiak*	*fika*
'coffee'	*ko(f,v)(i,e)*	*kofe*	*kofi*	*köf*	*kofi*
'candy'	*loli*	*lole*	*lole*	*löl*	*loli*
'paper'	*veva*	*pepa*	*pepa*	*peap*	*pepa*
'sugar'	*suka*	*suka*	*suka*	*suak*	*suka*
'wish'	–	*finagalo*	*finangalo*	*fiangar*	*fiangaro*
'tree sp. (*Pisonia grandis*)'	–	*pu'avai*	(PPn **puka-wai*)	*puakvai*	*puakvai*

The 'strange' sound of Rotuman (in comparison with its close relatives) is largely due to its metathesis. As can be seen from Table 1, metathesis involves more than just the final vowel and the preceding consonant changing place.

The earliest written records of the language show that metathesis was present then in Rotuman. Lesson (landed on Rotuma on 1 May 1824) gave *talian* [θa'lyaŋ] for 'ears', the short form of contemporary *faliga*; Bennett (1830, published 1831) gave *Fangwot* [fa'ŋwɔt] for the short form of the place name *Fag'uta*; and Turner (1845, published 1884) gave *lium* ['lyɔm] for 'five', the short form of *lima*.

Metathesis is still a productive process. It is applied to most loan words, too. But it cannot be seen in isolation; it stands at the core of a morphological process in Rotuman. This may be called short-form derivation.

3 Short-form derivation

3.1 Scope of application: which words have short forms for what?

With the exception of mostly monosyllabic grammatical particles[2] (i.e. prepositions like *'e* and *se*, conjunctions such as *ma, ka, la*, etc.) and a handful of other words, every Rotuman content word (lexical morpheme) has two forms.[3] These have been referred to as 'long and short form' (Biggs 1959:24, 1965:388; Milner 1971:418), 'primary vs. secondary form' (Churchward 1929:283), 'complete and incomplete phase' (Churchward 1940:13ff.), 'absolute and construct case' (Hocart 1919:257), and 'proper & original form vs. altered or construct form of the words' (Hale n.d.:123).

When is which form used? One hundred and fifty years ago, Hale (1846:469) had already recognised some regularity in this process: 'A general law appears to be that when a word stands by itself, not followed by another on which it depends, it must terminate in a vowel, and this appears to be the proper and original form of most of the words; but when combined, in any way whatsoever, with other words, an alternation takes place by which the

2 These are also excepted from another rule, namely that each word must contain at least two morae, i.e. two vowels or a long vowel (Blevins 1994:491).

3 Some words have even more: *mi'a* and *miạ'*, *mi'e* and *mie'* 'red' (Churchward 1940:87).

concluding syllable is so transposed or contracted as that the consonant should be the final letter'.

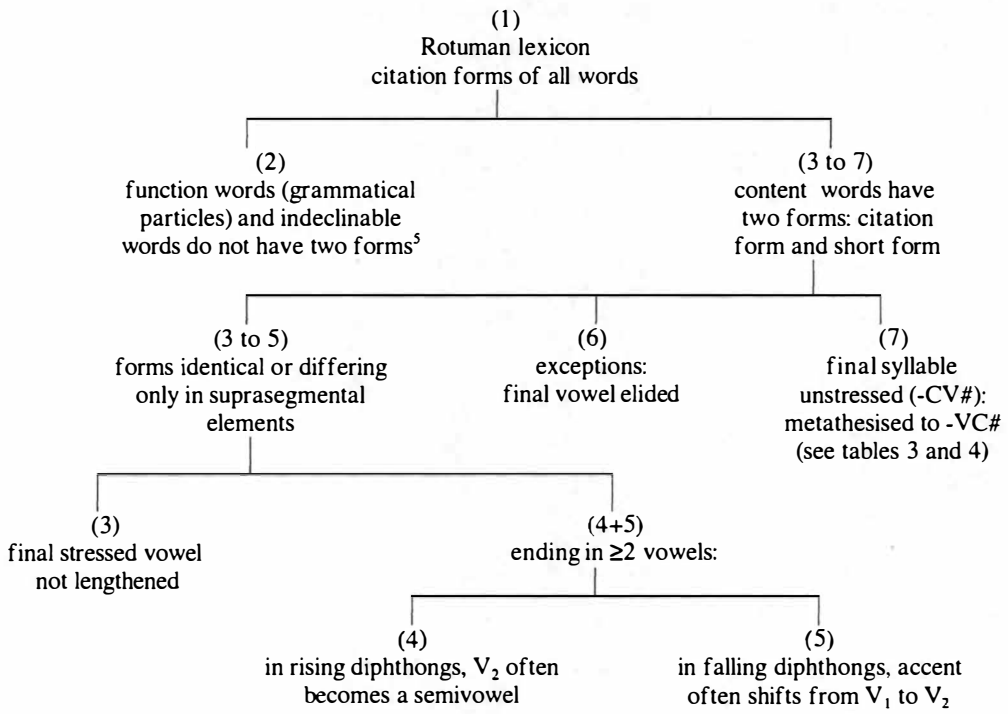
To put it in simple words, within a phrase all words except the last one are in their short form. All morphemes in a compound word except the final one are always in their short form, the final morpheme being in its long form only if the word is phrase-final. The shape of this final element shows whether a word or compound or a whole phrase is definite and specific (citation form) or indefinite and unspecific (short form).[4] The meaning of the words usually remains unchanged — contrary to what Vamarasi (1991:2) claimed.

Table 2: Examples of the two forms

complete phase	gloss	incomplete phase	gloss
hóla	'spread out, spreading'	*hoál ta*	'the spread(ing)'
hoál 'épa	'spread out mats'	*hoál 'eáp ta*	'the mat-spreading, spread mat'
hoál 'eáp fúri	'spread out zigzag mats'	*hoál 'eáp für ta*	'the spread(ing)'

One can group Rotuman words according to the way their short forms are derived.

Diagram 1: Dividing the Rotuman lexicon according to short-form derivation



4 Compare Churchward (1940:14, 88ff.), Besnier (1987:203ff.) and Blevins (1994:493).

5 Polysyllabic particles resemble in their form either the complete (*'ita-ke, 'ea-ke, ko-ta, se-minte*) or the incomplete phase (*kat, kal, sin, mar, kam*). Colloquially *ra*, the second part of the negation, is cliticised to the preceding word and shortened to *-r*, e.g. *gou kat 'inea-r* < *gou kat 'inea ra* 'I don't know'. The contrary can be observed in poetry and songs: *kat* > *kate*, *sin* > *sini* (cf. Table 13).

Notes:

re (2) Indeclinable words form a closed group within the Rotuman lexicon (see list in Churchward 1940:86.5). Nowadays it is expanded by the import of loan words from English and Fijian.

re (3) Final vowels are lengthened only when they carry the stress. In the short form they fall automatically into a stressed slot within a phrase so that no lengthening (in order to attract the accent) is required. For example:[6] *uakē* [ua'ke] 'brawl' → *uakĕ* in *uaké ta* [ua'keta] 'the brawl'.

re (4) Biggs (1959:24ff.) regarded final rising diphthongs as belonging to two different syllables and then concluded that they merged in the short form into one syllable whereby the less sonorant of the two became the semivowel. For example: *vo̯i* ['vɔi] > ['vɔy].

re (5) Words ending in falling diphthongs build their short form usually by shifting the accent from the penult to the final syllable (Churchward's third declension): *koría* 'sailing boat, sailing boats in general' → *koriá he* 'a sailing boat'.

re (6) Pronouns (except the dual forms), pronominal and directional suffixes (like *-me*, *-na*) form their short form irregularly by deleting the final vowel (Churchward 1940:85ff.). Within group (7) there are many cases where the short form looks like the long form without its final vowel. In reality it has coalesced with the vowel of the preceding syllable after metathesis, though sometimes leaving no traces (examples under rule 2 in the following section).

re (7) All other content words have the canonical shape of the ending $-V_1CV_2\#$ in their citation form and build their short form via metathesis. How this works is demonstrated in Tables 3 and 4.

Table 3: Endings of the short forms in the conventional spelling

	V_2 =	/i/	/u/	/e/	/o/	/a/
V_1 =						
/i/		iC	iC	ieC	ioC	iạC
/u/		üC	uC	ueC	uoC	uạC
/e/		eC	eC	eC	eC	eaC
/o/		öC	oC	öC	oC	oaC
/a/		åC	ạC	äC	aC	aC

6 Vamarasi's (1991:6) view that 'words with long vowels have no short form at all', cannot be upheld since words like *tāope* do have a short form (*tāöp*). She was probably referring to *final* long vowels.

Table 4: Endings of the short forms in contemporary pronunciation

V_2 =	/i/	/u/	/e/	/o/	/a/
V_1 =					
/i/	iC	iC	yeC	yuC	yɔC
/u/	ɨiC	uC	weC	woC	wɔC
/e/	eC	eC	eC	eC	yaC
/o/	øC	oC	œC	oC	waC
/a/	æC	ɔC	æC	aC	aC

The tables above should be read as follows. On the far left (V_1 stands for the vowel of the penultimate syllable of the long form) and above each table (V_2 stands for the final vowel) I have listed the five original vowels of Rotuman, /a e i o u/. I think that the language had only these five common Oceanic vowels when this process started in Pre-Rotuman, and will try to argue for my assumption in the next section. Thus suppose V_1 = /u/ and V_2 = /a/, that is a morpheme ends in /uCa/ in the citation form, then its short form is phonemically /uaC/ but will be written *uạC* and pronounced [wɔC]. For example: *hula* ['hula] → *huạl* ['hwɔl] 'moon'.

Other phonological processes have applied later and changed the participating vowels even further to blur the picture (see §3.5). Some authors see it exactly the other way round: that ablaut happened *before* the creation of short forms, e.g. *hạfu* < Pre-Rotuman *hɔθu (Biggs 1965:388; see Table 6).

3.2 Review of previous explanations

Various linguists have come up with different scenarios and systems of rules to explain how the short forms developed out of the above 25 citation forms with an unstressed final syllable. I believe that the core element of this transformation (or incomplete phase formation) is metathesis of the unstressed final syllable: the final vowel and the immediately preceding consonant change place.

Let me present the main attempts by earlier writers at explaining this derivation and then give my comments as well as my own system of rules (in §3.3).

3.2.1 Churchward

Churchward began his grammar by listing the phonemes and then straight away describing the two forms or phases. He regarded the citation form as the original form. He grouped Rotuman words into four declensions according to the way they form their short forms. The first two declensions both end in -VCV. In his first declension the short form is created by deleting the final vowel of the citation form. In addition, in declension 1b to 1d the vowel of the penultimate syllable is changed into the corresponding umlaut. When the final vowel of the citation form is lower than the penultimate one, metathesis of the final vowel with the preceding consonant occurs (second declension). When a citation form ends in two or more vowels, the (stressed) penultimate vowel will be somewhat shortened (third declension). Words ending in a long vowel as well as indeclinable words remain unchanged (fourth declension).

More recently, people have seen metathesis as the first stage of short-form derivation; but for Churchward, the starting point was the deletion of the unstressed final vowel. Unfortunately he does not list the reverberations this has on the vowel of the root syllable (the penultimate or stressed one of the most common shape of Rotuman words, i.e. two-syllable morphemes). The rule defining his second declension has been copied many times because it is correct (Milner's third rule, Vamarasi's first one and Geraghty's fourth); but it does not apply to the occurrence of metathesis, but rather to the preservation of the vowel pair created by metathesis. He did not mention that the first of them becomes a diphthong later on.

3.2.2 Biggs

Biggs (1959 and 1965) saw the underlying principles. The base of all short forms is the interchange/exchange/swap (metathesis) of the final two phonemes of the citation form. Later the accent moved from the penult to the final syllable and the short form lost one syllable, either because the less sonorant of the two now adjacent vowels became a semivowel or because two similar vowels merged into a single one. This coalescence preserved the rounding of one of the vowels[7] and the fronted position of the other in the resulting umlaut. For example:

Table 5: Development of the short form according to Biggs (1965:389)

pre-Rotuman citation form	metathesised form	contemporary short form	in Churchward's spelling	gloss
/óta/	/óat/	[wat]	*oat*	'sago'
/láje/	/læej/	[lætʃ]	*läj*	'coral'
/séru/	/séur/	[sör][8]	*ser*	'comb'
/'úli/	/'úil/	[ʔül]	*'ül*	'skin'
/kámi/	/káim/	[kœm][9]	*kåm*	'dog'

Some corrections to the bottom row are called for: the short form of /kámi/ is *kåm* [kæm] and not [kœm]. Actually he should have written them /kɔmi/ and /kɔim/, since he claimed that [æ a ɔ], the three allophones of /a/ had already been created in Pre-Rotuman.[10] The short form *kåm* [kæm] can be much easier derived from a hypothetical interim stage of /káim/ than from /kɔim/ (see under rule 7 in §3.5.2).

In his example /kámi/ 'dog', the root vowel /a/ is today rounded [ɔ]; but when the fashion of creating short forms started, it was not rounded yet and consequently the resulting umlaut in *kåm* [kæm] is not rounded. So his rule that rounded vowels [ɔ o u] merging with front vowels [e i] into a front rounded umlaut [œ ø üi] is not correct, since *u* and *e* did not merge in the short form: *pure* > *puer* 'cowry shell'.

---

7 The /a/ in /kámi/ is not rounded and therefore neither is the resulting umlaut in *kæm*. A rule prescribing that the rounded vowels [ɔ o u] and the front vowels [e i] merge into a rounded front umlaut [œ ø üi] is therefore not correct because *u* and *e* in the short form do not merge: *pure* > *puer* 'cowry shell'.

8 This variant pronunciation or vowel coalescence does not apply to all cases of the ending *-eCu*, though for most which end in the other high vowel (see Table 9).

9 Biggs (1959:26) earlier transcribed it as [kɔ̈m], but it should be [kæm].

10 I wonder how a loan word *kamia* ('come here') can be labelled pre-Rotuman?

3.2.3 Milner

Milner (1971) also recognised that the underlying vowels were /a e i o u/ and that metathesis was the origin of the multitude of vowels in Rotuman. He gave the following rules for the derivation of the short form (1971:418ff.):

(3.1) If V_1 is low (i.e. /a/) and V_2 is a front or high vowel (i.e. /e i u/), V_2 is elided and V_1 changes into a mid-high allophone (i.e. /æ ɔ/); it becomes /ɔ/ before /u/ and /æ/[11] before /e i/.

(3.2) If V_2 and V_1 are identical or if V_1 is not higher than V_2, then V_2 is deleted (i.e. for endings like eCo, iCu, eCu, aCo).[12] If I had regarded the ending of short forms in [ɔC] derived from */auC/ as caused by elision (rule 2b) instead of vowel merger + ablaut formation (rule 3b), a rule could be formulated which is as nice and simple as Milner's second rule.

(3.3) If V_2 is lower than V_1, metathesis occurs (Churchward's second declension).

(3.4) If V_2 is front and V_1 back and V_1 not lower than V_2, then V_2 is deleted and V_1 becomes the front rounded allophone of V_1 'as in ['futi] ⟶ [füt], ['mose] ⟶ [møs], ['ofi] ⟶ [øf], /lagi/ ['lɔŋi] ⟶ [lœŋ].'[13] In all cases back vowels would turn into their rounded equivalents in front position (cf. my comments following rule 7 in §3.5.2 and under Table 5).

In his view (Milner 1971:421), the phonetic equivalents of the five allophones generated by rules 1 and 4 were the following:

ä [æ] *å* [œ] *ạ* [ɔ] *ü* [ü] *ö* [ø]

As mentioned above, Churchward's *å* is pronounced [æ] (cf. also Besnier 1987:209; Biggs 1959:24) and it is not rounded; *ö* is [œ] or [ø], so the correct series should be:

ä [æ] *å* [æ] *ạ* [ɔ] *ü* [ü] *ö* [ø, œ]

3.2.4 Cairns

For Cairns (1976), two alternative orderings of the rules were equally probable:

either 1. fronting, 2. metathesis, 3. umlaut formation, 4. elision

or 1. metathesis, 2. vowel coalescence, 3. elision.

3.2.5 Anttila

Anttila (1989) relied only on 'Churchward's ... sometimes confusing description ... [and] imprecise characterization of Rotuman vowels and stress' (Besnier 1987:202). He assumed rightly that the Rotuman vowel system was not only expanded because of short form creation

---

11 By error he also wrote [œ] instead of [æ].

12 Cairns (1976:275) and also Vamarasi (in her rule 2) believed that 'metathesis only occurred if the final vowel was lower than the penultimate one'.

