PACIFIC LINGUISTICS

School of culture, History and Language

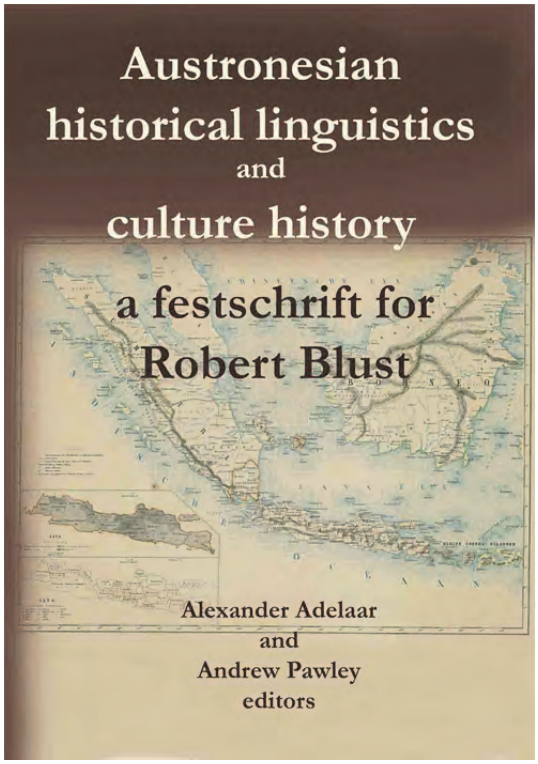
College of Asia and the Pacific

THE AUSTRALIAN NATIONAL UNIVERSITY

Out of print books in pdf format

This PDF document may be copied, printed and stored for use in libraries and for personal use. It may not be reproduced for sale or distribution.

Pacific Linguistics Home Page:
http://www.pacling.com



601 Austronesian historical linguistics and culture history:
a festschrift for Robert Blust

Alexander Adelaar and Andrew Pawley, editors

This book brings together new work on Austronesian historical linguistics and culture history to honour Robert Blust. The memoirs in Part 1 reflect on Blust's groundbreaking contributions to these fields over the last 40 years. The remaining 26 chapters contain contributions by leading Austronesianists on a wide range of topics that broadly match Blust's own research interests. The chapters in Part 2 ('sound change') examine issues in the historical phonology of Austronesian languages. Those in Part 3 ('grammatical change and typology') deal with morphological and syntactic reconstruction at various levels, from Proto Austronesian down. Methodological and substantive issues in the genetic classification of Austronesian languages are treated in Part 4 ('subgrouping') and in several chapters in other sections. Chapters in Part 5 ('culture history and lexical reconstruction') investigate ways in which the close analysis of lexicon, in conjunction with different kinds of non-linguistic evidence, can throw light on the history of Austronesian-speaking peoples.

Several chapters in the volume propose significant revisions to currently accepted reconstructions of PAn phonology and/or morphosyntax. Others focus on the historical development of languages of particular regions, including Taiwan, the Philippines, Borneo, Java, the Strait of Malacca, Sulawesi, the Moluccas, New Guinea, the Solomon Is., Vanuatu, Polynesia and Micronesia.

2009 ISBN 9780858836013 554 pp

Austronesian historical linguistics and culture history:

a festschrift for Robert Blust

Pacific Linguistics 601

Pacific Linguistics is a publisher specialising in grammars and linguistic descriptions, dictionaries and other materials on languages of the Pacific, Taiwan, 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.

FOUNDING EDITOR: Stephen A. Wurm

EDITORIAL BOARD: John Bowden and I Wayan Arka (Managing Editors), Mark Donohue, Nicholas Evans, David Nash, Andrew Pawley, Malcolm Ross, Paul Sidwell, Jane Simpson, and Darrell Tryon

EDITORIAL ADVISORY BOARD:

Karen Adams, *Arizona State University*
Alexander Adelaar, *University of Melbourne*
Peter Austin, *School of Oriental and African Studies*
Byron Bender, *University of Hawai'i*
Walter Bisang, *Johannes Gutenberg-Universität Mainz*
Robert Blust, *University of Hawai'i*
David Bradley, *La Trobe University*
Lyle Campbell, *University of Utah*
James Collins, *Universiti Kebangsaan Malaysia*
Bernard Comrie, *Max Planck Institute for Evolutionary Anthropology*
Soenjono Dardjowidjojo, *Universitas Atma Jaya*
Matthew Dryer, *State University of New York at Buffalo*
Jerold A. Edmondson, *University of Texas at Arlington*
Margaret Florey, *Monash University*
William Foley, *University of Sydney*
Karl Franklin, *SIL International*
Charles Grimes, *SIL International*
Nikolaus Himmelmann, *Westfälische Universität Münster*
Lillian Huang, *National Taiwan Normal University*
Bambang Kaswanti Purwo, *Universitas Atma Jaya*
Marian Klamer, *Universiteit Leiden*
Harold Koch, *The Australian National University*
Frantisek Lichtenberk, *University of Auckland*
John Lynch, *University of the South Pacific*
Patrick McConvell, *The Australian National University*
William McGregor, *Aarhus Universitet*
Ulrike Mosel, *Christian-Albrechts-Universität zu Kiel*
Claire Moyse-Faurie, *Centre National de la Recherche Scientifique*
Bernd Nothofer, *Johann Wolfgang Goethe-Universität Frankfurt am Main*
Ger Reesink, *Universiteit Leiden*
Lawrence Reid, *University of Hawai'i*
Jean-Claude Rivierre, *Centre National de la Recherche Scientifique*
Melenaite Taumoefolau, *University of Auckland*
Tasaku Tsunoda, *University of Tokyo*
John Wolff, *Cornell University*
Elizabeth Zeitoun, *Academica Sinica*

Austronesian historical linguistics and culture history:

a festschrift for Robert Blust

Edited by
Alexander Adelaar and Andrew Pawley

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

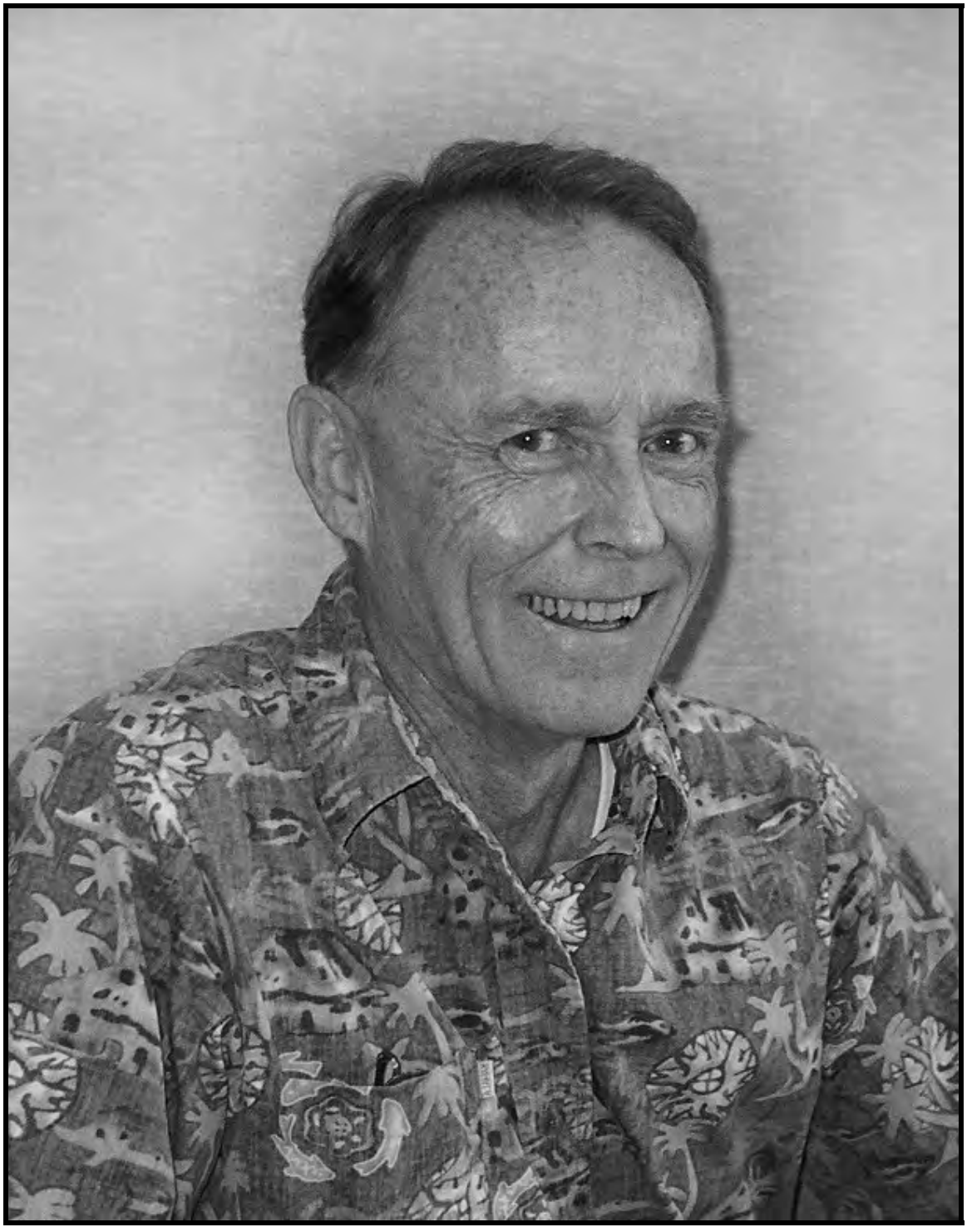
Copyright in this edition is vested with Pacific Linguistics

First published 2009

National Library of Australia Cataloguing-in-Publication entry

Title:	Austronesian historical linguistics and culture history: a festschrift for Robert Blust / Alexander Adelaar and Andrew Pawley, editors.
ISBN:	9780858836013 (pbk.)
Notes:	Bibliography.
Subjects:	Blust, R.A. Austronesian languages. Historical linguistics. Culture history.
Other Authors/ Contributors:	Adelaar, Alexander. Pawley, Andrew. Blust, R.A.
Dewey Number:	499.2

Copyedited by Anthony Jukes, Rachel Hendery and Jeanette Coombes
Typeset by Jeanette Coombes
Cover design by Julie Manley
Cover photo: Map, *Nederlandsch Oost Indië* by J. Kuyper, Leeuwarden, printed by Hugo Suringar (1866)
Printed and bound by Addcolour Digital Pty Ltd, Fyshwick, Canberra