13 The resulting umlaut in *låg* (< *lągi*) is [æ], and not [œ]. The same error cropped up in Biggs (see §3.2.2 above).

but other processes as well. But he was not right in claiming (1989:110 and 114) that, 'it is vital that the process of the 'raising umlaut' has occurred *before* creation of the short form' (see my reasons in §3.5.2). Another of his assumptions is highly improbable, namely 'that the raising of /a/ to [æ] was completed before the fronting of /a/ before /i/ had occurred' (Anttila 1989:64). Both are results of the vowel merger after metathesis: /a/ + /e/ > /a+e/ > *ä* [æ], i.e. the raising[14] of /a/ after its merger with /e/, and /a/ + /i/ > /a+i/ > *å* [æ], i.e. the raising of /a/ after its merger with /i/.

The short form is said to be built according to three rules, the first one of them '(fronting with umlaut formation and deletion of the final vowel) being complementary to the other two changes (metathesis and vowel shortening); so all three occupy the same position in the relative chronology'. For one thing, this is impossible since it requires metathesis to bring the two vowels next to each other before they can be shortened or changed into the umlaut; on the other hand, vowel coalescence does not involve palatalisation only (the opposite process applies to /e+u/ and /a+u/).

3.2.6 Besnier

According to Besnier (1987:205) the 'incomplete forms of Rotuman words are derivable from the complete forms through the following four processes:

(6.1) a rule of metathesis inverting the order of the last vowel of the word and of the immediately preceding consonant, if there is one;[15]

(6.2) a rule of vocalic assimilation that reduces certain vocalic pairs obtained through metathesis to a single vowel whose phonological characterisation is a combination of the distinctive features of the two vowels in the underlying pair;

(6.3) a rule of reduction that changes the first vowel of other vocalic pairs into a glide, thus reducing the underlying pair to a diphthong; and

(6.4) a rule of length reduction that shortens clusters of similar vowels obtained from rule (6.1) to single vowels.'

After the application of rule (6.1) various vowel pairs develop (1987:208ff.):

(6.3.1) 'Vowel clusters consisting of a high vowel (/i u/) followed by a non-high non-back vowel (/e a/) reduce to a monosyllabic diphthong[16] consisting of a glide that corresponds in roundness to the first underlying vowel, followed by the round vowel [ɔ].' This explanation is valid for the patterns /iCa/ and /uCa/, but unfortunately Besnier failed to identify endings with *-e* as variants ('narrow versions', see §3.5.3) of *-a*.

(6.3.2) 'Vowel clusters consisting of the mid-high vowels (/e o/) followed by /a/ or of the high vowels (/i u/) followed by /o/ reduce to a monosyllabic diphthong consisting of a glide that corresponds in roundness to the first underlying vowel, followed by the second underlying vowel.'

---

14 The main thing is fronting and not — if it can be discerned at all — raising (cf. Table 8).

15 More relevant is that the final vowel is unstressed.

16 Isn't that what defines a diphthong, that it unites two vowels in one syllable?

Churchward had described this reduction rule (6.3) in a more elegant way in his second declension, realising that the ablaut of /a/ to [ɔ] in V_2 was not part of short-form derivation, but a later development.

(6.2.1) A back vowel (/o u/) followed by /i/ and also /o/ followed by /e/ (i.e. oCi, oCe, uCi) are reduced to a single front rounded vowel [ø ü] whose height preserves the height of the first underlying vowel of the pair. The formation of umlaut is cyclical, i.e. it spreads to preceding identical vowels.

(6.2.2) A low vowel (/a/) followed by the high vowel /i/ is reduced to [ɛ].[17]

(6.4) All other vowel pairs were reduced to the first underlying vowel. He includes also umlauts such as *käs* < *käse* and *haṣ* < *haṣu*, since he assumed that the ablauts had already been present in the citation form before the derivation of short forms: 'Clusters whose underlying form is /a(C)e/ will be fed into the metathesis rule as [ɛ(C)e], and those of underlying form /a(C)u/ and /a(C)i/, as [ɔ(C)u] and [ɔ(C)i] respectively' (Besnier (1987:207).

Besnier used two secondary vowels [ɔ] and [ɛ] as starting points or underlying vowels next to the five basic vowels in his scheme of rules. The ablaut of the root vowel /a/ in the citation form is due to more recent partial regressive assimilation or copying of the pronunciation of the short form. Endings like [ɔC] or [æC] have not been created by elision of the final vowel but rather through metathesis and vowel coalescence. (See my rules 1 and 3 in §3.3)

3.2.7 Vamarasi

Vamarasi (1991:2) listed three processes with which to derive the short forms from the long ones, 'depending on the order of the vowels in the penultimate and final syllable of a word' (for words ending in CV). The order is less important than their quality. The three processes are metathesis, umlauting and vowel loss.

(7.1) 'Metathesis of final CV to VC occurs when the final vowel is lower than the penultimate. The resulting vowel combination develops into a diphthong' (Churchward's second declension).

(7.2) 'The final vowel is deleted and the penultimate vowel becomes an Umlaut, if a back vowel in the penultimate syllable is followed by a front vowel in the final syllable. This rule must apply after metathesis because the combination of V_1 = /u/ and V_2 = /e/ is taken care of by metathesis rather than Umlaut formation.'

(7.3) 'The final vowel is deleted, if it is identical with the preceding one or ...', and here she simply listed all other combinations without being able to summarise them.

In her analysis, the accent shift from the final syllable and the reduction of syllables in the short form are missing. The creation of umlauts in (7.2) is not a contrasting alternative to metathesis but rather a further step, once metathesis brought two vowels into the immediate vicinity of each other.

---

17 I hear it as [æ].

3.2.8 Geraghty

Geraghty (1995:933ff.) gave the following rules:

(8.1) Metathesis of final vowel and of the preceding consonants

(8.2) a. Reduction of double vowels if $V_1 = V_2$ (e.g. *ala* → *aal → *al*)

(8.2) b. Vowel elision: V_2 is elided except if V_1 is higher (e.g. *hifu* → *hiuf → *hif* (identical with Milner's second rule)

(8.3) Creation of semivowels: if V_1 is higher than V_2, it becomes the appropriate semivowel (e.g. *aire* → *aier* → [a'yer])

(8.4) Umlaut: a non-front vowel followed by a vowel which is not lower, becomes an umlaut:

-aCi → *-aiC → -æC
-oCe → *-oeC → -øC[18]
-oCi → *-oiC → -øC
-uCi → *-uiC → -üC

(8.5) Syllable reduction

(8.6) Accent shift from penultimate to final syllable.

Geraghty's model (1995:933ff.) is the best so far presented. Nevertheless, the order of his rules 8.3, 8.5 and 8.6 does not seem logical. The penultimate syllable had to lose the accent first (rule 8.6) before being deleted (rule 8.5). This accent shift in turn is the most probable trigger for the now unstressed penultimate vowel to become a semivowel (rule 8.3).

Table 6 gives a simplified comparison of the major attempts at devising and ordering rules of short-form derivation as far as they can be squeezed into a rigid frame:

Table 6: Different orders of rules for short form derivation

rule no.		process	Geraghty 1995	Vamarasi 1991	Besnier 1987	Cairns 1976	Milner 1971	Biggs 1965	Churchward 1940
1		Metathesis	1	1	2	2	3	2	3
2a		Deletion of V_2 if identical with V_1	2	3	5	4	2		1
2b		Elision of V_2	3	3	5	4	1		1
3		Vowel merger:	5	2	3	3		4	2
	3a	umlaut formation			3	3	4		
	3b	ablaut formation					1		
4		Accent shift	7	1				3	
5		Semivowel development	4	1	4			4	
6		Syllable reduction	6	1				5	
7a		Backing + rounding of /á/ to [ɔ] before /u/			1		5	1	
8		Fronting of /á/ to [æ]			1	1	5		
9a		Backing + rounding of /á/ to [ɔ] after /u/					5		

[18] The resulting vowel should be transcribed as [œ].

10a	Raising + fronting of unstressed /a/ to [e] after /i/			-				
11	Extension of ablaut rules (7a, 9a, 10a) to the other high vowel (7b, 9b, 10b)			1			1	

3.3 My attempt at devising and ordering rules

In my opinion, words with unstressed final syllable (i.e. ending in $-V_1CV_2$) form their short forms according to the following rules:

Rule 1: Metathesis

V_2 and C are interchanged: $V_1CV_2\# \longrightarrow V_1V_2C\#$

Rule 1 applies to all 25 cases (in Tables 3 and 4). Blevins (1991:2) ordered the rules in the same way, suggesting that the derivation of the incomplete phase included apparent metathesis of a final CV pair with following assimilation and deletion. Anttila (1989:64), on the other hand, assumed that there was no 'direct evidence for the interim form, i.e. first metathesis, *futi* → **fuit*, and then **fuit* → *füt*'. But what else do short forms such as *hoál, puér, suák, tapiók* show? Only metathesis can have created the results of rule 5 and the umlauts from rule 3 equally well.

Milner (1971:422) also recognised this and proposed to spell the short form always in a phonemic way with $-V_1V_2C\#$ finally, irrespective of its current pronunciation. That would require only the five basic vowels and no special characters for the umlauts and ablauts.

Table 7: Milner's proposal for a new spelling of the Umlauts

citation form in Churchward's spelling	short form in Churchward's spelling	pronunciation	short form in Milner's spelling	gloss
mose	*mös*	[mœs]	*moes*	'sleep'
futi	*füt*	[füt]	*fuit*	'pull'
ạsu	*ạs*	[ɔs]	*aus*	'steam'
ạti	*åt*	[æt]	*ait*	'gather (shellfish)'

Milner's spelling, however, would create additional homographs in the short form:

Table 8: Homographs as a consequence of Milner's proposed phonemic spelling

Milner's proposal	pronunciation	could be mistaken for	pronunciation	explanation
moes	[mœs]	*'omoes*	[ʔo'moes]	citation form of *'omoe* + interrogative suffix *-s*
fuit	[füt]	*fuit*	['fuit]	citation form of *fui* + indefinite article *-t*
aus	[ɔs]	*'aus*	['ʔaus]	short form of *'ausa*
ait	[æt]	*'ait*	['ʔɔit]	short form of *'aitu*
tair	[tær]	*tạir*	[tɔ'ir]	short form of *tạiri*

Rule 2: Deletion of the unstressed vowel V_2 (applies to 5 + 6 = 11 cases)[19]

consisting of

Rule 2a: Shortening double vowels into a single one:

$V_1CV_1\# \longrightarrow *V_1V_1C\# \longrightarrow V_1C\#$

and

Rule 2b: Deletion of the unstressed vowel V_2:

$V_1CV_2\# \longrightarrow *V_1V_2C\# \longrightarrow V_1C\#$

V_2 is deleted if V_1 is not further back than V_2 and if V_2 is not lower than V_1. When two identical vowels come to be next to each other, this does not result in a long vowel, but the double vowel is reduced to a single one. That is why I group these cases under elision and not coalescence (rule 3).

It is less probable that a stressed vowel is elided and therefore I think that after metathesis, the stress was still on V_1. I guess the accent shift was the fourth step (rule 4).

Table 9: Deletion of final vowel after metathesis

citation form	metathesised form	short form	examples
/áCo/	→ /*áoC/	→ [aC]	*rako, rak*
/éCi/	→ /*éiC/	→ /eC/ (often [əC, øC])	*fesi, fes*
/éCu/	→ /*éuC/	→ /eC/ (often [əC, øC])	*seru, ser*
/íCu/	→ /*íuC/	→ [iC]	*hifu, hif*
/éCo/	→ /*éoC/	→ [eC]	*he'o, he'*
/óCu/	→ /*óuC/	→ [oC]	*folu, fol*
/áCa/	→ /*áaC/	→ [aC] (not āC)	*fara, far*
/éCe/	→ /*éeC/	→ [eC]	*sere, ser*
/íCi/	→ /*íiC/	→ [iC]	*miji, mij*
/óCo/	→ /*óoC/	→ [oC]	*ono, on*
/úCu/	→ /*úuC/	→ [uC]	*lumu, lum*

It could be argued that the short-form endings at the top of the above table are the result of vowel merger (rule 3) rather than deletion, since /é/ followed by a high vowel (or a deleted one) is often pronounced like a high rounded central vowel; see Table 10.

Table 10: minimal pairs through vowel elision

citation form	short form + article	pronunciation	gloss
sere	*ser ta*	['serta]	'(the) knife'
seru	*ser ta*	['sørta]	'(the) comb'

19 Similar to Churchward's first declension, cases 1–11 in Table 14.

Hocart (1919:255; Grace 1959:27ff.) gave mid-high allophones of /e/ and /o/, narrowed by a following high vowel. Churchward heard an allophone of /á/ when it was followed by a glottal consonant and /o/ and called it 'posterior a'.

Rule 3: Vowel merger (applies to 6 cases)

V_1 and V_2 merge or coalesce

In other words, V_2 gives some of its qualities to V_1 and then drops out. This occurs either if V_2 is more fronted and not lower than V_1, or if V_2 is high and V_1 low. Thus there are two sub-rules:

Rule 3a: Umlaut formation (applies to 3 cases: cases 15–17 in Table 14)

V_1 and V_2 merge into an umlaut if V_1 is a rounded back vowel [o u] and V_2 a front vowel which is not lower than V_1

Rule 3b: Ablaut formation (applies to 3 cases: cases 12–14 in Table 14)

V_1 and V_2 merge into an ablaut if V_1 is low (= [a]) and V_2 is a high (= [i u]) or a front vowel (= [i e])

Thus five umlauts and ablauts result from metathesis and subsequent vowel merger:

Table 11a: Umlaut and ablaut formation through vowel merger (I)

citation form	metathesised form	short form	examples
/óCi/	→ /*óiC/	→ öC [øC]	*mori, mör*
/óCe/	→ /*óeC/	→ öC [œC]	*tole, töl*
/úCi/	→ /*úiC/	→ üC [üC]	*kuji, küj*
/áCu/	→ /*áuC/	→ ạC [ɔC]	*hạfu, hạf*
/áCi/	→ /*áiC/[20]	→ åC [æC]	*sạsi, sås*
/áCe/	→ /*áeC/	→ äC [æC]	*päre, pär*

In other words:

Table 11b: Umlaut and ablaut formation through vowel merger (II)

V_1	+	V_2		merged vowel	change	feature
a	+	e, i	→	ä, å	V_1 is fronted	+front
a	+	u	→	ạ	V_1 is rounded and backed	+round +back
o	+	e, i	→	ö	V_1 is fronted and stays rounded	+front
u	+	i	→	ü		

Diagram 2 shows the five new vowels which have arisen out of V_1 due to metathesis and subsequent merger with V_2:

20 And not *ɔ̈iC.

Diagram 2: Umlaut and ablaut creation

It becomes evident from this diagram how exceptional the backing of /a/ to [ɔ] is.

Within this rule, the 'real' umlauts (*ö* and *ii*) constitute a special group. The assimilation of preceding identical vowels ('spreading') is obligatory in their case, but for the ablauts *ä* and *å* only optional, for [ɔ] formed out of *a+u* unusual (see Blevins (1991:2):[21] *pulufi* → *pülüf*, *konosi* → *könös*; but *haharągi* (not **hąhąrągi) → *haharåg* (not **håhåråg), *kanąpu* (not **kąnąpu) and *hagäe* (not **hägäe), *'anasi* (not ** 'ąnąsi) → *'anås* (not ** 'ånås).

Cairns (1976:274) had not heard the language spoken and thus claimed incorrectly that '[æ] occurs only in exactly those short forms that also contain an [æ] in their citation form'. [æ] in short forms goes back to an original /a/ in the citation form with a following front vowel which later changed into one of the ablauts [æ] or [ɔ] (see § 3.5). This process has also produced minimal pairs in the short form rendering it impossible today to re-develop the underlying citation form unambiguously from every short form:

Table 12: Minimal pairs through ablaut

gloss	short form		citation form ending in a front vowel		
'banana'	*får*	[pær]	/pari/	*pąri*	['pɔri]
'guard'	*pär*	[pær]	/pare/	*päre*	['pære]

The unstressed high vowels in final position are deleted after metathesis, usually with umlauting of the root vowel. Similarly, in Tongan, 'high vowels have productive voiceless allophones if they (1) are short and unstressed, (2) follow a voiceless consonant, (3) are situated in final position of a morpheme and (4a) stand at the end of an utterance or (4b) precede a voiceless consonant. The low vowel /a/ is devoiced under conditions 1,3,4 (though only following /h/), but oddly not the mid vowels' (Feldman 1978:137).