Robert Blust

Table of contents

Contributors to this volume x

Acknowledgements xiii

Part 1: About Bob

1 Reflections on Bob Blust's career
Alexander ADELAAR and Andrew PAWLEY 3

2 Fêting Bob's career to date
Byron W. BENDER 17

3 Thoughts on learning that Bob Blust has reached festschrift age
George W. GRACE 19

4 The publications of Robert A. Blust 23

Part 2: Sound change

5 Structure-preserving sound change: a look at unstressed vowel syncope in Austronesian
Juliette BLEVINS 39

6 Irregular sound change and the post-velars in some Malakula languages
John LYNCH 57

7 In search of an historical Sea-People Malay dialect with *-aba-*
Waruno MAHDI 73

8 The sounds of Southeast Babar
Hein STEINHAUER 91

9 Motherese and historical implications
Shigeru TSUCHIDA 107

10 The Proto Austronesian laryngeal
John WOLFF 115

Part 3: Grammatical change and typology

11 The various origins of the passive prefix *di-*
Alexander ADELAAR 129

12 Relative-clause bracketing in Oceanic languages around the Huon Gulf of New Guinea
Joel BRADSHAW 143

13 The history of the Tukang Besi pronominals
Mark DONOHUE 163

14 Verbal aspect and personal pronouns: the history of aorist markers in north Vanuatu
Alexandre FRANÇOIS 179

15 Austronesian typology and the nominalist hypothesis
Daniel KAUFMAN 197

16 Start and finish: some grammatical changes in Toqabaqita
Frantisek LICHTENBERK 227

17 On the Eastern Polynesian 'direct object' marker *i*
Yuko OTSUKA 247

18 Inclusory constructions and their development in Philippine languages
Lawrence REID 267

19 Proto Austronesian verbal morphology: a reappraisal
Malcolm ROSS 295

20 Possession in South Halmahera — West New Guinea: typology and reconstruction
Rene van den BERG 327

21 Reassessing the reconstruction of plural affixes in PAn: evidence from the Formosan languages
Elizabeth ZEITOUN 359

Part 4: Subgrouping

22 Austronesian language phylogenies: myths and misconceptions about Bayesian computational methods
Simon GREENHILL and Russell GRAY 375

23 Linguistic differences among Siraya, Taivuan and Makatau
Paul JEN-KUEI LI 399

24 Out-of-Borneo subgrouping hypothesis for Rejang: re-weighing the evidence
Richard McGINN 411

25 The position of Makuva among the Austronesian languages in East Timor and Southwest Maluku
Aone van ENGELENHOVEN 425

Part 5: Culture history and lexical reconstruction

26 Words of Eastern Polynesia: is there lexical evidence for the origin of the East Polynesians?
Paul GERAGHTY 445

27 Some clan names of the Chuukic-speaking peoples of Micronesia
Jeff MARCK 461

28 Kinship terms in Bungku-Tolaki languages: inheritance, innovation and borrowing
David MEAD 489

29 Cat's cradle: a disappointing field for lexical reconstruction
Meredith OSMOND 509

30 The role of the Solomon Islands in the first settlement of Remote Oceania: bringing linguistic evidence to an archaeological debate
Andrew PAWLEY 515

Contributors to this volume

ALEXANDER ADELAAR
Asia Institute
University of Melbourne, VIC 3010
Australia
karlaa@unimelb.edu.au

BYRON W. BENDER
Department of Linguistics
University of Hawaii at Manoa
1890 East-West Road
Honolulu, HI 96822
USA
bender@hawaii.edu

JULIETTE BLEVINS
Max Planck Institute for
Evolutionary Anthropology
Deutscher Platz 6
D-04103 Leipzig
Germany
blevins@eva.mpg.de

JOEL BRADSHAW
University of Hawai'i Press
2840 Kolowalu Street
Honolulu, HI 96822
USA
bradshaw@hawaii.edu

MARK DONOHUE
Department of Linguistics, RSPAS
The Australian National University
Canberra, ACT 0200
Australia
mark@donohue.cc

ALEXANDRE FRANÇOIS
LACITO - CNRS
7 rue Guy Môquet
F - 94801 Villejuif
France
Alexandre.Francois@vjf.cnrs.fr.

PAUL GERAGHTY
Department of Literature and Language
University of the South Pacific
PO Box 1168
Suva, Fiji
geraghty_p@usp.ac.fj

GEORGE GRACE
Department of Linguistics
University of Hawaii at Manoa
1890 East-West Road
Honolulu, HI 96822
USA
grageorge@gmail.com

RUSSELL GRAY
Department of Psychology
University of Auckland
Auckland 92019
New Zealand
rd.gray@auckland.ac.nz

SIMON GREENHILL
Department of Psychology
University of Auckland
Auckland 92019
New Zealand
s.greenhill@auckland.ac.nz

DANIEL KAUFMAN
Department of Linguistics,
Graduate Center
City University of New York
Mailing address: 534 LaGuardia Pl., NY NY 10012, USA
dak37@cornell.edu

PAUL JEN-KUEI LI
Institute of Linguistics, Academia Sinica
128, Sec. 2, Academia Road
11529 Nankang, Taipei,
Taiwan
paulli@gate.sinica.edu.tw

FRANK LICHTENBERK
Department of Applied Language Studies and Linguistics
University of Auckland
Auckland 92019
New Zealand
f.lichtenberk@auckland.ac.nz

JOHN LYNCH
University of the South Pacific
PMB 9072
Port Vila
Vanuatu
johnlynch123@gmail.com,
lynch_j@vanuatu.usp.ac.fj

WARUNO MAHDI
Fritz-Haber-Institut der Max-Planck Gesellschaft
Faradayweg 4-6
14195 Berlin
Germany
mahdi@fhi-berlin.mpg.de

JEFF MARCK
Department of Linguistics, RSPAS
The Australian National University
Canberra, ACT 0200
Australia
jeff@jeffmarck.net

RICHARD McGINN
Department of Linguistics
383 Gordy Hall
Ohio University
Athens, OH 45701
USA
mcginn@ohio.edu

DAVID MEAD
Sulawesi Language Foundation
director@sulang.org

MEREDITH OSMOND
Department of Linguistics, RSPAS
The Australian National University
Canberra, ACT 0200
Australia
meredith.osmond@anu.edu.au

YUKO OTSUKA
Department of Linguistics
University of Hawai'i at Manoa
1890 East-West Road
Honolulu, HI 96822
USA
yotsuka@hawaii.edu

ANDREW PAWLEY
Department of Linguistics, RSPAS
The Australian National University
Canberra, ACT 0200
Australia
andrew.pawley@anu.edu.au

LAWRENCE A. REID
5-1-1-1701 Nyoidani
Minoo-shi
Osaka 562-0011
Japan
reid@hawaii.edu

MALCOLM ROSS
Department of Linguistics, RSPAS
The Australian National University
Canberra, ACT 0200
Australia
malcolm.ross@anu.edu.au

HEIN STEINHAUER
Faculty of Humanities, Indonesian Studies
Leiden University
PO Box 9515
2300 RA Leiden
The Netherlands
H.Steinhauer@hum.leidenuniv.nl

SHIGERU TSUCHIDA
4-42-1 Kugahara
Ota-ku, Tokyo
Japan 146-0085
tshige@beige.plala.or.jp

RENÉ VAN DEN BERG
SIL International
PO Box 226
Ukarumpa EHP 444
Papua New Guinea
r.vandenberg@sil.org.pg

AONE VAN ENGELENHOVEN
Faculty of Humanites
Leiden University Centre for Linguistics
Leiden University
PO Box 9515
2300 RA Leiden
The Netherlands
aonevanengelenhoven@gmail.com

JOHN U. WOLFF
Department of Linguistics
203 Morrill Hall
Cornell University
Ithaca, NY 14853-4701
juw1@cornell.edu

ELIZABETH ZEITOUN
Institute of Linguistics, Academia Sinica
128, Sec. 2, Academia Road
11529 Nankang, Taipei
Taiwan
hsez@gate.sinica.edu.tw

Part 1

About Bob

1 *Reflections on Bob Blust's career*

ALEXANDER ADELAAR and ANDREW PAWLEY

1 A Blustery gale hits Austronesia

At the end of the 1960s Austronesian historical linguists felt the onset of a fresh breeze blowing out of Hawai'i. The breeze quickly became a strong wind and then grew to gale force, blowing new data, analyses, syntheses, critiques, and other objects across the Austronesian world and beyond.

This was a welcome gale. For a good many years the higher levels of Austronesian comparative linguistics had been drifting in the doldrums. Since the groundbreaking reconstructions of early Austronesian phonology and lexicon made by Otto Dempwolff in the 1920s and 30s, not a single new lexical reconstruction had (to our knowledge) been attributed to Proto Austronesian. There had been some significant refinements in the understanding of early Austronesian historical phonology, chiefly through the work of Isidore Dyen, but as far as reconstruction of lexicon was concerned it was as if Dempwolff had exhausted the possibilities. There had been no systematic attempts to reconstruct morphological paradigms or syntax. And in the area of high-order subgrouping, critical to reconstructive work, Dempwolff's proofs for an Oceanic subgroup remained the only secure achievement. There was no consensus about the identity of the primary subgroups of Austronesian and, as a consequence, no agreement about the most likely precise location of Proto Austronesian itself and the initial directions of the Austronesian dispersal.

All that changed in the 1970s and the main agent of change was the storm emanating from Hawai'i. The name of this phenomenon was Robert (Bob) Blust.

Several decades have passed since the 1970s but the force of the gale has not diminished. On the contrary it has grown stronger and has shaken up almost every corner of Austronesian historical studies. When we sat down and classified Bob's more than 200 publications by subject matter it turned out that those reporting his own research can be divided into about eight main categories according to their primary focus: (1) historical phonology (mainly but not only Austronesian), (2) Austronesian subgrouping, (3) Austronesian etymologies and semantic reconstruction, (4) Austronesian historical morphology and syntax, (5) culture

Alexander Adelaar and Andrew Pawley, eds
Austronesian historical linguistics and culture history: a festschrift for Robert Blust, 3–15.
Pacific Linguistics, 2009.
© This edition vested with Pacific Linguistics.

history, (6) descriptions of Austronesian languages, (7) general surveys of Austronesian languages, and (8) method and theory. In addition, there are: (9) edited volumes, (10) reviews and comments, and (11) miscellanea (e.g. obituaries, personal memoirs, etc.).

Most of us slow down in our sixth and seventh decades. Bob's rate of publication since he turned 55 (averaging eight publications a year) is double what it was when he was in his vigorous thirties.

In this introductory chapter we will reflect on the ups (and occasional downs) of Bob's career, commenting on his *modus operandi* as a scholar and speculating on what makes him tick. First of all, how does he manage to produce such a volume of high quality publications year after year? Does he have time to do anything else? Then one has to ask: Why does he do it? What demons drive him? What are the questions he is seeking answers to?

But let us not run ahead of ourselves. First, mention should be made about where Bob has been located during different stages of his life and career.

2 Chronological notes

Robert Andrew Blust was born in Cincinatti, Ohio, on May 9, 1940. When he was four the family moved to Long Beach, California where Bob's father worked as a sheet metal mechanic at the naval shipyard and Bob's mother as secretary to the Commander of Long Beach Naval Station.

It is the good fortune of Austronesian linguistics that Robert Blust did not become a biologist, anthropologist, novelist or poet by profession. As a small boy Bob was often left to play by himself after school in a field near the family home and became fascinated with the animals and plants to be found there. He began to read widely about natural history and to collect minerals. At the age of nine or ten he entered an American Indian phase, reading every book in the local library about Native Americans and their cultures and, over the next several years he wrote short stories and several novels about them. (Having realized how badly Native Americans had been treated by the US Government. he got into trouble at school for refusing to do the Pledge of Allegiance.) Between 15 and 17 he was a prolific writer of poems and at 17 won a national prize for poetry in a field of 115,000 entrants.

On leaving high school Bob spent several years looking for his niche, without finding it. He studied English, French and German at junior colleges, entered a pre-Med program at UC-Riverside and moved to San Francisco wanting to be a writer. Then Uncle Sam called. We must give the US Army much of the credit for Bob's ultimate choice of career. When he did his military service in 1963–66 he was mainly based in Hawai‘i, attached to the Signal Battalion in the 25th Division at Schofield Barrack. After a month or two working in the motor pool he found it boring and asked to be placed in the language division. The first language that came to Bob's notice was Indonesian. He began a training in this, the instructor being an Indonesian army officer recruited for the purpose, and was deployed to Jakarta for two months, where he greatly increased his fluency in the language. Meanwhile, the US Government had cut its ties with the Sukarno administration and the Indonesian army officer was no longer available to be an instructor. When Private First-class Blust came back to Hawaii he was sent to the University of Hawai‘i to take advanced courses in Indonesian, before returning to the language division to be an instructor, a job he did for the rest of his military service.

On leaving the army Bob resumed his studies as an undergraduate at the University of Hawai‘i, where he majored in Anthropology, completing a BA with honors in 1967. But

the seeds of a love affair with Austronesian languages had already been planted and in September 1967 he entered the graduate program in linguistics at the University of Hawai'i at Manoa. The Department of Linguistics there had been going for just a few years but was already the major centre for the study of Austronesian languages of Oceania, with a rapidly growing Faculty that included Oceanic specialists such as George Grace, Samuel Elbert, Byron Bender, Albert Schütz and (for a time) Bruce Biggs, and was beginning to establish a strength in Philippine linguistics.

It can't have been long before the Department realised that they had someone special. When you first meet Bob Blust socially he may seem like an ordinary fellow, who will talk to you about his children, baseball and the high cost of housing. But once he starts discussing Austronesian historical linguistics you soon see that this is someone who not only has a great passion for the subject but combines a truly awesome command of data and theory with a formidable clarity and fluency of thought. It was no different when he was a graduate student (the second editor, who joined the UH Faculty in 1973, speaks from first hand experience but see also other memoirs in this volume.)

Of course, it helped that Bob had distinguished teachers, including George Grace as his principal mentor, but their recollection is that he became an accomplished Austronesianist almost overnight. He devoured much of the Austronesian literature swiftly and early on in his graduate studies began churning out papers that were technically sophisticated and highly polished. By the time he finished his PhD he had a dozen or so in print. Far from being planned in cold blood to enhance his CV, these articles were the result of Bob being seized by excitement at some new discovery in the course of his research. Indeed, George Grace recalls being a bit concerned that Bob was writing so many articles he didn't have time to finish his doctoral thesis, which after all would be his main ticket to a career in linguistics. Another distraction was Bob's romance with Elaine Holder, a graduate student from Brazil, which led to a lengthy visit to Brazil and to marriage.

It was during his years as a graduate student that Bob undertook his first spell of linguistic fieldwork: some six months in Sarawak in or about 1969–70. There he collected extensive wordlists and phonological and grammatical data on more than 40 languages of Borneo, most of them previously largely undescribed. This fieldwork furnished crucial data for his doctoral thesis, which focused on the phonological history of the North Sarawak subgroup, and for many subsequent descriptive and comparative studies.

In 1974, after completing his doctorate, he travelled with Elaine to Australia to take up a postdoctoral fellowship in the Research School of Pacific Studies at The Australian National University in Canberra. Unlike most comparative Austronesianists, who specialize in one branch or region of this enormous language family, Bob had read the literature on every branch. However, he had not previously had first-hand encounters with Oceanic languages and he now took the opportunity to do so. Three months very productive fieldwork in Papua New Guinea, mainly based in Manus, yielded quite extensive lexical materials and some grammatical data for each of the 25 or so Austronesian languages of the far flung Admiralty Is. group and on several other languages. His informants were principally older high school students in Manus. These materials provided the basis for a monograph on the history of the Proto Oceanic palatal consonants and for a number of descriptive and comparative papers on languages of the Admiralties and nearby regions.

After two years in Australia Bob took a teaching job at Leiden University, in the Department of Languages and Cultures of Southeast Asia and Oceania, moving there

towards the end of 1976. At Leiden he mainly taught Austronesian comparative–historical linguistics. In a program that was predominantly oriented to philology and literature his courses did not attract too many students but those who joined his classes were inspired by his teaching. Among them was the first editor of this volume, who wrote his PhD thesis under Bob's supervision. Others included René van den Berg and Aone van Engelenhoven, both contributors to this volume. There were also students visiting from other universities, such as Hans Schmidt from the University of Hamburg, who studied Rotuman, and James Collins, who was finishing a PhD thesis at the University of Illinois on the classification of Central Maluku languages.

Another of Bob's duties was teaching and supervising Indonesian PhD students in the Indonesian Language Development Project, both in Bogor (Indonesia) and Leiden. He taught them in Indonesian, which he spoke well.

A daughter, Lani, had been born to Bob and Elaine during their stay in Canberra. A second daughter, Karen, arrived early in their time in Leiden.

It took time for Bob to get used to ways of doing things in the Dutch academic world but his eight years at Leiden were very productive. It was during this period that Bob wrote several of his most important contributions to the culture history (or, if you prefer, the historical anthropology) of Austronesian-speaking societies. Although, for some reason Bob never became fluent in spoken Dutch, he was an avid reader of Dutch scholarly works in Austronesian linguistics and Indonesian structural anthropology. Through his study of Austronesian social organization he embarked on lively discussions with various anthropologists, which have continued. His keen understanding and appreciation of Dutch scholarship in these areas became obvious at a symposium in Leiden in 1983, dedicated to the anthropologist J.P.B. de Josselin de Jong and his concept of an 'Indonesian field of anthropological study'. Bob was able to show how the genetic relationship between Austronesian languages was fundamental to J.P.B. de Josselin de Jong's concept of an anthropological study field, an insight which, as it turned out, had escaped some of the organizers of the symposium.

In 1984 Bob returned to the Department of Linguistics, University of Hawai'i at Manoa, as an Associate Professor. Whereas the 1960s and 70s had been a golden era for Austronesianists at the University of Hawai'i, with plenty of federal funding for major descriptive and comparative research projects and a steady flow of outstanding students keen to work on Austronesian languages, the economic and intellectual climate had changed. Bob returned to a Department where relatively few graduate students chose to specialize in Austronesian linguistics. However, his presence as a teacher and researcher has helped the University of Hawai'i to remain a force in this field. Bob has remained at the UH, becoming Full Professor in 1987 and serving as Chair from 2005–08.

In 1994 Bob spent a sabbatical year in Taiwan, based at Academica Sinica in Taipei. This allowed him to do fieldwork on several Austronesian languages of Taiwan, including Kavalan, spoken on the east coast, and Pazeh and Bunun in central Taiwan. Later in the year a friend from Academica Sinica took him to the village at Sun Moon Lake in central Taiwan where the last few fluent speakers of the Thao language reside. He began to gather materials for a dictionary of Thao and returned to Taiwan in the summers of 1995, 1996 and 1999 to continue this project (more on this in §8).

His choice of sabbatical location was no doubt related to the central role of the Taiwan languages in reconstructing the history of Austronesian but there was another, more personal factor in his choice. Some years earlier Bob and Elaine had parted ways. In 1992 he had married Laura Chang, who hailed from Taiwan and had done her PhD in linguistics at the University of Hawai'i. Laura kept her job in Taiwan for some years, until their daughter, Jasmine, was born in 1999.

It is time now to reflect on Bob's major achievements as a scholar. There are many and needless to say, we can only skim the surface.

3 The Austronesian family tree

We mentioned that as recently as 1970 Austronesianists didn't have much idea about the higher-order subgrouping of the family. As a working hypothesis many scholars assumed a primary split between an Eastern branch (by then generally called Oceanic), containing the Austronesian languages of Melanesia, Polynesia and Micronesia, and a Western branch, containing the rest. But the notion of a Western Austronesian group was always problematic because no innovations uniting these languages had been noted. (In the early 1960s Isidore Dyen had proposed a radically different classification, based on lexicostatistical evidence, which recognised some 40 first-order subgroups and isolates. The fact that more than 30 of these putative primary branches were in Melanesia was taken as evidence that Melanesia was the primary dispersal centre of Austronesian. However, although the radical parts of this classification were taken seriously by a few outsiders, they cut no ice among Austronesian historical linguists.)

The subgrouping picture changed dramatically in the 1970s, by which time several linguists had began to look more closely at some of the little-known Austronesian languages of Formosa (now known as Taiwan). Many of these languages had died out following the colonization of Taiwan beginning in the 17th century but 14 have survived. An important breakthrough came with a short, dense monograph published in 1973 by the Norwegian scholar Otto Christian Dahl. Dahl argued, chiefly on phonological grounds, for a primary division between one or more Formosan groups on the one hand, and a vast 'Extra-Formosan' group into which all other Austronesian languages fall.