Rule 4: Accent shift (applies to all cases)

In Rotuman, the accent is usually placed on the penultimate syllable (of the citation form). In the short form, it shifts to the final syllable. The accent shift to the right is a decisive marker of Rotuman metathesis. Based on the incomplete description by Churchward,[22] Cairns (1976:273) cited examples such as *tíko* → *tíok*, but the short form of words like /tíko/ is [tyok] derived from */tiók/. Later (1976:274) he even formulated 'a rule that assigns

21 Later, the root vowel /a/ in long forms was also changed: before unstressed high vowels (*á(C)u* and *á(C)i*), it was backed and rounded to *ą* [ɔ] (rule 7) and before /e/ (*á(C)e*), it was fronted to *ä* [æ] (rule 8).

22 For example: 'The stress seems to be levelled out, [...] *fora* becomes *foar*, which is pronounced almost, though perhaps not quite, as one syllable, the stress being evenly distributed' (Churchward 1940:86).

stress to every penultimate vowel'. This is correct, but only in the citation form, and only when the final vowel is short (Churchward 1940:85). All Cairns' (1976:275) examples of short forms have the accent on the final instead of the penultimate vowel.

The accent shift cannot have occurred before the elision of V_2 (rules 2 and 3) because a stressed vowel is less likely to be dropped than an unstressed one. The reduction of V_1 to the corresponding semivowel (rule 5) can best be explained if V_1 lost the stress first and then the accent shifted to V_2.

Rule 5: Semivowel formation (applies to 8 cases: cases 18–25 in Table 14)

If V_1 is higher than V_2, it will be changed into the corresponding semivowel (i.e. front vowels to [y], back vowels to [w]), as illustrated in Table 13.

Table 13: Semivowel formation

ending of citation form	ending of metathesised form (rule 1)	ending of short form	
		accent shift (rule 4)	semivowelformation (rule 5)
/íCe/	→ /*íeC/	→ iéC	→ [yeC] (often [yəC])
/íCo/	→ /*íoC/	→ióC	→ [yuC]
/íCa/	→ /*íaC/	→ iáC	→ [yaC]
/úCe/	→ /*úeC/	→ uéC	→ [weC] (often [wəC])
/úCo/	→ /*úoC/	→ uóC	→ [woC]
/úCa/	→ /*úaC/	→ uáC	→ [waC]
/éCa/	→ /*éaC/	→ eáC	→ [yaC]
/óCa/	→ /*óaC/	→ oáC	→ [waC]

Biggs and Besnier described this correctly. Besnier (1987:211ff.) and Blevins (1994:492) made a little mistake though. They mistook the cases of ablaut with *-e* (deriving from *-a*) for the citation forms of short forms ending in *-ạC* [ɔC] (see § 3.5.3). The short forms given by Besnier were correct, namely [tyɔf], [hwɔŋ] etc., but they derive from an underlying long form ending in /a/ (*tifa* and *huga*) and not their variants (*tife* and *huge*). Words ending in /iCe/ and /uCe/ have short forms ending in /ieC/ and /ueC/ respectively. Thus:

Besnier (1987)		should be corrected as follows:	
citation form	short form	citation form	short form
iCe	[yɔC]	iCa	→ [yɔC]
		iCe	→ [yeC][23]
uCe	[wɔC]	uCa	→ [wɔC]
		uCe	→ [weC]

23 Often these endings of the short forms are also pronounced with shwa: [yəC] and [wəC].

Rule 6: Syllable reduction (applies to all 25 cases)

The deletion of the final syllable is a result of the elision of V_2 (rules 2+3) or the weakening of V_1 into a semivowel (rule 5).[24]

Summary

In Table 14 I summarise in which cases and in which order the above-mentioned rules apply.

Table 14: Order of the rules to create the endings of the short forms

	ending of citation form	rules 1.	2.	3.	4.	5.	examples
1.	áCa	*áaC	[aC] (not āC)				*fara, far*
2.	éCe	*éeC	[eC]				*sere, ser*
3.	íCi	*íiC	[iC]				*miji, mij*
4.	óCo	*óoC	[oC]				*ono, on*
5.	úCu	*úuC	[uC]				*lumu, lum*
6.	áCo	*áoC	[aC]				*rako, rak*
7.	éCo	*éoC	[eC]				*he'o, he'*
8.	íCu	*íuC	[iC]				*hifu, hif*
9.	óCu	*óuC	[oC]				*folu, fol*
10.	éCi	*éiC	eC				*fesi, fes*
11.	éCu	*éuC	eC				*seru, ser*
12.	áCe	*áeC		äC [æC]			*päre, pär*
13.	áCi	*áiC		åC [æC]			*sạsi, sås*
14.	áCu	*áuC		ạC [ɔC]			*hạfu, hạf*
15.	óCe	*óeC		öC [œC]			*tole, töl*
16.	óCi	*óiC		öC [øC]			*mori, mör*
17.	úCi	*úiC		üC [üC]			*kuji, küj*
18.	éCa	*éaC			eáC	[yaC]	*pera, pear*
19.	íCa	*íaC			iáC	*yaC [yɔC]	*lima, liạm*
20.	íCe	*íeC			iéC	[yeC]	*aire, aier*
21.	íCo	*íoC			ióC	[yoC]	*tiro, tior*
22.	óCa	*óaC			oáC	[waC]	*mofa, moaf*
23.	úCa	*úaC			uáC	*waC [wɔC]	*usa, uạs*
24.	úCe	*úeC			uéC	[weC]	*pure, puer*
25.	úCo	*úoC			uóC	[woC]	*ulo, uol*

[24] Compare a similar development in Tongan: 'The frequency and regularity of voiceless vowels in words of all origins [including loan words], moreover, lead one to believe that Tongan is well on its way to developing closed syllables' (Feldman 1978:138).

The following tables show which phonological processes apply to which vowel pairs. The abbreviations mean: **E** = elision of the final vowel (rule 2b), **S** = shortening of identical vowels (rule 2a), **M** = semivowel development after metathesis (rule 5), **U** = umlaut (rule 3a), **A** = ablaut (rule 3b).

Table 15a-c: endings of short forms, arranged by phonological processes

V_1 =	V_2 = /i/	/u/	/e/	/o/	/a/
/i/	S	E	**M**	**M**	**M**
/u/	U	S	**M**	**M**	**M**
/e/	E	E	S	E	**M**
/o/	U	E	U	S	**M**
/a/	A	A	A	E	S

V_1 =	V_2 = /i/	/e/	/u/	/o/	/a/
/a/	**A**	**A**	**A**	E	S
/o/	**U**	**U**	E	S	M
/u/	**U**	M	S	M	M
/e/	E	S	E	E	M
/i/	S	M	E	M	M

V_1 =	V_2 = /e/	/a/	/i/	/o/	/u/
/e/	S	M	**E**	**E**	**E**
/a/	A	S	A	**E**	**E/A**
/i/	M	M	S	M	**E**
/o/	U	M	U	S	**E**
/u/	M	M	U	M	S

3.4 Metathesis or not?

Is it really metathesis which is involved, or is it rather the anticipation of the final vowel with its subsequent deletion as Anttila thought? Since metathesis is still productive, one can analyse what happens to loan words.

The shape of many terms borrowed from English with their closed final syllable resembles a short form in Rotuman. Consequently they are first adopted as such into the Rotuman lexicon, which is completely natural, since in the spoken language the short forms are much more common than the long ones. 'This altered or construct form of the words is the one in which they are the most commonly heard' (Hale 1846:469). Hocart (1919:263) remarked on this practice: 'This is the natural tendency of the White Man; for as most words in any sentence are in the construct, it is the form he learns first; when the absolute [case] does occur he does not take much notice of it.' The Catholic missionaries also had initially written all words in their short form — whereas the Methodists first wrote only long forms.

The derivation process is inverted here. When people need a citation form of these loan words, they have to create it completely anew using the same rules in the reverse order. In most cases this simply means to add (rule 2a) an echo vowel (a copy of the vowel of the final

syllable) to the final consonant: i.e. from *kap* 'cup' we make *kapa*; compared to *kapa* 'copper' which was incorporated as a citation form into Rotuman.

English source word	borrowed as	later created form
'cup'	*kap*	long form: *kapa*
'copper'	*kapa*	short form: *kap*

If the vowel of the final syllable of the loan word is not one of the five basic vowels, it is regarded as an umlaut or ablaut and re-cut into two vowels according to the above-mentioned rules. Semivowels are formed back into their corresponding high vowels. This is another example how the process of metathesis was reversed (see also Hocart 1919:258).

Table 16: Examples of later created citation forms ('back formation') of loan words[25]

rule no.	English source word	short form		citation form	
2a	onion	*'anian*	[ʔa'nyan]	/'aniana/	*'aniana*
3b	horse	*hạs*	[hɔs]	/hasu/	*hạsu*
3b	Christmas	*kesmås*	[kes'mæs]	/kesmasi/	*kesmạsi*
3b	nurse	*nås*	[næs]	/nasi/	*nạsi*
3b	salmon	*sämän*	[sæ'mæn]	/samane/	*sämäne*
3c	shirt	*söt*	[søt]	/soti/	*soti*
5	tapioca	*tapiok*	[ta'pyok]	/tapiko/	*tapiko*[26]
4, 5	whip	*uef*	[wəf]	/ufe/	*ufe*
4, 5	wharf	*uạf*	[wɔf]	/ufa/	*ufa*
4, 5	watch	*uạj*	[wɔtʃ]	/uja/	*uja*

Today, even native speakers have difficulties in trying to remember the citation form or re-develop it correctly from short forms:

Table 17: Incorrect back-formation of short forms

short form		citation form		
spelling	pronunciation	erroneous back-formation	correct form	gloss
kekes	[kə'kəs, ki'køs]	['kosi]	*kekesi*	shellfish sp.
es	[øs]	['osi]	*esu*	papaya
ser	[sør]	['sori]	*seru*	comb

25 Several (directly inherited) pronouns have also created their citation form later, e.g. *gou-a* < *gou* 'I' (< PCP **au*), *'äe-a* < *'äe* 'you (2SG)' (< PCP **koe*), *sei-a* < *sei* 'who?' (< PCP **z(a,e)i*), *tei-a* < *tei* 'where?' (also *iris-a*, *'aus-a*, *'ạmis-a*, etc. Churchward (1940:159) assumed correctly that here *-a* was suffixed later to construct citation forms analogous to the other pronouns.

26 Geraghty (1995:933) illustrated his rule of metathesis unfortunately with this example, *tapiko* > *tapiok*, whereas it had developed in exactly the opposite direction.

Some unexpectedly different final vowels in Rotuman may have been the result of erroneous back-formation. Another form (reflex of a protoform or related to forms in neighbouring languages) can equally well be re-developed out of the current short forms.

Table 18: Comparative evidence for an originally different final vowel

short form		current	via	comparable	gloss
spelling	pronunciation	citation form	rule	form	
alel	['a'lel]	*alele*	2b	PCP **ʔalelo*	'tongue'
'anås	[ʔa'næs]	*'anạsi*	3b	PCP **kanace*	'mullet'
teg	['teŋ]	*tegi*	2b	PCP **degu*	'nod'
köt	['køt]	*koti*	3a	SF *kote*	'coat'
kapkap	[kap'kap]	*kapkapo*	2	PPn **kapakapa*	'flutter'
lag	['laŋ]	*laga*	2	PEO **laŋo*	'fly (n)'
mak	['mak]	*maka*	2	Tongan *mako*	'dance, sing'

Table 19: More comparative evidence for an originally different final vowel

related or proto form	Early Rotuman	current short form	via rule	current citation form	gloss
PCP **viza*	**hisa*	*his*	2b	*hisi*	'how many?'
PCP **giza*	**kisa*	*kis*	2b	*kisi*	'when?'

Hale (1846:470) had noted 150 years ago: 'The law which prevails in the Polynesian dialects by which two consonants never occur without a vowel between them, does not apply to this tongue'. In Rotuman, consonant pairs could arise at morpheme boundaries within compound words: 'Compound words generally have two successive consonants at the point of suture, which gives the language a very un-Polynesian sound' (Hocart 1919:257).

All morphemes or elements of a compound are usually in their short form save the last one. Its shape (or 'phase') determines whether the whole word is definite or indefinite (Churchward 1940:88).[27] Consonant pairs are created by metathesis of the non-final elements: *fion-garo* 'wish' (< PPn **fina-galo*), *fat-manava* 'heart' (< PPn **fatu-manava*), *täh-roro* 'fermented coconut sauce' (< PPn **tahi* + **lolo*), *puakvại* 'tree sp.' (< PPn **puka-wai*).

Blevins (1991:3, no.11) correctly broke up *tutkai* 'thin' into *tutV* + *kai* and *fofkoa* 'snail sp.' into **fofV* + *koa*.[28] Loan words like *jeslē* 'chisel' (Blevins 1994:493) were incorporated into the Rotuman lexicon as compounds, e.g. consisting of two mora, **jese+lē*, and then regularly transformed into *jeslē*.

Contrary to this rule, Churchward's (1940:88) first rule, there is a number of polysyllabic words in Rotuman with non-final elements apparently in their citation form. These constitute about one fourth of all lexemes of more than two syllables. Churchward (1940:89, 156.6) was able to explain a few of these exceptions as contractions (*ferehiti* < *fer 'e hiti, figalelei* <

---

27 Therefore Cairns (1976:272) was not right claiming that 'citation forms have only open syllables.' Compare the examples of *takmül'ạki* < *taka* + *muli* + *-'ạki*, *a'vavhina* < *a'a-* + *vava* + *hina*, etc.

28 According to rule 2, V can here only mean a rounded back vowel [o, u].

fiongar lelei). Still others are onomatopoetic (*kararā* 'snore', *mururū* 'rush' et al.), while a third group is borrowed (*kato'aga* 'feast', *matuataliga* 'hammerhead shark', *tamamu'a* 'cheeky', etc.). But other polysyllabic loan words such as *fatmanava* (< *fatu+manava*) 'heart', *firmoto* 'tree sp.' (< PPn **fīlimoto*), *fütporo* 'football' (< *futi+poro*) adhere to the rule. Modern loan words like *rakapī* 'rugby' are not subjected strictly any more to this and other morphological rules of Rotuman.

In songs and poetry, the non-final morphemes are also transformed into their long form (Churchward's 'plenary phase') because open syllables are more convenient for singing and recitals.

Table 20: Examples of poetic forms ('plenary phase', Churchward 1940:100)

short form	citation form	poetic long form	gloss
fatmanav	*fatmanava*	*fatumanava*	heart
kạt'åk	*kạt'ạki*	*katu'ạki*	mourn
fürmaria	*fürmaria*	*furimaria*	happy

3.5 Other processes responsible for vowel change (Ablaut)

Aside from metathesis, other phonological processes were and are active in Rotuman which also produce allophones of /a/. In my opinion, they are partly analogous to short form creation, but must have started later.

3.5.1 Partial regressive assimilation

In §3.2.3 it was shown how the allophone [ɔ] was created out of *a+u* and [æ] from *a+e* and *a+i* in the course of metathesis. Later the pronunciation of the short form was copied onto the root vowel [a] in the underlying *citation form*:

citation form				short form			citation form
/áCu/	→	/ạCu/ [ɔCu]	analogous to	/ạC/ [ɔC]	←	/*áuC/	← /áCu/
/áCe/	→	/äCe/ [æCe]	analogous to	/äC/ [æC]	←	/*áeC/	← /áCe/

Even without an intervening consonant, i.e. in vowel pairs and diphthongs, /a/ assimilates partly before a following /u/ or /e/:

/áu/ → /ạu/ [ɔu, ɔw]
/áe/ → /äe/ [æe, æy]

We can formulate rules 7 and 8 accordingly (cf. Geraghty 1995:934):

Rule 7a /á/ → [ɔ] /_(C)u

/a/ in stressed syllable assimilates to an indirectly or immediately following /u/ to become [ɔ].

Rule 8 /á/ → [æ] /_(C)e[29]

When indirectly or immediately followed by /e/, /a/ in stressed syllable assimilates partly to become [æ] (exceptions see Churchward 1940:76).

---

[29] Similarly Blevins' (1994:492) rule IV: /á/ → æ /_C_0e and rule V: /a/ → æ /_e.

Besnier (1987:206) stated correctly that this is an 'assimilatory process independent of the process of metathesis', but he forgot to mention a crucial condition for the application of this process: namely that the /a/ in question must be stressed. This applies to only half of his examples under (6): /váve/ and /váe/ became *väve* ['væve] and *väe* [væy], but /faéga/ is not stressed on the /a/ and only turned into *fäega* [fæ'eŋa] because it probably derived from /**fáe* + *ga*/ (cf. *päré-ga* 'protection' < *päre* 'guard'). The fourth example is *mané'a* 'play', which is not pronounced *[mæ'ne?a]; in the short form *maneá'* there is a variant [mæ'nya?].[30]

3.5.2 Backing

In words ending in /áCi/, the pronunciation of the root vowel in the short form (*-åC* [æC] < *-aiC < -áCi*) is not copied to the long form, but the stressed /á/ is backed to [ɔ] by a following /i/ just as if it were followed by the other high vowel [u]:

/áCi/ → ạCi [ɔCi] analogous to /áCu/ → ạCu [ɔCu]

Similarly in diphthongs and vowel pairs:

/ai/ → ại [ɔi, ɔy] analogous to /áu/ → ạu [ɔu, ɔw]

Thus we can write another rule:

Rule 7b /á/ → [ɔ] /_(C)i

and merge **rule 7a** (backing of /a/ to [ɔ] before /u/)[31] and **7b**:

Rule 7 /á/ → [ɔ] /_(C)V_{high} or /á/ → [ɔ] /_(C){i,u}

Expressed in binary features:

[ɔ]	←	adjacent to *i*	*a*	adjacent to *u*	→	[ɔ]
		[i]		[u]		
[+back]		[–back]	[-back]	[+back]		[+back]
[+rounded]		[–rounded]	[–rounded]	[+rounded]		[+rounded]
[–high]		[+high]	[–high]	[+high]		[–high]

By keeping Churchward's spelling it is possible to derive the citation form from the short form, which is not possible when following phonetic transcription: *tåk* [tæk] < *tạki* vs. *täk* [tæk] < *täke*.

While the shift from /a/ (low central vowel) before /u/ (high back vowel) to [ɔ] (low back vowel) is phonologically natural and constitutes a partial assimilation, the contrary can be said for /a/ before /i/ (high front vowel) becoming [ɔ]. This sound change prompts me to conclude that metathesis occurred first and the other phonological processes afterwards (according to the **rules 7 and 8**). The derivation of the ending [æC] via *[aiC] from the long form /áCi/ seems more natural than a development via *[ɔiC] and [ɔCi] from /áCi/. Later,

---

30 In some cases, the ablaut also crops up in unstressed syllables: *hạmúa, fä'ére*, et al.

31 Similar to Blevins'(1994:492) rule III: /a/ → ɔ /_C_0{i,u}.

after metathesis had become well-established in Rotuman, **rule 7b** created an ablaut in the citation form from /áCi/ to [ɔCi], parallel to the one of /áCu/ to [ɔCu].[32]

While the ablaut of /a/ in **rule 7** until now only occurred when a stressed /a/ was followed by high vowels (rising diphthong), the same ablaut also appears in the short form if an originally unstressed /a/ receives the stress after metathesis and follows immediately after a high vowel (falling diphthong):

citation form		short form			example		
		phonemic		phonetic			
/úCa/	→	/uáC/	→	[wɔC]	*hula*	→	*huạl*
/íCa/	→	/iáC/	→	[yɔC]	*pịja*	→	*piạj*

These rules can be written as follows:

Rule 9a /á/ → [ɔ] / u_C

Rule 9b /á/ → [ɔ] / i_C

and summarised into one:

Rule 9 /á/ → [ɔ] / V_{high}_C or /á/ → [ɔ] / {i,u}_C

The question is how /úaC/ changed into contemporary [wɔ́C]? Two different ways seem possible:

Table 21: Possible paths of Rotuman ablaut

Ending of complete phase	rule	outcome	rule	outcome	rule	outcome	rule	Ending of surface form in incomplete phase
/úCa/	1 →	/úaC/	4 →	/uáC/	5 →	/waC/	9a	[wɔC]
					9a →	[uɔ́C]	5	

After metathesis and accent shift, either /a/ assimilated partly to the preceding /u/ and then the unstressed high vowel was weakened to a semivowel (rule 5) or vice versa. Since high vowels are at the centre of so many morphophonemic changes in Rotuman, I assume that /u/ first caused /á/ to be backed (rule 9a) before becoming itself weakened to [w] (rule 5). Later the rule was applied to the other high vowel as well.

These forms are in contrast with words with vowel pairs in their root syllable whose final (non-high) vowel is deleted in the short form. A minimal pair is for example:

	short form					citation form	gloss
	surface form			underlying form			
	puạk	[pwɔk]	←	**/puák/**	←	/púka/	'vine, rope'
vs.	*puak*	[pwak]	←	***/puaák/**	←	/puáka/	'pig'

Perhaps the formerly geminate vowel prevented /a/ in the bottom line from being backed.

[32] Anttila (1989:63) separated Churchward's *ạ* into /ạ/ from a/_i and /ɔ/ from a/_u and thought both were 'unambiguously analysable as sequences of two morphophonemes.' Their origin is certainly different, but phonetically both Umlauts are identical and were written by Churchward the same way (*ạ*).

3.5.3 e-forms

A final (and **unstressed**) /a/ in the long form of words with a stressed high vowel in penultimate syllable (root syllable) is often raised and fronted to [e]. Churchward (1940:14, 87ff.) called this *e*-form 'narrow version'.[33] 'The *a*-form conveys the idea of bigness or plurality, while the *e*-form conveys the idea of smallness or singularity' (Churchward 1940:87). In the incomplete phase, both *a*- and *e*-forms occur if there is no intervening consonant (other than /h ʔ/); otherwise, only the *a*-form occurs:

Table 22: *e*-forms

complete phase		incomplete phase		gloss
a-form	*e*-form	*a*-form	*e*-form	
pija	*pije*	*piạj*	–	'rat'
huga	*huge*	*huạg*	–	'mind'
i'a	*i'e*	*iạ'*	*ie'*	'fish'
tupu'a	*tupu'e*	*tupuạ'*	*tupue'*	'immortal man'
keia	*keie*	*keia*	*keie*	'poor'
hanua	*hanue*	*hanua*	*hanue*	'land'

Rule 10 /a/ $\dashrightarrow$ [e] /V_{high}(C)_

consisting of

Rule 10a /a/ $\longrightarrow$ [e] / i(C)_

Rule 10b /a/ $\dashrightarrow$ [e] / u(C)_

I think this rule started out as partial progressive assimilation (rule 10a), similar to the Tongan regressive case (see below). Later it was extended to instances of the other high vowel (rule 10b), although it cannot be called assimilation any more, quite the contrary. This extension is parallel to the rule 7 of backing and rounding /a/ to [ɔ] which started off as a partial regressive assimilation (rule 7a) before it was applied to instances of /a/ before the other high vowel ([i]) as well (rule 7b). Chronologically, rules 10b and 7b may inhabit the same slot.

Until now, most phonological changes concerned stressed /a/, but here it is the unstressed /a/ in final position. Raising or partial assimilation of /a/ to preceding high vowels seems comparable to vowel raising in Tongan, which also affects /a/ only and mainly results in partial regressive assimilation to a following high vowel: Proto Tongic **a* became Tongan [e] next to front vowels and [o] next to back vowels (Pawley 1966:57; Biggs 1971:483; Tovey 1993). Tovey counted 75 cases of raising, of which only 16% constituted progressive assimilation, the remainder being regressive (even though he confused the terms). In Tongan as well as Rotuman, it is not relevant whether there is an intervening consonant or not. 'Unstressed /a/ is often raised in the environment of a non-low vowel, a rule which yields morphophonemic alternation in reduplications, e.g. *pelepela, monumanu, hinehina* and

[33] In §3.2 I mentioned that Besnier (1987:208) had been mistaken in listing *tife* and *huge* as examples of roots with final /e/. Their citation form ends in /a/, but the forms with an *e*-ablaut are used much more often on the surface. The same applies to his example No. 10 on p.208: When he took *tife* and *huge* as the underlying roots of *tifa* and *huga*, why not do the same to the other three examples, *pije, puke, kuruge*?

compounds, e.g. *ta'e* 'excrement', *te'elango* 'fly's excrement, wax, candle' ' (Geraghty 1995:939). Compare some examples of Tongan ablaut:

Table 23: Loan words in Rotuman without Tongan ablaut o/e < *a

Rotuman	Tongan	East 'Uvean	Samoan	gloss
kạrkạru	*kolukalu*	*kalukalu*	*'alu'alu*	'jellyfish'
matua'rau	*motulau*	–	*matūlau*	'fish sp. (*Parupeneus*)'
mạunu	*mounu*	*mounu*	*māunu*	'bait'
telua	*luo*	*luo*	*lua*	'hole (in ground)'
'ạitu	*'eitu/matupu'a*	–	*aitu*	'God'

Since Rotuman has borrowed heavily from Tongan and other West Polynesian languages, it might be conceivable that some morphophonemic changes were also due to Tongan influence. But which change was triggered by the Tongan ones?

Table 24: Comparison of Tongan and Rotuman vowel changes

Rule	Tongan	Rotuman	Rule
T1	/a/ → [o] / _(C){o,u}[34]	/á/ → [ɔ] / _(C)u	R7a
T2	/a/ → [e] / _(C){e,i}	/á/ → [æ] / _(C)e	R8
		/a/ → [e] / i(C)_	R10a
T3	devoicing of final vowel	elision of final vowel	R2

Rule	changes	environment/conditions	assimilation
T1	backing + raising + rounding	before a back rounded vowel	regressive
R7a	backing + rounding	before a high back rounded vowel	regressive
T2	fronting + raising	before a front vowel	regressive
R8	fronting	before a mid-high vowel	regressive
R10a	fronting + raising	after a high front vowel	progressive
T3	devoicing of high vowel[35]	if the high vowels (1) are short and unstressed, (2) follow a voiceless consonant, (3) are situated in final position of a morpheme and (4a) stand at the end of an utterance or (4b) precede a voiceless consonant	lenition
R2	elision of final vowel	V_2 is deleted if it is identical with V_1 or if V_1 is not further back than V_2 and if V_2 is not lower than V_1	progressive

I think that R7a and R8 are simply extensions of vowel changes from the short form to the less frequently used long form. Rotuman ablaut formation is not comparable to Tongan ablaut but constitutes an independent parallel development. Deletion of final vowel (R2) is

---

34 'In Tongan *a* unaccented becomes *o* when the next vowel is *u*, though a consonant may intervene. (I first saw this rule formulated by A.M. Hocart "Man", vol. XV, p.149 note). This change is still going on: *tanu* 'bury', has a passive *tanu-mia*, which is frequently pronounced *tonu-mia*' (Collocott 1922:187).

35 The low vowel /a/ is devoiced under conditions 1, 3, 4 (though only following /h/), but oddly not the mid vowels.

part of a larger process (short-form derivation) which started several centuries ago whereas devoicing of final vowels in Tongan seems to be a more recent development.

4 Morphological aspects of suffixation and metathesis

Rotuman affixes can be distinguished in several ways, among them being whether they are directly inherited or borrowed and whether they are still productive or fossilised. Suffixes especially can be further grouped into those which are attached to the short form or incomplete phase of a root word and others which cling on to the complete phase or citation form of a root. The last case violates a fundamental rule of Rotuman morphology, that all morphemes in a compound must be in their incomplete phase except the last one. That is why suffixes have to be included in a discussion of Rotuman short-form derivation since some of them can be attached to the complete phase of a root.

Table 25:. Suffixation and stress placement

	root word		gloss	productive ones: suffix type	added to: citation form	added to: short form	gloss
	citation form	short form					
accent remaining unmoved after suffixation	*húla*	*huạ́l*	'month'	INDEF	*húle-t*		'a/one month'
	hạ́ni	*hản*	'woman'	INTERROG	*hạ́ni-s*		'which woman?'
	tá'a	*tá'*	'that (2P)'	ORN	*tá'a-g*		'that'
	hó'a	*hoá'*	'carry'	DIR	*hó'a-me*		'bring here'
	máka	*mák*	'sing'	PRON	*(iris) máka-risa*		'(they) are singing'
	fúti	*fút*	'pull'	TR	*fúti-a*		'pull'
	fúti	*fút*	'pull'	(DIR+)PERF	*fúti-me-a*		'have pulled here'
	fóra	*foár*	'tell'	PERF		*foár-'ia*	'already told'
accent shifting to penult of suffixed form	*móse*	*mös*	'sleep'	NOM	*mosé-ga*		'place to sleep'
	móse	*mös*	'sleep'	MODIF		*mös-'ạ́ki*	'put to sleep'
				fossilised ones			
accent remaining unmoved after suffixation	*mára*	*mar*	'suffer'	PERF		*már-tia*	'have suffered'
accent shifting to penult of suffixed form	*múri*	*mür*	'end (n)'		*muriá'a*		'end (vi)'
	púlu	*pul*	'gum'	TR	*pulúfi*		'stick to'

Why is that so? It is remarkable that the only (productive) suffixes which can be attached to the short form of a root are borrowed. The suffixes on the left are unlikely to have been borrowed (though *-(á)ga* is ambiguous) whereas *-'ạ́ki* '(relational, causative, instrumental,

prepositional, durative, moderative)' and *-'ia* '(resultative)'[36] are obvious borrowings from Tongan, even including fine nuances of meaning (see also Biggs 1965:414). Consequently one explanation of the different behaviour of Rotuman roots at suffixation is that suffixes attached to the complete phase are older than the ones connected to the incomplete phase of nouns and verbs. They were productive already at a time when there were no short forms and metathesis (yet) in Rotuman. After this fashion caught on, bases before the 'old' suffixes were not subjected to metathesis whereas the two 'new' suffixes were attached to the incomplete phase of Rotuman roots — which is understandable because the short form of a word is the much more frequently used surface form.

Incidentally, all suffixes on the bottom left in the above table which cause the accent to shift to the penultimate syllable of the suffixed word form their incomplete phase regularly. But many of the suffixes attached to the complete phase of a root (top left column of Table 25) do not form their short forms regularly; often, a final vowel is simply elided which would be metathesised elsewhere. It seems to me as if some present citation forms were later creations:

Table 26: Regular and irregular short forms of suffixes

rule no.	short form	current citation form	expected form	gloss
–	*máka-ris*	*máka-risa*	**maka-rias*	'(they) are dancing'
–	*ófi-en*	*ófi-ena*	**ofi-ean*	'finished'
–	*há'u-m*	*há'u-me*	**ha'üm*	'arrive here'
1+4	*mose-ág*	*mosé-ga*		'place to sleep'
1+3b	*fäeag-'åk*	*fäeag-'ą́ki*		'tell'
1+3a	*pülüf*	*pulúfi*		'stick (v)'

Is this sufficient evidence that the two 'phases' (citation form and short form) had already existed before the Rotumans started to borrow words and morphological elements from Tongan?

5 The stress factor

Table 25 also showed that only two of the currently productive suffixes cause the accent to shift from the penultimate syllable of the root to the penult of the suffixed word. Only the nominalising and the modificatory suffixes do so. Why these two? Perhaps it is because they carry more weight than the others: (1) they consist of at least one syllable even in their short forms, and (2) one (the modificatory suffix) changes the meaning considerably and the other (the nominalising suffix) causes a change in word class.

5.1 Word accent

I agree with Biggs (1959:24) and Blevins (1991:1, 1994:497) that vowel length is not phonemic in Rotuman, but accent is. Just as in Fijian (Schütz 1985:54) and Tongan, there is no long vowel 'in a position where it would receive the stress if it were a normal vowel'

36 Compare *ia* in East 'Uvean.

'(Churchward 1953:10). It was shown earlier (see last paragraph of §3.5.1; cf. Churchward 1940:75) that the accent is responsible whether *vowel assimilation* occurs in Rotuman or not.[37]

The difference in stress between citation forms and short forms is striking (Churchward 1940:75; Hocart n.d. [1913]:4897). 'Long forms are always stressed on the penultimate syllable, short forms on the last syllable. Since forms differing only in the position of stress occur, it is considered to be phonemic, e.g. */fáfa/* 'await' vs. */fafá/* 'challenge' (Biggs 1959:24, also 1965:388). In a few cases minimal pairs (of homographs) can be found, i.e. two forms which differ only in their stress pattern (without additional lengthening), e.g. *'ío-m* 'look here!' (< *'ío* + directional suffix *-me*) vs. *'ióm* 'drink' (< *'imo* + metathesis).

Churchward (1940:75) saw the causal chain exactly the other way round: 'Words ending in a long vowel, however, take the accent on the final syllable. Except ... when a suffix is added the accent remains where it was' — although then it will not be on the penultimate syllable any more.