At this point Bob Blust entered the fray. Over the next two decades he wrote a series of papers that, taken together, provided evidence for a complex sequence of splits in the roots of the Austronesian family tree (bearing in mind that historical linguists perversely invert their tree diagrams so that the roots sit at the top.) He added to and refined the set of innovations defining the putative Extra-Formosan group, which he renamed 'Malayo-Polynesian' (MP). He argued for several new intermediate nodes within MP, reducing Oceanic to the status of a fourth-order subgroup of Austronesian. He proposed a group, Eastern MP, which unites Oceanic, comprising more than 450 languages, with about 50 Austronesian languages of south Halmahera and west New Guinea. He then argued that Eastern MP forms a higher-order group with Central MP, a putative subgroup of about 150 languages centered in the Moluccas. He called this huge higher-order group 'Central-Eastern MP'.

The remaining MP languages spoken in the Philippines, Indonesia, Malaysia and Madagascar, Blust collectively termed 'Western MP', not because they were demonstrably a subgroup but as a convenient geographic label for some 500 MP languages whose high-order relationships remain unclear. Along with certain other scholars he argues for a

Philippines subgroup comprising almost all the languages of the Philippines plus certain languages of north Sulawesi. He also worked on the relationships between Malay, Acehnese, Chamic and Moken-Moklen, and between Malay, Javanese, Sundanese and Madurese.

Perhaps his most startling claim, advanced in the 1990s, has been that the 14 surviving Taiwanese languages fall into as many as nine different branches sharing no common ancestor more recent than Proto Austronesian. Such a distribution of subgroups would overwhelmingly favour Taiwan as the location of PAn.

Not all of Bob's high-order groups are equally well accepted. He himself recognizes that some (particularly Central-Eastern MP and Central MP) rest on more tenuous grounds than others. This situation is unsurprising given that (a) Malayo-Polynesian speakers appear to have spread very far and very fast in the 2nd millennium BC, leaving little time for well-defined intermediate nodes to form on the tree, and (b) in certain regions complex dialect chains formed within which innovations spread unevenly. Still, it is fair to say that the Blust subgrouping has formed the point of departure for all subsequent work on the roots of the Austronesian tree. The case for Taiwan as the primary dispersal centre for Austronesian is now generally accepted by specialists in this field. See, for example, Malcolm Ross's paper in this volume, which argues for a slightly more complex layering of subgroups in Taiwan, one that even more emphatically favours Taiwan as the homeland.

Bob's subgrouping efforts have not been confined to the higher levels. He has for instance written a number of papers treating the subgrouping of Borneo languages and several works arguing for subgroups within Oceanic. However, it is the high-order groupings that have been central to attempts to reconstruct the direction and dating of the dispersal of the Austronesian languages, and that in some cases are critical in choosing between competing hypotheses about the phonological shape of PAn and PMP etyma.

One particular subgrouping controversy, not of Bob's making, led him to develop a simple but very ingenious method of assessing the value of lexicostatistical evidence. We mentioned above that in the 1960s Isidore Dyen put forward the controversial hypothesis that Melanesia was the most probable primary dispersal centre of Austronesian, given that this was the region where by far the greatest concentration of groups and isolates sharing very low cognate percentages with all other Austronesian languages is to be found. During the 1970s Dyen continued to maintain that the lexicostatistical evidence was a serious competitor to the evidence of shared innovations in phonology, which points to a very different classification. At the 3rd International Conference on Austronesian Linguistics in Bali in 1981, Bob responded by taking as the starting point lexical reconstructions attributable to Proto Malayo-Polynesian (PMP) for the Swadesh 200 meaning list (plus about 50 other meanings) and showing that Malayo-Polynesian languages vary greatly in their retention rates. By far the lowest retention rates are to be found in Melanesia. It should be added that comparison of diverse languages from different regions of Melanesia will yield a set of lexical reconstructions that matches closely the PMP set. The problem now becomes to explain why some languages have replaced their basic lexicon much, much faster than others. In the case of Melanesia there are some obvious possible factors.

4 Lexical reconstruction and the Austronesian comparative dictionary project

In the final volume of his *Vergleichende Lautlehre des austronesichen Wortschatzes* Dempwolff reconstructed some 2000 roots to a level he called Proto Austronesian, but

which we would now call Proto Malayo-Polynesian or Proto Extra-Formosan. It remains something of a mystery why, in the 30 years after Dempwolff's 1938 study, no new etymologies were proposed at this level. It is true that the pool of scholars working in Austronesian historical linguistics during this period was small but cognate sets supporting additional reconstructions would not have been hard to find. One has to conclude that people were simply reluctant to look for them. Bob had no such inhibitions. His first set of new etymologies, 443 of them, was published in 1970, followed by further substantial additions in 1972 and 1973, while he was still a graduate student, with further additions in 1980, 1983–84, 1986 and 1989. These works more than doubled the body of PMP etymologies to be found in Dempwolff as well as adding many new cognate sets attributable to other levels, such as PAn and Proto Oceanic (on which see below).

How did he do it? By old-fashioned methods. No computer-assisted searches. Bob's lexical database was obtained by going through dictionaries by hand, recording cognate sets on filing cards. One gets the impression that instead of settling down to watch a movie or read a novel in the evenings, Bob has always found it more relaxing to sit with a pile of dictionaries and a box of cards, entering cognate sets. Bob's memory for lexical information and sound correspondences is so good that this method probably does the job better than any computer-based search program could.

However, because his reconstructions with supporting cognate sets were scattered across many journal articles it made sense to try to consolidate them. Around 1990 Bob began a NSF-funded project which aimed to bring together all his own reconstructions and those of other contemporary scholars in an Austronesian comparative dictionary. The scale of the project was massive. Bob attempted a complete search of sources for about 120 languages, with reconstructions not only for PAn but for eight lower-order proto-languages, and projected a final product with more than 7000 main entries (exclusive of affixed and reduplicated forms), running to more than 4000 singe-spaced pages.

By the early 1990s tools for creating and searching large electronic databases were available. Bob was a somewhat reluctant convert to Internet technology but he undertook to create an electronic version of the dictionary that colleagues could access by application. By 1995 the database contained entries for over 5100 bases, together with many more reconstructed words built up by affixation, reduplication or compounding, along with added commentary. This amounts to about half of Bob's total card file. And there it remains until Bob finds the time to resume. In the meantime, the Austronesian comparative dictionary, though still incomplete, has been an invaluable resource for a number of major projects by other scholars.

5 Sound change and phonological reconstruction

The study of sound change is a domain of historical linguistics that is largely impenetrable to non-specialists but which underpins all reconstructive and subgrouping work. A great deal of Bob's work—more than 40 substantial papers and some monographs—has been in this field. He has done careful case studies of sound change in many languages of Indonesia and Malaysia but also in certain languages of Taiwan, the Philippines, Polynesia, Melanesia and Micronesia, as well as doing studies seeking to test hypotheses about universal principles governing sound change. The titles of a few representative papers give some idea of the technical nature of his contributions: 'Chamorro historical phonology'. 'A Tagalog consonant cluster conspiracy', 'A double counter-

universal in Kelabit', 'Ablaut in Northwest Borneo', 'Coronal-noncoronal consonant clusters: new evidence for markedness', 'More on the origins of glottalic consonants', 'Some remarks on stress, syncopation and gemination in Mussau' and 'Anti-gemination: canonical constraints in Austronesian languages'. Enough said?

There have been at least three schools of thought about how best to handle reconstruction of the phonemic inventory of PAn. There is what we might call the formulist school, espoused by Isidore Dyen. This school is not much concerned with how realistic the reconstructed phonological system is, but focuses on the task of identifying all sets of sound correspondences that are at least partly distinct, and assigning a separate symbol (or symbol with subscript) to each such correspondence set. Other schools place more weight on figuring out how the attested sound correspondences fit into a realistic phonological system. There is the minimalist approach, espoused by John Wolff, which seeks to reduce the number of protophonemes posited by Dempwolff and Dyen and create a symmetrical phoneme system. It does so by arguing that some putative proto-phonemes were allophones conditioned by the environment, others reflect secondary developments (that is, borrowing or analogical changes) and still others reflect errors in the way the proto-form has been reconstructed. Blust's approach lies in the middle. He is too much concerned with the phonetic value attributed to reconstructed phonemes but also with constraints on phonotactic patterns. He places great weight on the disyllabic syllable structure and consonant configurations that are characteristic of the more conservative living Austronesian languages, while using doublets as a way to deal with certain unexpected sound correspondences that seem not to warrant positing separate proto-phonemes. Bob's approach to phonological reconstruction is continuous with his interest in processes of sound change.

6 Culture history

It may surprise some who think of Bob solely as a hardcore linguist to find among his *opera* titles such as: 'The origins of dragons', 'The fox's wedding', 'Pointing, rainbows and the archeology of mind', 'The limits of the thunder complex', 'Rats ears, tree ears, ghost ears and thunder ears in Austronesian languages' and 'Linguistic evidence for some early Austronesian taboos'. But it happens that, except for historical phonology, Bob has written more papers on culture history than any other domain.

His work in this field is diverse. The above-mentioned set of papers all deal with beliefs about the supernatural which occur in a wide range of societies around the world, but which initially came to Bob's attention in the course of his comparisons of Austronesian languages. The papers all reflect his longstanding concern with the more general methodological problem of how one can arrive at reliable culture-historical inferences by studying the distribution of cultural data (other than linguistic and archaeological data). The main problem is how to separate common heritage from parallel development (convergence) and borrowing.

In the case of beliefs about the supernatural, Bob tackled this methodological problem by first charting the distribution of beliefs and practices, such as the widespread taboo on pointing at rainbows and other elements of the rainbow complex, across regions and language families and then assessing the chances that particular distributions were due to one or another determinant.

In researching the nature and distribution of particular beliefs Bob has shown formidable enterprise and thoroughness. For example, in quest of beliefs about rainbows he not only patiently searched through a vast anthropological and folklore literature but constructed questionnaires which he sent to umpteen institutions and individuals on every continent. And he buttonholes every potential informant he encounters. (On a visit to Hawai'i, when AP was travelling with a man from the highlands of New Guinea, both were invited to lunch by Bob. Bob soon turned the conversation to rainbows and was delighted when this man said of course bad things will happen to you if you point at rainbows.)

Another set of Bob's papers have to do with reconstructing the movements of the Austronesian speaking peoples and the way of life of early Austronesian speaking societies. The two disciplines that have most to say about these matters are archaeology and historical linguistics. Bob has been at pains to point out ways in which evidence from these two disciplines can be complementary, corroborative or contradictory. Bob has spelled out many inferences about perishable items of material culture, social organization and belief systems of Proto Austronesian and Proto Malayo-Polynesian speakers that can be drawn from lexical reconstructions, inferences that cannot be obtained from archaeological assemblages, which are typically restricted to non-perishable materials. And we have already referred to his use of subgrouping in determining the most likely homeland of the Austronesian family.