The **suffixes** can be grouped into those that let the accent remain on the penult of the root and others which demand it to shift to the penult of the suffixed word. For example:

Table 27: Stress placement in affixed forms

	root		gloss	affix	affixed form		gloss
	phonemic	phonetic			phonemic	phonetic	
accent shifting to penult of suffixed form	/mose/	['mose]	'sleep'	NOM	/mose-ga/	[mo'seŋa]	'bed'
	/hanisi/	[ha'nisi]	'love'	RECIP	/hai-hanisi-ga/	[hɔihani'siŋa]	'love one another'
	/fere/	['fere]	'fly'	MODIF	/fer-ʔaki/	[fer'ʔɔki]	'fly with'
accent remaining unmoved after suffixation	/ʔíhi/	['ʔihi]	'invite'	DIR	/ʔíhi-me/	['ʔihime]	'invite here'
	/kéle/	['kele]	'see'	PRON	/kéle-na/	['kelena]	'look at'
	/ravá/	[ra'va]	'defeated'	TR	/ravá-tia/	[ra'vatia]	'have been defeated'

With some high-frequency suffixes (the nominalising, reciprocal and modificatory suffixes), stress shifts to remain on the penult of the suffixed form. But the accent remains unmoved on the root in spite of pronominal, directional or transitive suffixation. It is remarkable that the accent only shifts right to the (new) penultimate syllable of the now compounded word after adding suffixes which were certainly (in the case of *-'ạki*)[38] or possibly (in the case of *-(á)ga*, *-á'a*) borrowed from Tongan. My explanation for this is that these suffixes were not borrowed in isolation but first as parts of compound loan words and regarded as a stress unit in Rotuman; therefore stress was placed on the penult. Once they had borrowed a sufficient number of root words and compounded forms, the Rotumans could identify the suffixes as separate morphemes and use them productively themselves (cf. Ross' observation that it is much more likely for a language to borrow content words than bound morphemes).

---

37 In rare cases unstressed /a/ is also assimilated: *fạvíri* (< /faviri/), *hạmúa* (< /hamua/), *kạría* (< /karia/), *tạíri* (< /tairi/), etc. (see also Churchward 1940:76).

38 Compare Fijian *-yaki*, which is the only one of the disyllabic transitive suffixes which attracts the main accent (Arms 1974:95).

In the Tongan source words themselves 'the accent moves in suffixed forms to the right in order to remain on the penult. Similarly prepositions with two vowels build a stress unit together with the following article *-e* and let the accent move. The same applies if aspect markers like *'óku* '(present)', *ná'a* '(past)', *kúo* '(perfect)' and the conjunction *péa* 'and' are followed by SG- and DL-markers' (Feldman 1978:134ff.).

There would be a nice and clean-cut separation, as depicted in the table above, if directly inherited suffixes like the directional markers or pronominal elements would only be affixed to the citation form and if the accent would remain unmoved after suffixation, whereas in all other cases borrowed suffixes were always attached to short forms. Alas, the accent before *-'ia* does not shift although this suffix looks like a loan; also alimental pronominal suffixes can be attached to the short form in rare cases.[39]

5.2 Stress shift and Tongan definitive accent

As mentioned above, stress falls regularly on the penultimate syllable of words in their complete phase but on the final syllable in the incomplete phase. However, both rules can be reconciled because a new stress group is created: the short form usually does not stand on its own but must be followed by a dependent morpheme (article, adjective, adverb or the like). A noun phrase can be changed as follows:

citation form		gloss	short form + article		gloss
mori	['mori]	'orange(s)'	*mör ta*	['mørta]	'the oranges (specific ones)'
ṭaku	['tɔku]	'doctor(s)'	*ṭak ta*[40]	['tɔkta]	'the doctor'

The newly created (larger) accent unit is once again stressed on the penult. It is reflected in the indigenous way of spelling in that a noun in its incomplete phase is often written together with the following definite article *ta*: *Itu'muta* (< *Itu' mut ta*) 'the cut-off district', *ö'fata* (< *ö'fā ta*) 'the father', *hånta* 'the woman', *mumueta* 'the first', etc.

Aside from the large proportion of Tongan loan words in the Rotuman lexicon, there are further similarities between both languages, one of which is the accent shift and the derivation of the short form compared to the 'definitive accent' in Tongan.

To make a noun phrase 'definite or generic' in Rotuman, the primary stress shifts from the penultimate to the final vowel of the last word (Churchward 1953:6–10, 12, 25–27, 268–289). In Rotuman the long or citation form is semantically 'definite, specific or generic', whereas the short form signifies something indefinite or unspecific. On the other hand in Tongan, a noun phrase with regular stress on the penult is unspecific and made specific by putting a 'definitive accent' on the final syllable.

---

39 They are identical with the short form of the alimental possessive pronouns. For example: *mák-'e-n*, *hát-'e-u*, *nóh-'e-ris*, etc. in contrast to the continuous verb forms *mák-ana*, *háta-u*, *nóho-a-ris*, etc. As with the alimental possessives, the meaning of these compounded intransitive verb forms is something like 'have a turn to do X' or 'do one's share of X' (where X stands for the literal meaning of the root of the verb). The present generation of Rotumans (Ravai Shaw, pers. comm. 3/2000) does not see a difference in meaning between both verb forms with pronominal suffixes and considers the second form (such as *mák-'en*) to be a variant of the supposedly correct form (here *máka-na*) or regards the suffix as the alimental possessive pronoun without the possessed noun and writes it separated from the verb: *ia mak 'en* 'she danced (her part)', 'it was her turn to dance'.

40 The last syllable of the English source word was regarded in Rotuman as the postposed definite article.

There is also some formal overlap if the noun phrase contains several words: 'In both languages, moreover, the general rule is that when a nounal or pronominal group which is definite is extended by the addition of one or more qualifying words, the definitive accent (or its Rotuman counterpart) is carried on to the new end of the group. This ... is the most interesting and perhaps the most vital feature of the whole phenomenon' (Churchward 1953:269ff.).

The *functions* of the accent shift in both languages are very similar: 'The functions of the stress on the final syllable in Tongan, which are, broadly speaking, emphatic and definite, by contrast with stress on the penult, which is non-emphatic and indefinite, are analogous to the uses of the complete and incomplete phases of Rotuman respectively' (Milner 1971:418).

Besnier (1987:204) described definitive accent in Tongan as the 'phonemic gemination of the last vowel ... of an NP.' In Rotuman, if a vowel in final syllable is stressed, it is automatically lengthened. I cannot see the difference when Hovdhaugen calls a comparable phenomenon in Samoan — the 'locative accent' (Condax 1989) — *vowel lengthening*. The Samoan locative accent 'has nothing to do with the definitive accent we find in Tongan and Tokelauan. But the definitive accent in those languages is not so much an accent on words as on phrases. This type of accent can also sporadically be heard in Samoan, especially on Savai'i. It may be due to influence from Tongan or Tokelauan and it is a very marginal phenomenon in Samoan' (Hovdhaugen 1992:284, fn. 4).

A definitive accent is also productive in Niuafo'ou[41] and East 'Uvea which were both heavily influenced by Tongan, as well as Tokelau and East Futunan. The accent shift in these languages is said to be 'very old', whatever time frame this might mean, and cannot result solely from Tongan influence, since the languages in question had participated in this innovation already in its very early stage of development (Tsukamoto 1994:54).

Clark (1974) and Tsukamoto (1994:49ff.) saw the *origin* of the Tongan definitive accent in the deletion of the PPn deictic clitic **ra*. As part of a regular sound change, PPn **r* became Ø, and the unstressed vowel *a* assimilated as usual in Tongan to the preceding vowel. Similar scenarios apply to Niuafo'ou, East Futunan and East 'Uvean (Tsukamoto 1994:54).

ROTUMAN

**húla ta*	> **húal ta*	> *huál ta*	'the moon'	
**móri he*	> **móir he*	> *mőr he*	'a orange'	

TONGAN

**fále ra*	> **falé ra*[42]	> **falé a*	> **falée*	> *falé*	'that house',
**fále na*				> *falé na*	'that house over there'

Besnier (1987:204) saw similarities as well: 'The specific forms of words in both languages are posited as having arisen historically as the result of suffixal deictic particles that dropped after having attracted the strictly penultimate stress in both languages to the last vowel of words marked for specificity. Residues of such particles are found in both

---

41 'The Niua Fo'ouans raise the voice on final words and syllables, giving the language a cadence not unlike that of Samoan' (Collocott 1922:188).

42 Tsukamoto reconstructed PTo **ra* 'that one over there (deictic element, correlate to the 3rd P).' Clark derived it historically from a third deictic suffix **-e* beside *-ni* 'here' and *'-na* 'there'.

rather a postposed definite article; it might have derived from the shortened form of a demonstrative (*ta'a* 'that (correlate to the 2nd P)', *täe* 'that one over there (correlate to the 3rd P)').

It seems also that Rotuman and the West Polynesian languages developed the same tendency for an accent shift to the final syllable to express specificity soon after their separation from Central Pacific. Short-form derivation in Rotuman was not caused by Tongan influence but constitutes a regional speciality ('area phenomenon')[43] or a coincidental parallelism just like e.g. the sound change from **t* > *k* in Samoan and Hawaiian.

Comparable is the partly independent development of polite register in Western Polynesia according to Milner (1961:300): many polite expressions in Tongan have no equivalent in Samoan and vice-versa. 'From this evidence and from our knowledge of the existence of a small number of terms suitable for high chiefs or royalty and found in Samoa, Tonga, Wallis, and Futuna, it seems possible to draw the conclusion that in this area the majority of the terms of respect have evolved in each community since it became separately established in its present habitat, but that a few words especially those referring to high chiefs and royalty were perhaps in use before the Western Polynesians became separated.'

6 Metathesis and Rotuma's language history

When did Rotuman split up from the Central Pacific dialect chain? The earliest archaeological finds from Rotuma were dated to about 1700 years ago. Initially Pawley (1979:40) thought that Rotuman had separated from the Central Pacific dialects in Western Fiji after Polynesian had split from Tokalau-Fijian, because it shares most innovations with all other regions of Fiji. 'The separation of Rotuman must have preceded the Proto Fijian period. To be more exact, it must have happened before the spread across Fiji of those particular innovations that are not found in Rotuman ... it is characteristic of the whole Fijian group, with sporadic exceptions in Vanualevu, that the PCP rising vowel clusters **ae* and **ao* reduce to a single vowel, which may be *a*, *e*, or *o*, depending on stress placement in the original form. As Rotuman preserves the original clusters the spread of this innovation across Fijian can be dated after the divergence of Rotuman' (Pawley 1996:111).

The first settlers on Rotuma may have sustained contact with their former home for some time,[44] but then a long period of isolation or separate development ensued. I assume that the major changes in Rotuman started during this phase. Similarly Ross (pers. comm.) and Irwin (1992:174): 'Islands began to diverge faster in isolation, from the time effective communication between them slowed or ended, rather than when contact between them began'. Especially, the sound changes of Rotuman in this period of comparative isolation were much more drastic and numerous than the ones [which ones?] in the following centuries when contact with Polynesians and perhaps also Fijians was more intense, and also compared to changes in the other Central Pacific languages which had never lost contact with their neighbouring dialects and languages.

The trigger to some changes in the language might have come already at the time when the fore-runner of Rotuman had not yet completely separated from the other Central Pacific

---

43 In the Lau dialect of Fijian the final syllable is lengthened to express respect (Hocart 1929:49).

44 'Rotuman initially maintained its contacts with the geographically closest western Fijian dialects' (Green 1981:149).

dialects in Western Vanualevu. But its most distinguishing features Rotuman acquired after the separation from them: the far-reaching sound changes, the creation of short forms of all content words, lexical and semantic changes, syntactic changes such as the postposition of the article. The idiosyncratic development during the long period of isolation led to the fact that 'the Rotuman language is totally unintelligible to speakers of the Fijian and Polynesian languages, to which it is genetically most closely related' (Geraghty 1984:34).

The changes were not triggered by Polynesian influence. The creation of short forms can be seen as a parallel development to the independent genesis of the definitive accent in a few western Polynesian languages. The tendency to shift the stress from the penultimate to the final syllable to express emphasis or definiteness seems to have been dormant in the common ancestor language, even when no trace of it can be found in contemporary Fijian speech. The accent shift in west Polynesian languages is probably not due to Tongan influence (cf. Geraghty 1984:34 and Tsukamoto 1994:54). Creating short forms in Rotuman became a fashion before the first intensive contact with Polynesians around the 13th century.

I posit the following order of events:

1.	Early Rotuman	all lexemes in complete phase (open syllables, no consonant clusters)
2.	Middle Rotuman	the fashion of incomplete phase derivation developing parallel to the definitive accent in Western Polynesian languages
3.	Metathetic Rotuman	the incomplete phase became the more frequently used one of the two forms
4.	High Rotuman	large influx of Tongan and other Polynesian borrowings, even grammatical morphemes such as affixes; new suffixes like *-'ạki* and *-'ia* were attached to the short form since that was the most commonly used one.

References

Anttila, Raimo, 1989, *Historical and comparative linguistics.* Amsterdam and Philadelphia: Benjamins.

Arms, David, 1974, Transitivity in Standard Fijian. PhD dissertation, University of Michigan.

Bennett, George, 1831, A recent visit to several of the Polynesian islands. *United Services Journal* 33/2:198–202, 473–482.

Besnier, Niko, 1987, An autosegmental approach to metathesis in Rotuman. *Lingua* 73:201–223

Biggs, Bruce, 1959, Rotuman vowels and their history. *Te Reo* 2:24–26.

1965, Direct and indirect inheritance in Rotuman. *Lingua* 14:383–415.

1971, The languages of Polynesia. In Thomas A. Sebeok, ed. *Current trends in linguistics*, vol. 8: *Linguistics in Oceania*, 466–505. The Hague; Paris: Mouton.

Blevins, Juliette, 1991, The separation of vowels and consonants in Rotuman phonology. Paper presented to the Sixth International Conference on Austronesian Linguistics, Honolulu.

1994, The bimoraic foot in Rotuman phonology and morphology. *Oceanic Linguistics* 33:491–516.

Cairns, Charles E., 1976, Universal properties of umlaut and vowel coalescence rules: implications for Rotuman phonology. In Alphonse Juilland, ed. *Linguistic studies offered to Joseph Greenberg on the occasion of his sixtieth birthday*, vol. 2: *Phonology*, 271–283. Saratoga: Anima Libri.

Churchward, Clerk Maxwell, 1929, Definiteness and indefiniteness in Rotuman. *Journal of the Polynesian Society* 38/4:281–284.

1940, *Rotuman grammar and dictionary*. Sydney: Australasian Medical Publishing Co.

1953, *Tongan grammar*. Oxford: Oxford University Press.

Clark, Ross, 1974, On the origin of the Tongan definitive accent. *Journal of the Polynesian Society* 83:103–109.

Collocott, Ernest Edgar Uyvyan, 1922, The speech of Niuafo'ou. *Journal of the Polynesian Society*, 31:185–189.

Condax, Iovanna D., 1989, Tongan definitive accent. *Journal of the Polynesian Society* 98/4:425–450.

1990, Locative accent in Samoan. *Oceanic Linguistics* 29/1:37–48.

Feldman, Harry, 1978, Some notes on Tongan phonology. *Oceanic Linguistics* 17/2:133–139

Geraghty, Paul, 1984, Language policy in Fiji and Rotuma. In George B. Milner, David G. Arms and Paul Geraghty, eds *Duivosavosa/Fiji's languages: their use and their future.* Fiji Museum Bulletin No. 8.

1995, Rotuman. In D.T. Tryon, ed. *Comparative Austronesian dictionary*, part 1, fascicle 2:931–936. Berlin: Mouton de Gruyter. 4 vols.

Grace, George W., 1959, *The position of the Polynesian languages within the Austronesian (Malayo-Polynesian) language family.* Supplement to *International Journal of American Linguistics* 25/3. [Reprinted 1974. Chicago: University of Chicago Press.]

Green, Roger C., 1981, Location of the Polynesian homeland: a continuing problem. In Jim Hollyman and Andrew K. Pawley, eds *Studies in Pacific languages and cultures in honour of Bruce Biggs*, 133–158. Auckland: Linguistic Society of New Zealand.

Hale, Horatio, 1846, *United States exploring expedition ... 1838–1842 ...*, vol. 6: *Ethnography and philology*. Philadelphia: Lea & Blanchard.

n.d. [c. 1840], Field notes from Rotuma. MS.

Hocart, Arthur Maurice, n.d. [1913], Field notes from Rotuma. MS. [In Alexander Turnbull Library, Wellington, New Zealand.]

1919, Notes on Rotuman grammar. *Journal of the Royal Anthropological Institute* XLIX (July–December): 252–264.

1929, Lau Islands, Fiji. *BP Bishop Museum Bulletin.*

Hovdhaugen, Even, 1992, Phonetic vowel length in Samoan. *Oceanic Linguistics* 31/2:284.

Irwin, Geoffrey, 1992, *The prehistoric exploration and colonisation of the Pacific.* Cambridge: Cambridge University Press.

Milner, George B., 1961, The Samoan vocabulary of respect. *Journal of the Royal Anthropological Institute* 91:296–317.

1971, Fijian and Rotuman. In Thomas A. Sebeok, ed. *Current trends in linguistics,* vol. 8: *The languages of Oceania*, 397–425. The Hague: Mouton.

Pawley, Andrew K., 1966, Polynesian languages: a subgrouping based on shared innovations in morphology. *Journal of the Polynesian Society* 75:39–64.

1979, *New evidence on the position of Rotuman.* Auckland: University of Auckland, Department of Anthropology, Working Paper No. 56.

1996, On the position of Rotuman. In Bernd Nothofer, ed. *Festschrift in Honor of Isidore Dyen*, 85–119. Hamburg: Abera Network.

Schütz, Albert J., 1985, *The Fijian language.* Honolulu: University of Hawai'i Press.

Tovey, Peter, 1993, Ablaut in Tongan. Paper presented to the First International Conference on Oceanic Linguistics, Port Vila.

Tsukamoto, Akihisa, 1994, *Forschungen über die Sprachen der Inseln zwischen Tonga und Saamoa.* Münster and Hamburg: Lit.

Vamarasi, Marit Kana, 1991, Stress, long vowels and diphthongs in Rotuman. Paper presented to the Sixth International Conference on Austronesian Linguistics, Honolulu.

Walsh, A.C., 1982, *Migration, urbanization and development in South Pacific countries. Comparative study on migration, urbanization and development in the ESCAP region.* New York: United Nations.

Woodhall, Derek, 1987, *Geology of Rotuma.* Suva: Ministry of Lands.

Fijian reflexes of the Proto Austronesian phonemes

JOHN U. WOLFF

1 Background

Fijian is a chain of closely related languages, spoken in the Fiji Islands, numbering some thirty or more. The standard language is based on the dialect spoken on Bau and has influenced the other languages of Fiji. In this paper we will cite principally forms of Wayan, the Fijian spoken on Waya island, for the reason that a substantial dictionary has been prepared in manuscript form by Pawley and Sayaba (n.d.), which is more accurate and reliable than the small dictionaries of Bauan which are available to English speakers or those who have not done first-hand field work. In this paper we cite forms from Bauan when no Wayan cognates are listed in our sources.

The Fijian languages have been given a comparative study in some detail by Geraghty (1983). This reliable and entirely sensible work has provided the clues to solving many of the mysteries as to the historical origins of irregularities which the data presented. Indeed, one of the reasons for focusing on the Proto Austronesian (PAn) origins of the current Fijian forms is that a great deal of information is available on the vocabulary of the Fijian languages and the relationships of the forms to each other, a factor which renders these languages among the most important for providing data for a reconstruction the Proto Austronesian phonology and vocabulary.

A theme which runs through Geraghty's work, which is crucial to our understanding of the history of the forms which we study here, is that the thirty or more Fijian languages, together with Polynesian and perhaps also Rotuman to the east of Fiji, form a chain of dialects such that innovations have spread one at a time over all or part of the area. Some innovations are found only in the western areas, some are found only in the eastern Fijian dialects, and indeed a number of innovations have affected only eastern Fijian dialects and Polynesian languages, so that in fact the eastern Fijian dialects have many features in common with Polynesian as opposed to the western dialects (Geraghty 1983:354–382). If we were to list all the Fijian reflexes for each PAn etymon we would have a fairly sizeable number of doublets. In this study the doublets are mentioned only episodically, where the mention of these in some way clarifies how a certain change is to be viewed. The dialectal diversity and

John Lynch, ed. *Issues in Austronesian historical phonology*, 209–227.
Canberra: Pacific Linguistics, 2003.

the way in which the forms spread over the area on an individual basis account for a certain number of the double reflexes of single PAn phonemes which are manifested in the data. A more important source of the double reflexes, however, is the consonant gradation (that is, certain morphophonemic processes) which characterise the whole of the Oceanic group (see §2.1).

2 Changes which characterise Fijian in general

Chart 1 shows the Fijian consonant system as found in Wayan and Bauan. We present the transcription with the phonetic realisation in brackets in those cases where the realisation diverges from the sounds normally symbolised by the letter or letter sequences.

Chart 1: Wayan and Bauan consonants[1]

plain	v	t	c [ð]	r	k	kw
prenasal	b [mb]	d [nd]	s	dr [ndr]	q [ŋg]	qw [ŋg^{w}]
nasal	m	n			g [ŋ]	gw [ŋw]
semi-vowel, liquid	w	l	y			

2.1 Consonantal gradation

The term *consonant gradation* refers to a process which characterises the whole Oceanic group whereby the stop consonants and **c* may in initial and medial position manifest double reflexes in the same etymon. One of the reflexes is derived from a prenasalised consonant and the other from the consonant without prenasalisation. The same etymon may have a prenasalised reflex in one language and a plain reflex in another, or even in the same language the same etymon may be reflected in two lexemes, one of which contains the prenasalised grade consonant and another of which contains the plain grade consonant. In none of the Oceanic languages are the prenasalised grades reflected with prenasalised consonants in all positions, and in Fijian the prenasalised grade is reflected by a prenasal only in the reflexes of the velar stops **k* and **g*. Chart 2 shows the oral and the prenasalised grades of reflexes in Fijian. (The details of the developments are discussed in §4.1ff.)

Chart 2: Consonant gradation in Fijian

PAn	*b, *p	*t	*d	*c, *j	*k	*g
Fijian oral	v	t	r	c	k	k-, -c-
Fijian prenasalised	b [mb]	d [nd]	dr [ndr]	s	q [ŋg]	q- [ŋg-], -s-

[1] The labio-velars /k^{w}/, /g^{w}/, and very marginally /q^{w}/ are found in only some of the Fijian languages. They occur in Wayan, but not in Bauan, and only in few forms inherited from PAn.

2.2 Tendencies to form disyllabic roots

Fijian words which are not atonic (i.e. those that are not stressless) have a minimum of two morae; that is, they cannot have simply a short-vowel nucleus, but must have minimally a long vowel (written here as a double vowel). This means that monosyllabic roots which were inherited, or which developed by medial consonantal loss and subsequent contraction (§2.6), must be disyllabised or minimally have the short-vowel nucleus lengthened:

**gem* 'grab in fist'	>	*qoom-i* 'take hold of s.t. in hand'
**teɣeb* 'having a full measure, plenty'	>	*too* 'have plenty of water in it, thick not runny'

These monosyllabic roots may also disyllabise in other ways, as they do in other Austronesian languages: by doubling the monosyllable, by treating an affixed form as the root, or by adding a prothetic vowel of the same quality as the vowel of the root.[2] Examples are as follows:

**deɣeqec* 'move vigorously and rapidly	>	*roro* (< **ro*) 'flow, run in a stream'
**tan* 'set a trap'	>	*tai* (< **ta + i*) 'snare, trap to catch land animals'
**nem* 'six'	>	*ono* 'six' (< **noo* < **no*)

Further, roots of more than two syllables may be disyllabised by the process of syncope or other types of syllable loss. However, this process affects only some of the inherited roots of more than two syllables. Many of the inherited trisyllabics are retained as such. Principally, the disyllabisation takes place when there is the loss of a medial consonant (**ɣ*, **q*, **s* — see §4.2ff.). Examples are:

**baqeɣu* 'new'	>	*vou* 'new' (< **beu* < **beqeɣu* [with weakening of the first syllable — see last paragraph in this section]),
**aɣúsu* '*Casuarina equisetifolia*'	>	*cau* 'k.o. *Casuarina* (Bau)' (< **auu* [with loss of **-ɣ-* and **-s-* plus accreted *c-* — see §4.3])

Contraction does not invariably take place. The accentual pattern of the root may be the determining feature, but this is not certain. The following form, which does not manifest contraction, seems to have had the same accentual pattern as **aɣúsu* cited immediately above:

**sabáɣat* 'strong monsoon'	>	*cavaa* 'windstorm'

In some cases of roots with three or more syllables (but not all), the first syllable is lost to form a root of one less syllable (disyllabic root in the case of a trisyllabic, and trisyllabic in the case of roots of four syllables):

**siɣejaq* 'weary'	>	*oca* 'tired' (Bau) (< **ɣejaq*)
**qasulípa* 'millipede'	>	*aliva* 'millipede'
**qalíma* 'hand'	>	*lima* 'hand'
**aluten* 'glowing stick'	>	*lito* 'wave a firebrand to keep it alight'
**ɣuqañay* 'male'	>	*ta-gwane* 'male' (cf. fn. 13)

---

2 The addition of the prothetic vowel may in fact be an example of metathesis (§2.3) which took place following the process of vowel lengthening: that is, CVV > VCV.

In one case the first two syllables of a trisyllabic seem to have been lost:

**qaɣicam* 'k.o. fern with reed-like stem'	>	*caca* 'k.o. fern: *Phymatodes longipes*' (Bau)

In one case the first syllable of a disyllabic root seems to have been lost:

**qinep* 'spend night'	>	*noo* 'sit, stay, occupy space'

However, some trisyllabic roots continue as trisyllabics, although the antepenult in that case may be weakened. In the first two of the following examples the antepenult is weakened, but in the other examples all three syllables are retained with no weakening. Weakening of the penult was also assumed for the development of **baqeɣu* 'new' to *vou* 'new', given at the beginning of this paragraph:

**bitáquɣ 'Calophyllum inophyllum'*	>	*betau* 'k.o. tree: *Mammea odorata*'
**palawud* 'go to sea'	>	*volau* 'boat shed'
**taliŋa* 'ear'	>	*taliga* 'ear'
**qañítu* 'supernatural spirit'	>	*anitu* 'spirit, supernatural being'
**qabuca* 'large social grouping'	>	*avusa* 'descent group, clan'
**tambuɣi* 'triton shell'	>	*tavui* 'triton, helmet shell'

2.3 Metathesis

Metathesis is a minor process in the history of Fijian, but there are forms that clearly show metathesis. The following forms show consonantal metathesis:

**basaq* 'wet'	>	*sava* 'wash, cleanse'
**caliw* 'give in exchange'	>	*soli* 'be given'(< **cawli*)
**sacaŋ* 'gills'	>	*see* 'gills'
**tusuj* 'knee'	>	*turu* 'knee' (< **turus* < **tusur*)

We assume syllabic or vocalic metathesis in some forms:

**becuɣ* 'sated'	>	*vuuse* 'sated' (dialectally, *vuse*)
**qasulipan* 'millipede'	>	*aliva* 'millipede' (< **suqalipan* with loss of the first syllable)
**jiɣuc* 'sprinkle'	>	*sui* 'be watered, sprayed'

2.4 Consonant clusters

Fijian permits only open syllables, and there are no consonant clusters. The prenasalised stops listed in Chart 1 function as phonemic units. No clusters develop by doubling monosyllabic roots, for these lose the root-final consonant in the first syllable of the resulting form:

**bedbed* 'wind around'	>	*vovot-i* 'be wrapped in a sling to carry'
**dapdap 'Erythrina* spp.'	>	*rara 'Erythrina*'

PAn medial nasal-plus-consonant clusters may be inherited as prenasalised consonants, or they may be simplified.

**cuŋkit* 'poke'	>	*cukit-a* 'dig up with stick' (Bau)
**leŋkung* 'bent'	>	*loku* 'bent, twisted, folded'
**pantaɣ* 'platform, flat extension of house'	>	*vata* 'shelf, platform'
**senti* 'stop'	>	*oti* 'finish, complete, end'
**tambuɣi* 'triton shell'	>	*tavui* 'triton, helmet shell'
**ambay* 'wave the arms'	>	*abe* 'walk with the arms swinging'
**lampin* 'wrapper'	>	*labi* 'bundle of fish wrapped for cooking' (Bau)
**paŋudañ* 'pandanus'	>	*padra* 'pandanus' (< **pandan* < **paŋdan*)
**waŋkaŋ* 'boat'	>	*waqa* 'boat'

Nasal clusters which arise from the doubled monosyllabic roots are always simplified:

**diŋdiŋ* 'wall, partition'	>	*riri* 's.t. which blocks off the wind'

2.5 Development of labiovelars

Proto Oceanic (POc) developed labio-velars **p*^{w}, **b*^{w} and **m*^{w}, often reflected as labials with a velarised *w*-off-glide (Lynch 2002), and some of these labio-velars occur in forms which are reflected in Fijian. In one case these labio-velars are reflected as such in Wayan, but not in Bauan, where they lose their off-glide. Otherwise, they are reflected sometimes with a labial or sometimes with a velar articulation. No rules can be specified which predict how the labio-velars develop. Apparently the forms with labio-velars in POc developed in different ways in different areas, and these forms spread individually. The following list gives forms with PAn etymologies that developed POc phonemes with labio-velar articulations and that are reflected in Fijian:

**belaq* 'split'	>	POc **b*^{w}*elaq*	>	*bola* 'split' (but Bauan *kola* with a velar articulation)
**beɣek* 'pig'	>	POc **b*^{w}*oRok*	>	Viti Levu *qoo* 'pig'
**taw + mata*	>	POc **tam*^{w}*ata*	>	*tamata* 'man, person'
**taw + maɣuqañay*	>	POc **tam*^{w}*aqane*	>	*tamwane* 'male'
**um + inum* 'drink'	>	POc **m*^{w}*inum*	>	*gunu*[3] 'drink'(Bauan)
**qúmaŋ* 'hermit crab'	>	POc **qum*^{w}*aŋ*	>	*uga* 'hermit crab'

2.6 Sequences of vowels

Vowel sequences which were inherited or which developed by the loss of intervocalic consonants were contracted: like vowels became single vowels which may be long or short depending on whether the root is monosyllabic or not (see §2.2) or depending on other conditions which have not been determined:

**daɣat* 'open area'	>	*raa-raa* 'open fields'
**beɣecay* 'oar'	>	*voce* 'paddle a canoe'
**ɣabi* 'evening'	>	*ei avi* 'in the evening'
**juɣuq* 'broth'	>	*suu* 'broth' (Bau)

---

3 The *w*-off-glide affected the following vowel before being lost.

Sequences of **a* followed by **u* (or which developed as such) became **au* and then, if there was no affix or the root was not trisyllabic, changed further to /o/ as inherited /aw/ did (§3.5.3):

**dasuwen* 'leaf'	>	*roo* 'leaf' (< **daun*)[4]
**jaɣum* 'needle'	>	*i-sau* 'needle' (Bau)
**bitáquɣ 'Calophyllum inophyllum'*	>	*betau* 'k.o. tree: *Mammea odorata*'

**y* was apparently lost between **a* and **u* (§3.5.3), and the development of the resulting **au* was similar to that of the preceding paragraphs:

**kasiw* 'wood, tree'	>	*kau*[5] 'wood, tree' (Bau) (< **kayu* < **kaiw*)
**isekan* + **qisu* 'fish + shark'	>	*ikoo* 'shark' (< **ikau* < **ikayu* < **ika iu*)[6]

To explain the following form we additionally assume that the sequence **awa* which developed was treated like **au*.

**qañiɣuwan* 'honey bee'	>	*oni* 'k.o. bee' (< **awn* + *i* < **awan* + *i* < **ayuwan* + *i* < **ayiuwan* + *i*)

We also assume that a sequence **aui* > /au/:

**kaweit* 'hook'	>	*kaut-i* 'catch s.t. on a hook' (from **kauit-i*)

Sequences containing **e* assimilated to the first vowel of the sequence if this first vowel was stressed (that is, occurred in a form without a suffix), but to the second vowel of the sequence if the stress fell on the second vowel (that is, occurred in a suffixed form)[7] (Geraghty 1983:162):

**jaqet* 'bad'	>	*caa* 'bad'
**iqeɣuŋ* 'nose'	>	*g-icu* 'nose'
**iseq* 'urine'	>	*m-ii* 'urinate'[8]
**tuqed* 'stand upright'	>	*tuu* 'stand, be upright'

The following forms occurred suffixed, and the stress fell on the **e*. The reflex shows assimilation to the second vowel:

---

4 Bauan reflects *rau* 'leaf'. Presumably this is a reflex which developed from a suffixed alternant, where *rau* was pronounced in two syllables with accent on the second, before the suffix, whereas *roo* develops from a monosyllabic alternant appearing in the unsuffixed form, which then underwent lengthening of the vowel nucleus as a monosyllabic root (§2.2)

5 The development of /au/ is explained by the fact that the form which was generalised was found in compounds, where it would have a reflex containing /au/ because of the accentuation. This root has a reflex *kai* in Wayan, which is probably not directly inherited.