Among Bob's papers on the Austronesian diaspora, one of his most ingenious is 'The linguistic value of the Wallace Line'. This uses a major biogeographical boundary and the distribution of terms for animals confined to one or other side of this boundary as evidence for subgrouping and for determining directions of population movements during the early stages of the Austronesian linguistic expansion.

It must be gratifying to Bob to see that archaeological research, especially over the past 20 years, has yielded evidence consistent with the Taiwan homeland hypothesis, and has provided absolute dates for particular archaeological events that can be correlated with particular stages in the dispersal of Austronesian languages. It is now well established that Neolithic cultures first appear in Taiwan about 3500 BC, and develop regional variants there, before expanding southwards, first into the Philippines around 2000 BC and then to various parts of the Indo-Malaysian archipelago and across Melanesia in the 2nd millennium BC.

Bob has had less success convincing social anthropologists to accept his conclusions about early Austronesian social organization. In 1980 he published the first of a series of papers dealing with aspects of Proto Malayo-Polynesian society, using a conjunction of social structure typology and the distribution of cognate sets to reconstruct particular types of social groups, and kinship and marriage systems. The seminal paper in this series, 'Early Austronesian social organization: evidence from language', appeared in *Current Anthropology* and stirred up a quite a storm of criticism from the anthropologists. Contrary to the well-known work of the anthropologist George Murdock, Bob concluded that PMP society had unilineal descent groups and a rule of prescriptive matrilateral cross-cousin marriage. Central to his argument was the reconstruction of distinctive terms for cross-siblings ('brother (woman speaking)' and 'sister (man speaking)' and the similar manner in which these terms were replaced, independently, in various daughter languages ('cross-sibling substitution drifts').

He developed some of his arguments further in other papers but continued to meet with stiff resistance from—or worse, non-comprehension by many (not all) anthropologist colleagues. In 1992 Bob wrote to AP in rather despondent terms:

> My 'sibling terms' paper was kind of rejected by the *Bijdragen* … The sole criticism that I could make out … was that I am using 'old fashioned' kinship models (Lévi-Strauss is evidently now passé). I can't honestly see that the models make a lot of difference to my conclusions. All that needs to be agreed on is that wife-giving groups have superior status to wife-taking groups, and are to be regarded as classificatory 'males' (in dualistic classification schemes), and this is so redundantly attested in the ethnographic literature that it is pointless to argue about it. The rest of my argument is basically linguistics (why two sets of cross-sibling terms? why is one analyzable as *(child) + male/female, and why does this set appear to have the character of a drift in diachronic perspective?)
>
> The real problem so far as I can see is that no anthropologist who I have spoken to has understood the argument. It just goes past them. Their eyes glaze over somewhere around the third or fourth minute of my sometimes overly enthusiastic explanation, and then I know it is a lost cause.
>
> What gets me down a bit is that I believe more than ever … that the … cross-sibling substitution drift argument … is one of the most original and discipline bridging contributions of my entire career. Yet I always seem to lose the linguists with half of the argument, and the social anthropologists with the other half. I'll keep trying.

And he has.

7 Semantic reconstruction

To draw reliable inferences about prehistoric cultures from lexical reconstructions one needs reliable semantic reconstructions. A number of Bob's papers on early Austronesian technology and social organization focus on tricky problems of semantic reconstruction and methodologies for resolving such problems.

This was yet another domain where Bob crossed swords with Isidore Dyen on an issue of methodology. The basic problem was how to decide what a particular reconstructed lexical form meant when there is disagreement among daughter languages. For example, what if five different forms can be reconstructed for a proto-language, all of which mean 'house' in certain daughter languages but mean something else in other daughter languages? Must we conclude that the proto-language had five synonyms for 'house'? If languages A, B, C and D have cognate forms meaning 'town', 'garden', 'fence' and 'fortified place', respectively, how does one decide what the ancestral form meant? Together with an anthropologist, David Aberle, Dyen had written a book in which they proposed a rigorous mechanical procedure for arriving at what they called 'lexical reconstructions', a procedure that depended essentially on the distribution of form–meaning pairs across subgroups. You take a certain meaning, say 'brother', as given in the proto-language and then look for the cognate set whose distribution can be most strongly associated with this meaning.

In 1987 Bob published a brilliant paper 'Lexical reconstruction and semantic reconstruction: the case of Austronesian 'house' words' that addressed these questions. Unsurprisingly, he regarded the mechanical procedure proposed by Dyen and Aberle as putting the cart before the horse, because it began by taking the meaning as a given, when this should properly be the object of enquiry. The danger is that, as with lexicostatistics,

such apparently simple 'rigorous' methods based on problematic assumptions are likely to appeal to scholars in other disciplines, who lack a deep understanding of historical processes in language but want quick results. The results are likely to be untrustworthy.

For Proto Malayo-Polynesian (PMP) one can reconstruct at least five different lexical forms each of which has reflexes meaning 'house' in a number of daughter languages. At least three of these etyma would have to be glossed 'house' following Dyen and Aberle's method. However, each PMP form also has reflexes with a number of other meanings in different daughter languages, and often even in the same language, i.e. the cognate sets have different 'semantic profiles'. Blust's approach was to redefine the task as reconstructing the meaning, not of a single etymon treated in isolation, but of a system of related (and often polysemous) PMP terms representing an extensive semantic field, terms which in turn were embedded in a certain kind of cultural milieu (social and economic organization, settlement patterns, types and functions of buildings, etc.) This approach provides no mechanical procedure but in the case of the five 'house' words it enables a well-informed scholar to arrive at a fairly convincing set of semantic reconstructions and semantic changes for the total system, convincing in that it allows the semantic profiles of each cognate set be derived by a series of natural steps.

8 On fieldwork and Purgatory

All scholars who have had long careers are bound to spend time in Purgatory, where, to attain a state of grace they will be required to finish uncompleted projects of genuine value. Fieldworking linguists, for instance, will have to publish satisfactorily detailed reports on the primary data they have collected in the field or at the very least archive it so that future generations may use it. However, Bob's time in Purgatory is likely to be shorter than most. This is partly because he works fast but also because he has been more diligent than most in writing up the field data he has collected.

Bob is by no means solely an armchair linguist. He has gathered first-hand data on about 100 languages, most of them Austronesian languages spoken in Borneo and Melanesia. He has published descriptions of nine Borneo languages, two Oceanic languages and one Taiwanese language.

Bob's fieldwork, has, as far as we know, never been of the classical type in anthropological linguistics, in which the researcher spends long periods immersed in the community, becoming fluent in the language with the aim (not often fully achieved) of producing not only a lengthy reference grammar and annotated texts but a large dictionary. Instead, most of his work on single languages has occurred as part of extensive regional surveys, to collect data useful for comparative purposes. Typically he has worked intensively with one or a few informants for a few days, before moving on to another language. For example, during three months fieldwork in the Admiralty Islands in 1975 Bob collected extensive data from informants on some 30 languages: about 1000 words (each phonetically transcribed, it seems), plus grammatical data, on each. That amounts to a language every three days, without even allowing for time lost to other factors. One marvels not only at his own endurance but also at that of his informants. His fieldwork in Sarawak seems to have followed a similar pattern.

There is a notable exception to this generalization. In 2003 Bob published his dictionary of Thao, a highly endangered language of central Taiwan, referred to earlier in this memoir. This formidable tome of more than 1000 pages, containing some 13,000 entries

and sub-entries, is the outcome of fieldwork done between 1994 and 1999, using interpreters. Often Bob had to use a chain of interpreters, first English-Mandarin, then Mandarin-Taiwanese, as the elderly informants could speak only Taiwanese besides their native Thao. We don't know how many hours the compilation and production of this book took but Bob did the data-collecting part in just 451 'contact hours' with his informants, an amazingly short time for such a massive work. How do we know this detail? Well, Bob keeps records of such matters and in his CV enumerates contact hours for particular languages or sets of languages.

9 The dangers of trying to do many things at once and of knowing too much

In spite of his astonishing productivity, even Bob has sometimes bitten off more than he can chew. Which brings us to the saga of his book on the Austronesian family. After nearly 30 years the saga has had a happy ending (happy at least for Bob) but shows that trying to do many things at once can lead to major road blocks and that too much knowledge can be a dangerous thing.

Back in 1980 one of the editors of the present festschrift was approached by Cambridge University Press to write such a book as part of that publisher's series on language families of the world. He declined, and counseled CUP that the Austronesian family was so large and diverse, and the literature so extensive, that it needed at least two specialists to cover the ground, one to treat the large Oceanic subgroup and one or more others to deal with the remaining languages. He added that there was only one person with the breadth of knowledge to tackle the whole family on his own, Robert Blust, and that even he might well decide the task was too big.

CUP duly invited Bob to do the book alone and he accepted. It seems a limit of some 400–500 pages was set, which at the time must have seemed ample. Within a year Bob had drafted almost 300 pages but then got distracted by a succession of other projects. One constant kind of distraction was what we might call the 'spin-off' paper. In the course of doing a paper on topic A, Bob notices some things that suggest another exciting paper on topic B, and in the grip of enthusiasm, it becomes impossible for him to resist writing this paper, which in turns suggests a further paper, and so on. Of course this happens to all of us to some degree but Bob generates spin-off papers at a rate that few others can match.

At any rate, between one thing and another, twenty years went by before Bob made a really concerted effort to finish the book. But he soon ran into the problem of knowing too much and trying to fit it all in. He wrote to AP that:

> The … book has been going extremely well. But, as you guessed, it is getting long … I have about 445 pages on the computer, and am nearing the end of Chapter 5 ('Morphology'). But the blueprint calls for 12 chapters ... I've written very little on historical stuff yet—it's been mostly typology, with critical asides on various theoretical proposals. The earlier stuff I wrote has been completely transformed, and I am trying to include something in this book that will be of interest to almost everyone in the field (theoretical phonologists, typologists, cognitive linguists, general historical linguists, sociolinguists, etc.).

When Bob finally offered the manuscript to CUP in 2005 it had grown to around 900 pages. Not surprisingly, Cambridge asked for substantial cuts (though as it happened, it had in the meantime published some very large books on other language families and regions that contain far fewer languages than Austronesian does and that lack the kind of

detailed historical work that has been done for Austronesian). Bob reluctantly made some cuts but they were not enough for Cambridge and an impasse was reached. To cut a long story short, it fell to Pacific Linguistics to publish the book. Given Bob's encyclopaedic knowledge of issues in Austronesian linguistics and how these bear on concerns of general theoretical interest, and given his penchant for going into fine detail in the cause of being thorough, we are surprised that, in the end, he was able to restrict himself to a mere 800 pages.

10 Blust the correspondent, reviewer and biographer

You might think that Bob has managed to write all these books and papers only by exclusively devoting himself to his own research. Far from it. He has always been very generous in giving his attention, time and data to scholars consulting him. He almost always takes the trouble to reply; his replies can be grumpy on occasion but they do not leave the receiver empty-handed. Indeed they are often quite lengthy and invariably to the point. Bob's critical commentary is like his own compositions: notable for sharp analysis and lucid presentation.

Bob has been exemplary in reviewing the work of colleagues, whether for journals, publishers or funding agencies. He has, for instance, written more than 40 book reviews and notices. Most of us would begrudge taking the time to write so many reviews, but it seems that Bob regards reviewing not only as a service to the profession but actually enjoys it because it broadens his own store of intellectual capital. His reviews are invariably thorough and insightful. Often it is clear that he knows more about the subject of the book than the author does but he is quick to give credit where credit is due.

He has been no less exemplary in contributing to or editing works in honour of other senior scholars and in writing biographical memoirs. One thinks of the appreciations he has written of the work of Otto Dempwolff, Otto Christian Dahl, George Grace, Paul Benedict, Don Laycock and Jack Prentice, and the dedication of his book on Austronesian root theory to J.C. Anceaux.

11 A small token of appreciation

An appreciation of Bob's own scholarly contributions to the field is long overdue. We offer this festschrift as a small token of our esteem for this great scholar. The Blustery gale shows no signs of easing and his colleagues can look forward to many more years in which Bob blows new data, analyses, syntheses, critiques, and other objects across the world of Austronesian linguistics and beyond.

2 *Fêting Bob's career to-date*

BYRON W. BENDER

My first memory of Bob probably goes back to 1970, when I as department chair and he as graduate student had a conversation of some intensity. I don't recall much of what was said, but I know exactly where we were standing at the time, out near the elevators, above the stairs on fifth floor of Moore Hall. I think it involved an article of his that was to appear in the departmental working papers. What I do recall is his success in getting through to me his certainty and resolve as to his career—not his future career, but the one he was already well embarked upon. Only the future details remained to be filled in. I had known other graduate students who were well along the road to becoming professionals, but Bob was exceptional in this regard.

Segue to another spot in Moore Hall, an office with the number 573, occupied by Bob ever since shortly after his return to UH as Associate Professor in the mid-1980s, just across and up the hall from the one now shared by George Grace and me. Its shelves served me well when, as editor of *Oceanic Linguistics*, I had to check on details of bibliography or sort out language names for the annual index, and its occupant needed no card catalog to reach for the needed volume or file, or give me the information from memory. It isn't an especially desirable office. It has no windows, and is little more than a large closet. Its occupant certainly could have moved on to bigger and better as his career mounted, and especially when he served his stint as department chair, but I have a theory that one of the secrets of his success (and of his prodigious scholarly output) has been his decision not to disrupt and attempt to relocate the extensions to his memory stored there, each in its appointed place.

I don't believe that the other 'secrets' are really that difficult to discern, as it is clear that he is gifted with an unusual mind. His singleness of purpose, and the fact that his nose is kept to the grindstone 24/7 keep any of it from going to waste. One might even give Uncle Sam a bit of credit for sending him off to Indonesia on a language assignment at an early age. And I think that majoring in anthropology as an undergraduate has equipped him with a breadth of vision important for working at the prehistory of an area. Among the five major scholarly contributions he identifies in a CV that recently came to my attention are two that many readers of *Oceanic Linguistics* may not have been aware of and that stem

Alexander Adelaar and Andrew Pawley, eds
Austronesian historical linguistics and culture history: a festschrift for Robert Blust, 17–18.
Pacific Linguistics, 2009.
© This edition vested with Pacific Linguistics.

from this breadth: (1) 'the reconstruction of [the] prehistoric social organization [of the Austronesians]' as having 'had unilineal (probably matrilineal) descent groups, and a prescriptive rule of matrilateral cross cousin marriage,' and (2) '[how to explain] the origin of [a belief in] dragons', an exercise in reinventing a comparative method in cultural anthropology that is inspired and guided by the comparative method of linguistics.[1]

Finally, his wide circle of acquaintances among scholars in related disciplines as well as in linguistics has made him an ideal Review Editor for *Oceanic Linguistics*, a position he has held since 1999. I am happy to join this effort in fêting his contributions to date, and will close by saying, 'Keep up the good work, Bob.'

[1] Key publications from these two endeavors are 'Early Austronesian social organization: the evidence of language', *Current Anthropology* 21:205–226, 237–244 (with comments and reply), 1980; and 'The origin of dragons', *Anthropos* 95:519–536, 2000.

3 *Thoughts on learning that Bob Blust has reached festschrift age*

GEORGE W. GRACE

It's a shock to realise that Bob Blust has reached festschrift age. In fact, it seems he's reached an age that in some times and places has meant compulsory retirement. Fortunately, that's not the case here. He, of course, shows no sign of retiring or even slowing down; how much more a shock it would be if he did!

Anyway, I find myself confronted with a quite unwelcome question: How can this much time have gone by? It all seems so sudden. It seems only yesterday that he showed up in my office announcing his intention to begin graduate work in Austronesian linguistics. Has he (and, by inescapable implication, have I!) actually gone through so many years?

This rude awakening has set me looking back more closely at my recollections of the first years of my association with Bob. On the one hand, those years still seem very recent; but on the other hand, when I try to inspect the memories themselves more closely, they turn out to be badly blurred. That's to be expected, I realise, because I've rarely had reason to call on them. I've been struggling to locate and restore them, but readers should still be warned that memories can be treacherous.

Anyway, in my memory's eye I see Bob showing up one day in my office at the University of Hawai'i and announcing that he had recently finished a degree in anthropology at the University and wanted to go on to graduate work in Austronesian linguistics. At that time I'd never had the opportunity to direct a student in my field, and as I look back now, I don't seem to have thought ahead much about what responsibilities that would entail. There were really no developed programs in Austronesian linguistics anywhere and therefore nothing like a standard sequence of courses. But still that seemed to present no problem; I could just observe his progress through the linguistics program requirements and offer guidance when and where it was needed. However, in time I began to wonder whether that was working. Bob hardly ever seemed in need of guidance at all. He was remarkably self-sufficient, and this sometimes left me with an uneasy feeling I must be overlooking something I should be doing.

Alexander Adelaar and Andrew Pawley, eds
Austronesian historical linguistics and culture history: a festschrift for Robert Blust, 19–22.
Pacific Linguistics, 2009.
© This edition vested with Pacific Linguistics.

The fact was that Bob had already provided himself with a fairly good basic orientation in the field before he ever came into our program. I suspect this may have gotten started when he learned Indonesian, which he had mastered sufficiently that he was teaching it in the army. That may have been what gave him his first foot in the door, but whatever the original stimulus, he had apparently already done a certain amount of reading in the field on his own before he came to our department.

In any case Bob was very much what I believe they call a 'self starter'—he always seemed quite sure of what he wanted to do. He seemed well acquainted with the relevant literature, with what had been done and what kinds of research offered promise. And so he was able to pick his own research topics without any very specific guidance. I can't even remember how his dissertation research—which involved field work in Sarawak—was decided upon, planned, or financed. In particular, I can't recall having had much to do with it.

Bob's first (as far as I'm aware) professional article was published in 1969—several years before he finished his dissertation. It was a very substantial piece of work. It had something new to propose—the revision of a set of Proto Austronesian reconstructions on the basis of phonological developments in several Sarawak languages. Moreover, it was remarkably clearly written for a first article (Bob has always had the ability to produce clear, precise, and economical prose), and showed an impressive command of the relevant literature. It was in every way the work of a mature professional linguist. (In the interest of full disclosure, however, it should be mentioned that the conclusions of that paper have been revised in a paper he published in 2006).

Although Bob was the first student to begin studying Austronesian linguistics under me, he wasn't the first to finish. In fact, my main worry with him at this time was that he kept writing articles and not finishing his dissertation. Some of these articles made very substantial contributions to the field, but the research for them took him far beyond the topic of his dissertation. Of course, there is a tendency for each new research endeavor to raise more than one enticing new question that itself seems to demand to be pursued, and I worried that this process could only accelerate since it seemed to have no natural end. I came increasingly to be concerned that he would continue to be so captivated by all of the exciting new questions that waited to be investigated that he would never sit down and get the credentials he needed for a successful professional career.

But if his involvement in new research endeavors wasn't complication enough, he along the way met and fell in love with a Brazilian girl who was studying in Hawaiʻi at the time. Unfortunately she returned to Brazil, and Bob followed her there. I can't say how long he stayed there—it seemed then to be a long time. In fact, it was long enough that I wasn't at all confident that I'd ever see him again. However, eventually he did turn up back in Hawaiʻi, this time with a wife.

In spite of all the distractions, he did finally (in 1974) relieve my anxiety by completing his dissertation. At that juncture, he moved on to appointments first at The Australian National University and then at the University of Leiden in the Netherlands, before eventually returning (in 1984) to the University of Hawaiʻi, where he has remained ever since—this time as a member of the Linguistics Department faculty.

The shock of being so abruptly faced with the fact that Bob has reached festschrift age leaves me mentally scrambling to account for the time. What, I find myself asking, has Bob been doing to slip so many years past me? Of course, I have been aware all along that

he was doing things—research, publishing things, stuff like that—but could that really have used up so many years?

Still as I reflect further, I'm increasingly—however grudgingly—obliged to recognise that some of the things he has done must inevitably have consumed quite a bit of time. And further reflection reveals that there have, indeed, been a number of these time-consuming endeavors. To begin with (and it would surely be less than grateful for me to begin anywhere else) there was the festschrift he edited in my honor. The number of contributors, the general quality of the contributions, and the quality of the production exceeded anything I could have imagined hoping for. He must surely have devoted a lot of time to seeing this to its most successful conclusion.

But of course there have been a lot of other activities that have conspired to use up the years. The online Austronesian Comparative Dictionary must surely be described as the core project of all. It's a continuing project that already provides a very great increase in the number of available reconstructions and an enormous increase in daughter-language reflexes. In the same breath I should certainly mention his subgrouping of the family; he first reported its outlines in the 1970s, and has subsequently continued to refine it. It now has a near canonical status—its essential features accepted by all but a few holdouts.

Of course, this is not the place for an enumeration of Bob's contributions to knowledge, but no attempt—however superficial—to be fair about the time-consuming nature of Bob's activities should omit his field research, particularly in Borneo, the Admiralties, and Taiwan, and his analysis and reporting of the results in each case. As any field worker can readily attest, this would inevitably have demanded a lot of time.