6 The occurrence of /o/ rather than /au/ may be explained by the fact that the stress was on the first syllable. This does not explain why the /o/ was lengthened, however. For that there is no explanation.

7 Stress in Fijian falls on the penult of the phonological word. That is, if the root is unsuffixed, the stress falls on the penult of the root, but if the root is suffixed, the stress moves away from the penult — to the final syllable of the root if the suffix has one syllable.

8 Blust (1983–84) reconstructs PAn **miqmiq* to account for this form, on the basis of forms shaped /mimiq/ and /mimi/ in northern Philippine languages, /meqmeq/ 'urine' in Chamorro and /mimi/ in various Oceanic languages, all of which mean 'urinate'. I believe that all these forms are cognate with forms which reflect PAn **iseq*, in the affixed form **-um- iseq*. In Fijian the contracted monosyllabic root which developed manifests a lengthened vowel nucleus, as often happens in Fijian (§2.2 above), but in the other languages referred to the monosyllabic root which developed was doubled to form a disyllabic root.

**bayeq* 'abscess'	>	*boo* 'boil'[9]
**seyup* 'blow'	>	*uv-i* 'be blown up'
**taseyup* 'blow'	>	*tuuv-i* 'blow s.t. (e.g. conch)' (< **teseyup* with weakening of the first syllable [§2.2, last paragraph])

3 Vowels and diphthongs

Except for the syncopations and vowel contractions discussed above, Fijian reflects the inherited PAn vowel system fairly faithfully. There are few changes in vowel quality conditioned by the environment.

3.1 **i*

**i* > /i/ in all positions.

**cici* 'k.o. mollusc'	>	*cici* 'molluscs and oysters'
**ipil* '*Intsia bijuga/retusa*'	>	*ivi* '*Inocarpus fagifer*'
**sayin* 'wind'	>	*cagi* 'wind'
**iqeyuŋ* 'nose'	>	*g-icu* 'nose'
**ciwa* 'nine'	>	*ciwa* 'nine'

**i* > /e/ under conditions which are not totally clear. Most of the cognates of these forms in other Oceanic languages also reflect /e/, and this is a change which took place early in Proto Oceanic times. Forms which ended in **i* plus **y* or **g* show final /-e/ in Fijian.

**biybiy* 'lips'	>	*bebe* 'vagina'
**bukig* 'mountain, hill'	>	*buke* 'mound, hillock'
**tambig* 'flat vessel'	>	*tabe* 'flat open basket'

However, there are also other cases of **i* reflected by /e/. Most of these forms also show /e/ in other Oceanic languages:

**lawi* 'tail feathers'	>	*laawe* 'tail feathers'
**cakit* 'painful'	>	*ma-cake* 'disease of the tongue'
**tebiŋ* 'edge, embankment'	>	*tebe* 'broad edge (Bau)'
**tuli* 'deaf'	>	*tule* 'earwax'

In the following two forms the final *-e* gets assimilated to a preceding stressed /a/ by the rule of §2.6:[10]

**daqis* 'forehead'	>	*raa* 'forehead' (< **rae*[7])
**taqi* 'faeces'	>	*taa* 'faeces'

---

9 Normally roots in Fijian had penultimate stress, but in this case we may assume on the basis of cognates in other Fijian languages that this form developed a prothetic /a/ (Geraghty 1983:§4.5.6) and that the stress moved to the prothetic vowel as contraction took place — i.e. stress moved away from the penult: **báyeq >*ábae > *abo > boo* (with loss of prothetic vowel and lengthening of the vowel nucleus in a monosyllabic root).

10 Cognates with final /e/ occur in other Fijian languages or other Eastern Oceanic languages, evidence that a form with final /e/ preceded the forms attested in Wayan.

3.2 **a*

**a* > /a/ except in those cases in which **a* developed into a vowel sequence which contracted (§2.6):

**qataɣ* 'liver'	>	*ate* 'liver'
**jalan* 'way'	>	*cala* 'way, path'
**qañitu* 'evil spirit'	>	*anitu* 'spirit, supernatural being'
**taliŋa* 'ear'	>	*taliga* 'ear'

/a/ is lengthened in monosyllabics (§2.2). However, there are other cases where /a/ has lengthened for which there is no explanation:

**datu* 'chieftain'	>	*raatuu* 'head of clan, chief'[11]
**lawi* 'tail feathers'	>	*laawe* 'tail feathers'
**sabayat* 'strong monsoon wind'	>	*cavaa* 'windstorm'

In a few unexplained cases **a* is reflected as /e/. In some cases this is a purely Wayan development, and eastern Fijian languages reflect /a/, but in other cases /e/ is manifested in all the attested languages.

**dawat* 'reach, get as far as'	>	*rewa* 'be able to be done' (Bau *rawa*)
**gali* 'dig'	>	*keli* 'be dug'
**luwaq* 'vomit'	>	*lue* 'vomit' (Bau *lua*)
**ma* 'tongue'	>	*mee* 'tongue'
**pan* 'bait'	>	*bee* 'bait'
**sacaŋ* 'gills'	>	*see* 'gills'
**tulak* 'push away'	>	*tulek-i* 'move s.t. by pushing or pulling'

3.3 **e*

**e* is mostly reflected by /o/, except in the cases of syncope and contraction (§2.2).

**beŋi* 'night'	>	*bogi* 'night'
**deŋeɣ* 'hear'	>	*rogo* 'be heard'
**kuden* 'earthen pot'	>	*kuro* 'clay cooking pot'
**tebu* 'sugar cane'	>	*tovu* 'sugar cane'
**teken* 'stick to lean on'	>	*i-toko* 's.t. that gives support' (*tokon-i* 'support a proposition')

In some cases **e* is reflected by /e/. The change of **e* to /o/ or /e/ may have originally been dependent on accentual conditions, and there are cases of doublets. Some of these forms also reflect /e/ in other Oceanic languages.

**buɣes* 'squirt water out of mouth'	>	*bue* 'boil, bubble up'
**bekbek* 'pulverise'	>	*vevek-a* 'rusty, worm-eaten' (Bau)
**becuɣ* 'full, sated'	>	*vuse* 'sated'
**liceqes* 'nit'	>	*cile* 'nit' (Bau *lise*)
**lepa* 'tack'	>	*leva* 'tack'

---

11 See Geraghty (1983:§4.5.6) on lengthening of the vowel nucleus in a monosyllabic root.

In one unexplained case **e* is reflected as /i/:

**betuŋ* 'k.o. bamboo' > *bitu* 'bamboo'

The sequences **ye* and **ey*[12] are reflected as /i/ and **we* is reflected as /u/:

**wasyeɣ* 'water' > *wai* 'water'
**kaweit* 'hook' > *kaut-i* 'catch s.t. on a hook' (from **kauit-i*)
**weliq* 'do again, go back' > *uli* 'reply in speech'

3.4 **u*

**u* > /u/ in all environments, except in vowel sequences which developed by loss of consonants (§2.6):

**búbu* 'fish trap' > *vuvu* 'fish trap with narrow funnel entrance'
**buku* 'knot, node' > *buku* 'knob, node'
**manuk* 'bird' > *manu-manu* 'bird, animal'
**qañitu* 'evil spirit' > *anitu* 'supernatural spirit'
**qudaŋ* 'crustacean' > *ura* 'crustacean'

Unstressed **u* is elided or loses its vocalic quality (i.e. > [w]) in the antepenult:

**ɣuqañay* 'male' > *ta-g-wane* 'male'[13]
**kuapu* 'grouper' > *kwavu* 'rock cod'
**kua* 'say' > *kwa-i* 'say s.t.'

In Wayan **u* is lost after /m/ in unstressed syllables. The /m/ bears a mora when loss of /u/ causes the /m/ to be final in the word.

**ñamuk* 'mosquito' > *yam* 'mosquito'
**lumu* 'soft, tender' > *ma-lum* 'weak, soft'
**lumuk* 'oily' > *lum-i* 'anoint'

In a few cases **u* is reflected as /i/. There are cases of double reflexes, one showing /i/ and the other /u/. Fijian forms which reflect /i/ from **u* for the most part have cognates with /i/ in other Eastern Oceanic languages. The conditions for the development of /i/ are unknown.

**aluten* 'glowing stick' > *lito* 'wave a firebrand to make it flare'
**caluɣ* 'watercourse' > *sali* 'flow'
**lumut* 'moss' > *lumi* 'seaweed growing on sand'
**pulec* 'twist, wind' > *viloc-i* 'wring clothes' and *vuloc-a* 'twist s.t.' (Bau)
**puceg* 'navel' > *vico-vico* 'navel'

---

12 A reflex of the sequence **ey* is attested in only one form, which is not listed for Wayan. This is the Bauan form *cago-laya* '*Zingiber zerumet*' (< POc **laqiya* – cf. Geraghty 1983:129).

13 Lynch (2002:323–327) suggests metathesis as an explanation of the form *tagwane*. That is, **taw + *mayuqañay* > **taw + *maqañay* (with the loss of the antepenult of **ɣuqañay*) > **tamwanay* (with loss of **q* and metathesis of the [w] and [m] and subsequent change of **ay* > /e/ and **m^{w}* > /ŋ^{w}/). **m^{w}* must be reconstructed for Proto Oceanic because many Oceanic languages reflect **m^{w}* in this form.

3.5 Diphthongs

*3.5.1 *ay*

**-ay* is reflected as /e/:

**beyecay* 'oar'	>	*voce* 'paddle, row'
**matay* 'die'	>	*mate* 'die'
**qatay* 'liver'	>	*ate* 'liver'
**cakay* 'climb, ride'	>	*cake* 'climb, ascend'

In the one example of **ay* in a monosyllabic form which is stressed, the reflex is disyllabic /ei/:

**ay* 'who?'	>	*ei* 'who?'

*3.5.2 *uy*

**uy* is reflected as /i/:

**culuy* 'shoot'	>	*culi* 'banana or taro sucker'
**sapuy* 'fire'	>	*v-avi* 'cook in earth oven' (< **pa-sapuy*)

*3.5.3 *aw*

When unstressed, **aw* >/o/:

**danaw* 'lake'	>	*drano* 'lake, body of water inland'
**laŋaw* 'a fly'	>	*lago* 'a fly'
**namaw* 'sheltered water, deep place in river'	>	*namo* 'pool left on reef at low tide'

Under stress, **aw* > /au/ in roots of more than one syllable. /au/ is also the reflex when **au* develops because of loss of a medial consonant (§4.2ff.):

**lawud* 'seaward'	>	*toka-lau* 'east wind'
**taŋuq* 'put'	>	*tau* 'be put, given, located'
**taqu* 'be familiar, know'	>	*ma-tau* 'right hand', *maa-tau* 'familiar'

When the resulting form is monosyllabic **aw* > /oo/ (§2.2, first paragraph):

**baw* 'above'	>	*boo-boo* 'shore, inland'[14]

4 Consonants

Fijian reflects the PAn consonants fairly faithfully in initial and medial position with the exception of changes which have been made by all the Oceanic languages:

(1) consonantal gradation has developed (§2.1);

(2) **s* was lost without a trace.

---

14 The form with /oo/ developed, and subsequently it was doubled.

Further, Fijian made innovations together with many of the Eastern Oceanic languages:

(3) **p* and **b* merged;

(4) **j* and **c* have merged and **g* has merged with them in medial position;

(5) root-final consonants have been lost but their reflexes are frequently retained in affixed position (see §5).

Finally, Fijian underwent further consonantal developments:

(6) **ɣ* and **q* were lost entirely.

Chart 3 outlines these developments.[15] For the value of the Fiji transcriptions see Chart 1.

Chart 3: Fijian reflexes of PAn Consonants

PAn	Fiji (prenasalised and plain reflexes)	PAn	Fiji
**p*	*b, v*	**ɣ*	*Ø*
**t*	*d, t*	**m*	*m*
**k*	*q, k*	**n*	*n*
**q*	*Ø*	**ñ*	*y, n*
**b*	*b, v*	**ŋ*	*g*
**d*	*dr, d*	**s*	*Ø*
**j*	*s, c*	**w*	*w*
**c*	*s, c*	**y*	*y*
**g*	*k-, -c-*	**l*	*l*

4.1 Mergers

The mergers discussed here resulted in morphophonemes, not phonemes, for the resulting phonemes are in consonant gradation (§2.1) — that is, there is alternation between prenasalised and plain phonemes, over different languages or in the same language and dialect but in different lexemes.

4.1.1 Merger of *p *and* *b, *mp *and* *mb

**p* and **b* merged to /b/ [mb] or /v/. Examples with /b/:

**papan* 'board, planks'	>	*baba* 'board, planks'
**pan* 'bait'	>	*bee* 'bait' (/ee/ explained in §3.2, last paragraph)
**pulut* '*Urena lobata* (a plant producing sticky material)'	>	*bulu* 'adhere, be stuck on'
**kapit* 'be close'	>	*kabit-i* 'adhere'
**lepit* 'folded'	>	*lobi* 'folded lengthwise and crosswise' (Bau)
**beŋi* 'night'	>	*bogi* 'night'
**buku* 'knot on wood, node'	>	*buku* 'knob, node'

15 I assume a rather different phonology for PAn than is commonly assumed in the literature on the history of the An languages. See Chart IA in §2 of my paper in this volume, 'The sounds of Proto Austronesian'.

Examples with /v/:

**pat* 'four'	>	*vaa* 'four'
**piga* 'how many?'	>	*vica* 'how many?'
**paka-* 'causative prefix'	>	*vaka-* 'causative prefix'
**apa* 'what?'	>	*ava* 'what?'
**sipaɣ* 'spouse's sibling or cousin of same generation'	>	*iva* 'son-in-law'
**kuɣapu* 'grouper'	>	*kwavu* 'rock cod'
**batu* 'stone'	>	*vatu* 'stone'
**beɣecay* 'paddle'	>	*vose* 'paddle a canoe'
**biɣaq* '*Alocasia* sp.'	>	*via* 'giant taro'
**bulañ* 'moon, month'	>	*vula* 'moon, month'
**labaw* 'rat'	>	*ka-lavo* 'rat' (Bau)
**lebeŋ* 'bury'	>	*lovo* 'be buried, covered over'

The prenasalised bilabial stops are mostly inherited as nasal-grade consonants:

**ambay* 'wave the arms'	>	*abe* 'walk with the arms swinging'
**dampaɣ* 'be flat'	>	*raba* 'broad, wide'
**lampin* 'wrapping'	>	*labi* 'bundle of fish tied together for cooking'

but occasionally as plain-grade consonants (§2.4):

**tambuɣi* 'triton shell'	>	*tavui*[16] 'helmet shell, triton shell'
**tumpaŋ* 'pile up'	>	*tuvan-i* 'set in order, arrange'

4.1.2 Merger of *c, *j *and* *-g[17]

**c*, **j* and **-g-* merged to /s/ and /c/. Note that what we reconstruct as initial **g-* in PAn merged with **k-* rather than **c* and **j* (§4.1.3). First, examples reflecting the nasalised grade (with /s/):

**caluɣ* 'watercourse'	>	*sali* 'flow'
**cekcek* 'stuff, pack chock full'	>	*soso* 'stuffed, packed'
**cinaɣ* 'ray of light'	>	*sina* 'lamp, torch'
**cuɣi* 'bone'	>	*sui* 'bone'
**beɣecay* 'oar'	>	*vose* 'paddle a canoe'
**jalateŋ* 'nettle tree'	>	*salato* 'nettle tree'
**juɣuq* 'broth, liquid material'	>	*suu* 'broth, soup'
**ɣugaŋ* 'load cargo'	>	*usa* 'be shipped, carried as cargo'

---

16 The loss of the prenasalisation after the first syllable probably has to do with its unstressed position.

17 There is evidence that Oceanic did not merge the medial PAn **-g-* with **c*. Geraghty (1986:296–298) reconstructs for Proto Central Pacific (PCP), an ancestor of Fijian, a phoneme **z* which contrasted with the reflexes of **c* (PCP **c* and **s*). He suggests (1983:155–157) that some cases of this **z* are reflexes of PAn **-g-*, and later proposes that PCP **z* reflects PAn **j* and **g* (1986:32–303). This phoneme PCP **z*, however, merges in Fijian with the reflexes of PAn **c*.