Finally, I don't want to pass over certain ramifications from his core research that I believe deserve recognition that they don't always receive. These ramifications have extended in multiple directions: to culture-history, theoretical issues in phonology and linguistic change, and even a few topics that don't find a place in any ongoing scientific discourse or anywhere in the scientific division of labor. Anyway, these also will inescapably have accounted for a certain amount of time.

When this festschrift was first mentioned to me, my first impulse was to suspect that some trick—some telescoping of the years—had been played with the history of this period. However, as I contemplate what Bob has done, I find myself compelled to recognise that he has used the years to good effect, indeed that he probably needed all of them in their full measure.

Anyway, I must acknowledge that there's considerable satisfaction in being able to look at the contributions Bob has made in his professional career (or, as I prefer to think of it, this *first portion* of his career) and to refer to him as my student. But I must acknowledge that, although I'm credited with having directed his dissertation, that description accords somewhat imperfectly with my recollections. In my memory's scenario, the direction in which I was 'directing' led directly toward finishing the dissertation, while the direction he was actually taking followed a series of different tangents. In retrospect, I suppose the route he wound up following might be described as just a more circuitous one that nevertheless eventually led to the same destination. (But my memory is inclined to describe it as a more *distraction-filled* route that nevertheless in the long run *failed to prevent* his reaching that destination.)

Well whatever! Anyway, I can recall with satisfaction that Bob once referred to me as having taught him linguistics. This happened in the nineteen eighties when my son was taking an introductory linguistics course that Bob was teaching, and Bob remarked to me

that just as I had taught him linguistics, he was in turn teaching my son. That set me wondering about the role I had played in his linguistic education. Had I in some sense really taught him linguistics?

Again, whatever! If I'm given the credit for having taught him linguistics, I'll certainly not reject it—he has, after all, undeniably been taught well.

4 *The publications of Robert Blust*

Robert Blust's publications, numbering more than 200, are divided here into eleven categories according to their subject matter, plus an additional category consisting of translated collections.

1. Austronesian historical phonology
2. Austronesian subgrouping
3. Austronesian etymologies and semantic reconstruction
4. Austronesian historical morphology and syntax
5. Culture history
6. Principles of methodology and language change
7. Descriptive works (grammars, dictionaries, vocabularies etc.)
8. Surveys of Austronesian or its regions
9. Festchrifts and other edited books
10. Other articles and memoirs
11. Reviews and comments
12. Translated works

The content of many items cuts across two or more subject matter categories. In such cases, items are assigned to the category that is most prominently represented in the publication. In a few cases, where the claims of two categories are about equal, items are listed under both.

Abbreviations, and publication sites of journals

AA	*American Anthropologist* (Washington)
AF	*Anthropological Forum* (Nedlands, Western Australia)
AL	*Anthropological Linguistics* (Bloomington, Indiana)
AP	*Asian Perspectives* (Honolulu)
AU	*Afrika und Übersee* (Hamburg)
ANTHROPOS	*Anthropos* (St. Augustin, West Germany)
ARCHIPEL	*Archipel* (Paris)

Alexander Adelaar and Andrew Pawley, eds
Austronesian historical linguistics and culture history: a festschrift for Robert Blust, 23–35.
Pacific Linguistics, 2009.
© This edition vested with Pacific Linguistics.

BKI	*Bijdragen Koninklijk Instituut* (Leiden, Netherlands)
BRB	*Borneo Research Bulletin* (Williamsburg, Virginia)
BSOAS	*Bulletin of the School of Oriental and African Studies* (London)
CA	*Current Anthropology* (Chicago)
DB	*Dewan Bahasa* (Kuala Lumpur)
DIACHRONICA	*Diachronica* (Ottawa)
JCL	*Journal of Chinese Linguistics* (Berkeley)
JPS	*Journal of the Polynesian Society* (formerly Wellington, now Auckland)
JSEAS	*Journal of Southeast Asian Studi*es (Singapore)
JWP	*Journal of World Prehistory* (Columbus, Ohio)
Lg.	*Language* (Baltimore)
LINGUA	*Lingua* (Amsterdam)
LL	*Language and Linguistic*s (Academia Sinica, Taipei)
MT	*Mother Tongue* (Glouster, Massachusetts)
NUSA	*NUSA* (Jakarta: Atma Jaya University)
OCEANIA	*Oceania* (Sydney)
OL	*Oceanic Linguistics* (Honolulu)
PIL	*Papers in Linguistics* (Edmonton, Alberta)
PJL	*Philippine Journal of Linguistics* (Manila)
PL	Pacific Linguistics (RSPAS, The Australian National University, Canberra)
SMJ	*Sarawak Museum Journal* (Kuching)
TAPS	*Transactions, American Philosophical Society* (Philadelphia)
VKI	*Verhandelingen, Koninklijk Instituut* (Leiden)
WA	*World Archaeology* (London)
WPL	*Working Papers in Linguistics*, University of Hawaii (Honolulu)

1 Austronesian historical phonology

1969 Some new Proto-Austronesian trisyllables. *OL* 8:85104.

1970 A further note on marked order in consonant clusters. *WPL* 2.2:169–178.

1970 *i* and *u* in the Austronesian languages. *WPL* 2.6:113–145.

1971 A Tagalog consonant cluster conspiracy. *PJL* 2.2:85–91.

1973 The origins of Bintulu, ɓ, ɗ. *BSOAS* 36:603–620. Trans. by James T. Collins as ‘Asal-usul ɓ dan ɗ dalam Bahasa Bintulu’ in *DB* 28:173–196 (March 1984).

1974 The Proto-North Sarawak vowel deletion hypothesis. PhD dissertation. Honolulu, Department of Linguistics, University of Hawai‘i. 319pp.

1974 A double counter-universal in Kelabit. *PIL* 7.3–4:309–324.

1974 The Proto-Austronesian word for ‘two’: a second look. *OL* 13:123–161.

1976 A third palatal reflex in Polynesian languages. *JPS* 85:339-358.

1978 *The Proto-Oceanic palatals*. Polynesian Society Memoir 43. Wellington: The Polynesian Society. 183pp.

1979 Coronal-noncoronal consonant clusters: new evidence for markedness. *Lingua* 47:101–117.

1980 More on the origins of glottalic consonants. *Lingua* 52:125–156.

1981 The Soboyo reflexes of Proto-Austronesian *S. In Robert A. Blust, ed. *Historical linguistics in Indonesia*, 21–30. *NUSA*, vol. 10.

1981 Some remarks on labiovelar correspondences in Oceanic languages. In Jim Hollyman and Andrew Pawley, eds *Studies in Pacific languages and cultures in honour of Bruce Biggs*, 229–253. Auckland, Linguistic Society of New Zealand.

1982 An overlooked feature of Malay historical phonology. *BSOAS* 45:284–299. (Trans. by James T. Collins as 'Satu Ciri Fonologi Sejarah Bahasa Melayu yang Diabaikan' in *DB* 26.3:152–175, March 1982.

1983 A note on hypercorrection in Mongondow. *BKI* 139:459–464.

1984 On the history of the Rejang vowels and diphthongs. *BKI* 140:422–450.

1987 Rennell-Bellona *l* and the 'Hiti' substratum. In Donald C. Laycock and Werner Winter, eds *A world of language: papers presented to Professor S.A. Wurm on his 65th birthday*, 69–79. Canberra: Pacific Linguistics.

1990 Three recurrent changes in Oceanic languages. In J.H.C.S. Davidson, ed. *Pacific Island languages: essays in honour of G.B. Milner*, 7–28. London: School of Oriental and African Studies.

1995 Sibilant assimilation in Formosan languages and the Proto-Austronesian word for 'nine': a discourse on method. *OL* 34:443–453.

1996 Low vowel dissimilation in Ere. *OL* 35:96–112.

1996 Low vowel dissimilation in Oceanic languages: an addendum. *OL* 35:305–309.

1996 The Neogrammarian hypothesis and pandemic irregularity. In Mark Durie and M.D. Ross, eds *The comparative method reviewed: regularity and irregularity in language change*, 135–156. New York, Oxford University Press.

1997 Ablaut in Northwest Borneo. *Diachronica* 14:1–30.

1997 Nasals and nasalization in Borneo. *OL* 36:149–179.

1997 Rukai stress revisited. *OL* 36:398–403.

1998 Some problems in Thao phonology. In Shuanfan Huang, ed. *Selected papers from the Second International Symposium on Languages in Taiwan (ISOLIT)*, 1–20. Taipei: Crane.

1998 Seimat vowel nasality: a typological anomaly. *OL* 37.2:298-322.

1998 In defense of Dempwolff: Austronesian diphthongs once again. *OL* 37.2:354–362.

2000 Chamorro historical phonology. *OL* 39:83–122.

2000 Low vowel fronting in northern Sarawak. *OL* 39:285–319.

2001 Some remarks on stress, syncope and gemination in Mussau. *OL* 40:143–150.

2001 Language, dialect and riotous sound change: the case of Sa'ban. In Graham W. Thurgood, ed. *Papers from the Ninth Annual Meeting of the Southeast Asian Linguistics Society, 1999*, 249–359. Tempe: Arizona State University, Program for Southeast Asian Studies.

2002 Formalism or phoneyism?: the history of Kayan final glottal stop. In K. Alexander Adelaar and Robert A. Blust, eds *Between worlds: Linguistic papers in memory of David John Prentice*, 29–37. Canberra: Pacific Linguistics.

2003 The phonestheme N- in Austronesian languages. *OL* 42:187–212.

2003 Vowelless words in Selau. In John Lynch, ed. *Issues in Austronesian historical phonology*, 143–152. Canberra: Pacific Linguistics.

2004 Austronesian nasal substitution: a survey. *OL* 43:73–148.

2004 *t to *k*: an Austronesian sound change revisited. *OL* 43:365–410.

2006 Anomalous liquid: sibilant correspondences in western Austronesian. *OL* 45:210–216.

2006 The origin of the Kelabit voiced aspirates: a historical hypothesis revisited. *OL* 45:311–338.

2007 Òma Lóngh historical phonology. *OL* 44:1–53.

2007 Antiantigemination: canonical constraints in Austronesian languages. *Phonology* 24:1–36.

2007 The prenasalized trills of Manus. In Jeff Siegel, John Lynch and Diana Eades, eds *Language description, history and development: linguistic indulgence in memory of Terry Crowley*, 297–311. Amsterdam/Philadelphia: John Benjamins.