Now for examples reflecting the plain grade (with /c/):

**caŋa* 'branch'	>	*caga* 'span, vagina'; also *saga* 'crotch' (Bau)
**cici* 'k.o. mollusc'	>	*cici* 'molluscs and oysters'
**cucu* 'breast'	>	*cucu* 'suck on breast'
**kacaw* 'rafter'	>	*kaso* 'rafter'
**jalan* 'road, way'	>	*cala* 'track'
**jeket* 'stick'	>	*coko* 'netted, trapped, caught'
**qujañ* 'rain'	>	*uca* 'rain'
**qágan* 'name'	>	*aca* 'name'
**piga* 'how many?'	>	*vica* 'how many?'

There are two forms with medial **j* which reflect a prenasalised consonant in PAn. In the first case below, the medial **nj* is reflected with /-s-/ just as the **-j-* in prenasalised grade, but in the second case **-nj-* seems to have merged with **-nd-* and produces /dr/ (as in §4.1.5). There may have been some special fact about the histories of the two forms which accounts for the difference, but the probable explanation is that one of the two etymologies is incorrect:

**banjaɣ* 'row'	>	*basa* 'level, in a line with' (Bau)
**tinjaw* 'look at'	>	*tidro* 'peer, watch'

4.1.3 Merger of *k *and* *g-

Initially, **g-* falls together with **k* as it does in many Austronesian languages. **k* and **g* become /k/ in the plain grade and /q/ [ŋg] in the prenasalised grade. Examples of these reflexes in the prenasalised grade:

**giliŋ* 'roll over s.t.'	>	*qili* 'be rolled, kneaded into shape'
**gem* 'hold in fist'	>	*qoom-i* 'grab hold of in hands'
**kep* 'clasp'	>	*qov-i* 'envelop s.t. so that it is completely covered' (also with the plain grade: *kov-i* 'covered')

There are no examples of **-k-* in medial position reflected by a nasalised grade. There is one example of **ŋk* reflected with a prenasalised grade:[18]

**waŋkaŋ* 'canoe'	>	*waqa* 'canoe'

In other cases of **ŋk*, the nasal cluster is simplified (cf. examples in §2.4). Examples of **k* and **g-* with plain-grade reflexes are as follows:

**kacaw* 'rafter'	>	*kaco* 'rafter'
**kep* 'clasp'	>	*kov-i* 'covered'
**kuku* 'fingernail, claw'	>	*kuku* 'nail, claw'
**likud* 'back'	>	*likur-i* 'diverge from, veer from'
**tuk* 'strike, peck, beak'	>	*tuk-i* 'be hammered'
**gilig* 'side'	>	*kili* 'side, edge'
**guɣita* 'octopus'	>	*kuita* 'octopus' (Bau)

---

[18] This form is probably not inherited from PAn, but it spread through the Austronesian languages at an early enough time that it can be considered to exemplify what has happened to medial **ŋk*.

4.1.4 Plain and nasal reflexes of *t

**t* has the reflex /d/ [nd] in the nasal grade:

**tales* 'taro'	>	*dalo* 'taro'
**teɣac* 'hard wood'	>	*doa* 'heartwood'
**tuqus* 'true'	>	*duu* 'right and correct' (also *tuu-tuu* 'exactly at such-and-such a time')

but is reflected as /t/ in the plain grade:

**taɣaq* 'cut, chop'	>	*taa* 'be cut by chipping or slashing'
**taliŋa* 'ear'	>	*taliga* 'ear'
**telu* 'three'	>	*tolu* 'three'

The sequence **nt* was simplified or it was reflected the same way as the nasal grade of **t*:

**panta* 'flat area'	>	*vata* 'shelf, platform'
**punti* 'banana'	>	*vudi* 'plantain'

4.1.5 Plain and nasal reflexes of *d

**d* has the reflex /dr/ [ndr] in the nasal grade and /r/ in the plain grade. First, the nasal grade:

**cedu* 'hiccough'	>	*cedru* 'gasp for air'
**daɣeq* 'blood'	>	*draa* 'blood, sap'
**dánaw* 'lake'	>	*drano* 'body of fresh water inland'
**dañum* 'water'	>	*dranu* 'water'

Now for examples of the plain grade:

**dapdap* '*Erythrina* sp.'	>	*rara* '*Erythrina*'
**deŋeɣ* 'hear'	>	*rogo* 'heard'
**diŋdiŋ* 'wall, partition'	>	*riri* 's.t. which blocks from the wind'
**dusa* 'two'	>	*rua* 'two'
**kuden* 'earthenware pot'	>	*kuro* 'clay cooking pot'
**tuduq* 'leak, drip'	>	*turu* 'leak, drip'

The sequence **nd* is reflected like the nasal grade of medial **-d-*:

**andak* 'go up'	>	*cadra* 'upwards'
**paŋudañ* 'pandanus'	>	*vadra* 'pandanus' (< **paŋdan*)

4.2 Consonants which do not undergo gradation

In addition to the liquids and nasals, the consonants **ɣ*, **q*, and **s* fail to undergo consonant gradation. They disappeared without a trace in Fijian, and their loss was followed by vowel contraction (§2.6). **s* was lost prior to Proto Oceanic and to my knowledge leaves no trace in any Oceanic language. **q* was lost more recently and is reflected regularly with [ʔ] in Tongan, which is closely related to Fijian. But based on the fact that it has no nasal grade, we may deduce that **q* became [ʔ] before the phenomenon of consonant gradation

developed (Ross 1988:32). Similarly, by the same reasoning we may deduce that **ɣ* had become a spirant before the development of consonant gradation (Ross 1988:32). **ɣ* is widely reflected as [r] or [l] in many Oceanic languages, but it is lost in the Central Pacific languages.

*4.2.1 *ɣ*

**ɣ* is lost without a trace:

**ɣuqañay* 'male'	>	*ta-gwane* 'male' (< **taw* + **mayuqañay* — cf. fn. 13)
**ɣabut* 'uproot'	>	*cavut-i* 'pull out'
**ɣebaq* 'fallen in'	>	*ova* 'fall sideways'
**ɣugaŋ* 'load cargo'	>	*usa* 'be shipped, carried as cargo'
**sabaɣat* 'monsoon with heavy wind'	>	*cavaa* 'windstorm'
**ñeɣab* 'yesterday'	>	*ena-noa* 'yesterday' (Bau)
**wiɣi* 'left'	>	*ma-wii* 'left'

*4.2.2 *q*

**q* is lost without a trace:

**qatay* 'liver'	>	*ate* 'liver'
**qulu* 'head'	>	*ulu* 'head'
**qeney* 'sand'	>	*one* 'sand' (in placenames — Geraghty 1983:394)
**basequ* 'smell'	>	*bo-i* 'smell' (Bau)
**tuqed* 'stand upright'	>	*tuu* 'stand, be upright'

*4.2.3 *s*

**s* is lost without a trace:

**sacaq* 'grind down to sharpen'	>	*aca* 'filed, grated, sharpened'
**sipaɣ* 'spouse's siblings or cousins of the same generation'	>	*iva* 'son-in-law'
**seɣup* 'blow'	>	*uv-i* 'be blown up'
**dusa* 'two'	>	*rua* 'two'
**iseq* 'urine'	>	*m-ii* 'urinate'
**udesi* 'last'	>	*m-uri* 'after, following'

4.3 /c/ and /y/ accretion

Forms beginning with **a-* or which developed initial **a-* after the loss of **s* or **ɣ* in a few cases developed an initial /y/ which then changed to /c-/ (§4.5.3). This change probably took place in Proto Eastern Oceanic times before the loss of **q*, so that there are no examples of

this process with roots beginning with **qa-*. This process probably is a generalisation of a form with a [y] glide which developed after affixes ending in /i/ or the locative marker *i*.

**aɣusu* '*Casuarina* spp.'	>	*cau* 'k.o. casuarina' (Bau)
**ɣabut* 'uproot'	>	*cavut-i* 'pull out'
**sabaɣat* 'monsoon with heavy wind'	>	*cavaa* 'windstorm'

The process was not automatic, and in fact the number of forms which do not show *c*-accretion is far greater than those which show the accreted /c-/.[19] A study of all the examples might reveal the constraints which determined whether or not *c*-accretion could take place. Here are three examples of forms which did not undergo *c*-accretion:

**ambay* 'wave arms'	>	*abe* 'walk with the arms swinging'
**ɣabíi* 'night'	>	*ei avi* 'yesterday'
**sacaq* 'grind down to sharpen'	>	*aca* 'filed, grated, sharpened'

In some current dialects a *y*-glide is automatically inserted before word-initial /a/. The occurrence of /y/ in these cases is non-contrastive and has nothing to do with /c/-accretion.

4.4 Nasals

**m*, **n*, and **ŋ* are reflected faithfully in initial and medial position by /m/, /n/, and /g/ [ŋ] respectively.

**mamaq* 'chew'	>	*mama* 'chew'
**manuk* 'bird'	>	*manu-manu* 'bird, animal'
**matey* 'dead'	>	*mate* 'dead'
**lima* 'five'	>	*lima* 'five'
**nunuk* '*Ficus benjamina*'	>	*nunu* '*Ficus*' (Bau)
**kan* 'eat'	>	*vaa-kan-i* 'feed'
**ŋucuq* 'mouth'	>	*gusu* 'mouth of animal'
**beŋi* 'night'	>	*bogi* 'night'

**ñ* is reflected by /n/ except in three forms inherited from PAn. In these three forms (but only in western Fijian languages), **ñ* is reflected by /y/. In Bauan and other eastern Fijian languages, **ñ* has merged completely with **n*.

**meniyak* 'oil'	>	*moya* 'brain' (< **meñak*)
**ña* 'third person genitive marker'	>	*ya* 'third person possessive'
**ñamuk* 'mosquito'	>	*yam* 'mosquito'

The conditions for the merger of **ñ* with **n* are unknown.[20]

---

19 The accretion of /c-/ is presumed to have taken place in Proto Eastern Oceanic times (Geraghty 1983, passim). In some cases Wayan fails to show *c*-accretion where other Fijian languages do. Geraghty assumes sporadic *c*-loss, whereas I treat these forms as though they never received *c*-accretion in the ancestor of Wayan.

20 The small number of forms which reflect /y/ from an earlier **ñ* is an indication that the change of **ñ* > /n/ was not carried through to the end.

4.5 **l*, **w*, and **y*

The phonemes **l* and **w* were stable, but **y* is lost in some environments.

*4.5.1 *l*

**l* is retained as /l/ initially and medially:

**laŋit* 'sky'	>	*lagi* 'sky, atmosphere'
**lemec* 'drown'	>	*lomoc-i* 'dunk'
**lima* 'five'	>	*lima* 'hand'
**bulañ* 'moon, month'	>	*vula* 'moon, month'

In four cases **l* was lost unaccountably:

**qaluɣ* 'deep part of river or pool'	>	*au* 'flow, run freely'
**caluwaki* 'sea urchin'	>	*caawaki* 'k.o. sea urchin'[21]
**malu* 'be ashamed'	>	*maa-dua* 'bashful', *si-maa* 'be ashamed'
**tuqelañ* 'bone'	>	*tua* 'bone' (also *tula* 'branch, twig')

There is one case in which **l* > /n/ in a syllable before another syllable beginning with /m/:

**limac* 'bail out'	>	*nimati* 'be bailed out'

*4.5.2 *w*

**w* is retained initially and medially as /w/ for the most part.

**waɣa* 'climbing vine'	>	*waa* 'creeper'
**wiɣi* 'left'	>	*ma-wii* 'left'
**dawat* 'reach, get as far as'	>	*rewa* 'be able to be done'

The sequence **we* > /u/ (§3.3, last paragraph). The sequence **uwa* is reflected in the same way as **ua* — that is, there is no contrast between [ua] and [uwa].

**banuwa* 'land, place'	>	*vanua* 'land, region, community'
**buwaq* 'fruit'	>	*bua* 'fruit'
**cuwan* 'dibble stick'	>	*sua* 'stab, pierce'

At the end of a word, **w* formed a diphthong with the preceding vowel. The only sequence found in the data with a final **w* is **-aw*, and **aw* became simplified to /o/ or /au/ according to the accentual conditions (§2.2, §3.5.3).

*4.5.3 *y*

Initially and medially, **y* > /c/. The best evidence for /c-/ as the reflex of **y-* comes from the process of *c*-accretion (§4.3). Otherwise, there is only one example in the data. There is one example of each of **uyu* > /ucu/ and **aya* > /aca/:

---

21 The change of **au* > to /aa/ is explained in §3.5.3.

**yuyu* 'k.o. crab' > *cucu* 'k.o. fish'
**layay* 'sail' > *laca* 'sail'

However, another item shows loss of intervocalic **y*, if the Fijian form is in fact connected:[22]

**dáya* 'inland' > *raa* (in *vua i raa* 'west wind' — in Waya coming from the ocean)

Between **a* and **u*, **y* is reflected by /-c-/ in some dialects. However, in the word for 'tree' (**kasiw* > **kayu*), **y* is reflected with Ø in Bauan, as here it typically occurred followed by a word for the specific type of tree — i.e. in atonic position):

**kasiw* 'wood, tree' > *kacu* (dialectal) and *kau* (Bau) 'wood, tree'

**y* is lost after **i* — that is, there is no contrast between /iyV/ and /iV/:

**liyus* 'go around' > *liu* 'front, precede, surpass'
**piyan* 'desire' > *via* 'auxiliary expressing desire'

4.6 Nasal substitution

A process of nasal substitution for initial root consonants is manifested by a small number of forms. By this process stop consonants are replaced by the homorganic nasals, and roots with initial vowels have a preposed /g/ [ŋ]:

**baksaw* '*Rhizophora mucronata*' > *mako* 'k.o. tileaceous tree'
**baŋa* 'gap stand open' > *maga* 'vagina' (Bau)
**pejem* 'close eyes' > *moce* 'sleep'
**putuc* 'cut off' > *mudu* 'cut off, amputate'
**take* 'show direction' > *i-naki* 'intent'

5 The stops in final position

Although the stops in final position were lost in post-Oceanic times, they are retained before an affix, and for the most part (but not invariably) the root-final consonant before the suffix in the current language reflects the inherited root-final consonant. There is evidence that the root-final consonants underwent special changes before getting lost.

First, **-j* fell together with **-d*. The evidence for this is from the following form which underwent metathesis — i.e. the final **-j* became medial. It manifests the same reflex as **-d* — that is, /r/:

**tusuj* 'knee' > *turu* 'knee'

Second, the final **-d* became devoiced to /-t/ (after the word for 'knee' underwent metathesis). Both final **-d* and final **-j* are reflected as /-t/ before suffixes:

**bedbed* 'spool, wind' > *vovoti* 'be wrapped in sling to carry on back'
**penej* 'plugged up' > *vonot-i* 'stop or plug up s.t.'

---

22 Further, the word for 'ginger', not attested for Wayan, is another case: it reflects /aa/ in some dialects (Geraghty 1983:129) and /aya/ in Bauan: **laqeya* 'ginger' > *cago-laya* 'ginger' (Bau) and *cago-laa* dialectally. Compare footnote 12 for an account of the development of this form.

However, there is a counter-case. The following form can be taken to have petrified the suffixed form and reflects the development of the root-final consonant as if it were intervocalic — that is, **-d-* > /-r-/:

**likud* 'back' > *likur-i* 'veer from, diverge from'

Further, final **-n* and **-ŋ* are neutralised. Evidence for this comes from the suffixed forms with final **-ŋ*, which reflect /-n/ before a suffix:

**lebeŋ* 'bury' > *lovon-i* 'cover over, bury'

Otherwise, the root-final consonants are usually reflected in the same way as they are medially in the plain grade but those that were lost are replaced by other consonants before the suffix. (Examples are given in the individual sections describing the development of each consonant phoneme.)

References

Blust, Robert A., 1983–84, Austronesian etymologies — II. *Oceanic Linguistics* 22–23: 29–149.

Capell, A., 1957, *A new Fijian dictionary.* Glasgow: Wilson Guthrie & Company Limited.

Geraghty, Paul A., 1983, *The history of the Fijian languages.* Oceanic Linguistics Special Publication No.19. Honolulu: University of Hawai'i Press.

1986, The sound system of Proto-Central Pacific. In Lois Carrington and S.A. Wurm, eds *Focal II: papers from the Fourth International Conference on Austronesian Linguistics,* 289–312. Canberra: Pacific Linguistics.

Lynch, John, 2002, The Proto Oceanic labiovelars: some new observations. *Oceanic Linguistics* 41/2:493–513.

Pawley, Andrew and Timoci Sayaba, n.d., *Words of Waya: a dictionary of the Wayan dialect of the western Fijian Language.* MS.

Ross, Malcolm D., 1988, *Proto Oceanic and the Austronesian languages of Western Melanesia.* Canberra: Pacific Linguistics.