2008 A reanalysis of Wuvulu phonology. *OL* 47:275–293.

2 Austronesian subgrouping

1974 Eastern Austronesian: a note. *WPL* 6.4:101–107.

1977 The Proto-Austronesian pronouns and Austronesian subgrouping: a preliminary report. *WPL* 9.2:1–15.

1978 Eastern Malayo-Polynesian: a subgrouping argument. In S.A. Wurm and Lois Carrington, eds *Second International Conference on Austronesian Linguistics: proceedings.* (Fasc. 1), 181–234. Canberra: Pacific Linguistics.

1983/84 More on the position of the languages of eastern Indonesia. *OL* 22-23:1–28.

1984 Malaita-Micronesian: an Eastern Oceanic subgroup?. *JPS* 93:99–140.

1991 The Greater Central Philippines hypothesis. *OL* 30.2:73–129.

1993 *S metathesis and the Formosan/Malayo-Polynesian language boundary. In Øyvind Dahl, ed. *Language – a doorway between human cultures*: *tributes to Dr. Otto Chr. Dahl on his ninetieth birthday*, 178–183. Oslo: Novus.

1993 Central and Central-Eastern Malayo-Polynesian. *OL* 32(2):241–293.

1995 The position of the Formosan languages: method and theory in Austronesian comparative linguistics. In Paul Jen-kuei Li et al., eds *Austronesian studies relating to Taiwan*, 585–650. Symposium Series of the Institute of History and Philology, Academia Sinica, No.3. Taipei.

1996 Some remarks on the linguistic position of Thao. *OL* 35:272–294.

1996 The linguistic position of the Western Islands, Papua New Guinea. In John Lynch and Fa'afo Pat, eds *Oceanic studies: proceedings of the First International Conference on Oceanic Linguistics*, 1–46. Canberra: Pacific Linguistics.

1998 The position of the languages of Sabah. In Ma. Lourdes S. Bautista, ed. *Pagtanáw: essays on language in honor of Teodoro A. Llamzon*, 29–52. Manila: Linguistic Society of the Philippines.

1998 A note on higher-order subgroups in Oceanic. *OL* 37:182–188.

1999 Subgrouping, circularity and extinction: some issues in Austronesian comparative linguistics. In Elizabeth Zeitoun and Paul Jen-kuei Li, eds *Selected Papers from the Eighth International Conference on Austronesian Linguistics*, 31–94. Symposium Series of the Institute of Linguistics (Preparatory Office), Academia Sinica, No.1. Taipei: Academia Sinica.

2001 Malayo-Polynesian: new stones in the wall. *OL* 40:151–155.

2005 The linguistic macrohistory of the Philippines: some speculations. In Hsiu-chuan Liao and Carl R. Galvez Rubino, eds *Current issues in Philippine linguistics and anthropology parangal kay Lawrence A. Reid*, 31–68. Manila: The Linguistic Society of the Philippines and SIL Philippines.

2007 The linguistic position of Sama-Bajaw. S*tudies in Philippine languages and cultures* 15:73–114.

2008 Is there a Bima-Sumba subgroup? *OL* 47:45–113.

3 Austronesian etymologies and semantic reconstruction

1969 Some new Proto-Austronesian trisyllables. *OL* 8:85–104.

1970 Proto-Austronesian addenda. *OL* 9:104–162.

1972 Proto-Oceanic addenda with cognates in non-Oceanic Austronesian languages: a preliminary list. *WPL* 4.1:1–43.

1972 Additions to 'Proto-Austronesian addenda' and 'Proto-Oceanic addenda with cognates in non-Oceanic Austronesian languages'. *WPL* 4.8:1–17.

1972 Note on PAN *qa(R)(CtT)a 'outsiders, alien people'. *OL* 11:166–171.

1973 Additions to 'Proto-Austronesian addenda' and 'Proto-Oceanic addenda with cognates in non-Oceanic Austronesian languages – II'. *WPL* 5.3:33–61.

1980 Iban antonymy: a case from diachrony? In D.J. van Alkemade, et al., eds *Linguistic studies offered to Berthe Siertsema*, 35–47. Amsterdam, Rodopi.

1980 Austronesian etymologies. *OL* 19:1–181.

1982 The Proto-Austronesian word for 'female'. In Rainer Carle, et al., eds *Gava'*: *studies in Austronesian languages and cultures dedicated to Hans Kähler*, 17–30. Berlin, Reimer.

1983/84 Austronesian etymologies – II. *OL* 22–23:29–149. 040.

1986 Austronesian etymologies – III. *OL* 25:1–123. 047.

1987 Lexical reconstruction and semantic reconstruction: the case of Austronesian 'house' words. *Diachronica* 4:79–106.

1989 Austronesian etymologies – IV. *OL* 28:111–180.

1995 Austronesian comparative dictionary. Unfinished work available online. Currently prints out to 2045 single-spaced pages.

1996 Notes on the semantics of PAN *-an 'locative'. In Marian Klamer, ed. *Voice in Austronesian,* 1–11. NUSA: Linguistic Studies of Indonesian and Languages in Indonesia 39. Jakarta.

1997 Semantic change and the conceptualization of spatial relationships in Austronesian languages. In Gunther Senft, ed. *Referring to space*: *studies in Austronesian and Papuan languages,* 39–51. Oxford: Clarendon Press.

2002 The history of faunal terms in Austronesian languages. *OL* 41:89–139.

2005 Liver and lungs: a semantic dyad in Austronesian languages. *OL* 44:537–543.

2007 Proto-Oceanic *mana revisited. *OL* 46:504–523.

2008 (Simon J. Greenhill, Robert Blust and Russell D. Gray) The Austronesian basic vocabulary database: from bioinformantics to lexomics. *Evolutionary Bioinformatic*s, 271–283.

To appear Five patterns of semantic change in Austronesian languages. To appear in a festschrift for Andrew Pawley. Canberra: Pacific Linguistics.

4 Austronesian historical morphology and syntax

1974 Proto-Austronesian syntax: the first step. *OL* 13:1–15.

1976 Dempwolff's reduplicated monosyllables. *OL* 15:107–130.

1977 A rediscovered Austronesian comparative paradigm. *OL* 16:1–51.

1979 Proto-Western Malayo-Polynesian vocatives. *BKI* 135:205–251.

1988 Beyond the morpheme: Austronesian root theory and related matters. In Richard McGinn, ed. *Studies in Austronesian linguistics*, 3–90. Ohio University Center for International Studies, Center for Southeast Asian Studies, Monographs in International Studies, Southeast Asia series, No.76. Athens, Ohio.

1988 *Austronesian root theory: an essay on the limits of morphology.* Amsterdam, John Benjamins. 190pp.

1989 The adhesive locative in Austronesian languages. *OL* 28:197–203.

1998 Ca- reduplication and Proto-Austronesian grammar. *OL* 37:29–64.

1998 A note on the Thao patient focus perfective. *OL* 37.2:346–353.

1999 A note on covert structure: Ca- reduplication in Amis. *OL* 38:168–174.

2001 Historical morphology and the spirit world: the *qali/kali- prefixes in Austronesian languages. In Joel Bradshaw and Kenneth L. Rehg, eds *Issues in Austronesian morphology: a focusschrift for Byron W. Bender,* 15–73.

2001 Reduplicated color terms in Oceanic languages. In Andrew Pawley, Malcolm Ross and Darrell Tryon, eds *The boy from Bundaberg: studies in Melanesian linguistics in honour of Tom Dutton*, 23–49. Canberra: Pacific Linguistics.

2001 Thao triplication. *OL* 40:324–335.

2002 Notes on the history of 'focus' in Austronesian languages. In Fay Wouk and Malcolm Ross, eds *The history and typology of Western Austronesian voice systems*, 63–78. Canberra: Pacific Linguistics.

2003 A note on monosyllabic roots in Kavalan. *OL* 42:239–243.

2003 Three notes on early Austronesian morphology. *OL* 42:438–478.

2005 A note on the history of genitive marking in Austronesian languages. *OL* 44:215–222.

2006 Supertemplatic reduplication and beyond. In Henry Y. Chang, Lillian M. Huang and Dah-an Ho, eds *Streams converging into an ocean: festschrift in honor of Professor Paul Jen-kuei Li on his 70th birthday*, 439–460. Taipei: Institute of Linguistics, Academia Sinica.

5 Culture history

1976 Austronesian culture history: some linguistic inferences and their relations to the archaeological record. *WA* 8.1:19–43. (Reprinted with minor additions in NUSA 3:25–37 (1977), and in Pieter van de Velde, ed. *Prehistory of Indonesia: a reader*, 217–241 (Dordrecht, Holland).)

1980 Early Austronesian social organization: the evidence of language. *CA* 21:205–226, 237–244 (with comments and reply).

1980 Notes on Proto-Malayo-Polynesian phratry dualism. *BKI* 136:215–247.

1981 Dual divisions in Oceania: innovation or retention? *Oceania* 52.1:66–80.

1981 Linguistic evidence for some early Austronesian taboos. *AA* 83.2:285–319.

1982 The linguistic value of the Wallace Line. *BKI* 138:231–250.

1983 Pointing, rainbows and the archaeology of mind. MS, 40pp.

1984/85 The Austronesian homeland: a linguistic perspective. *AP* 26.1:45–67.

1986/87 Language and culture history: two case studies. *AP* 27:205–227.

1991 On the limits of the 'thunder complex' in Australasia. *Anthropos* 86.4/6:517–528.

1991 Sound change and migration distance. In Robert Blust, ed. *Currents in Pacific linguistics: papers on Austronesian languages and ethnolinguistics in* honour of George W. Grace, 27–42. Canberra: Pacific Linguistics.

1992 Tumbaga in Southeast Asia and South America. *Anthropos* 87:443–457.

1993 Austronesian sibling terms and culture history. *BKI* 149:22–76. (Also published in A.K. Pawley and M.D. Ross, eds *Austronesian terminologies: continuity and change*, 1994, 31–72. Canberra: Pacific Linguistics.)

1994 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 Asian Linguistics Society*, 25–83. Tempe: Program for Southeast Asian Studies, Arizona State University.

1995 The prehistory of the Austronesian-speaking peoples: a view from language. *JWP* 9.4:453–510.

1996 Austronesian culture history: the window of language. In Ward H. Goodenough, ed. *Prehistoric settlement of the Pacific*, 8–35. TAPS 86.5. Philadelphia, American Philosophical Society.

1996 Beyond the Austronesian homeland: the Austric hypothesis and its implications for archaeology. In Ward H. Goodenough, ed. *Prehistoric settlement of the Pacific*, 7–140. TAPS 86.5. Philadelphia, American Philosophical Society.

1999 Linguistics vs. archaeology: early Austronesian terms for metals. In Roger Blench and Matthew Spriggs, eds *Archaeology and Language III: Artefacts, languages and texts*, 7–143. London and New York: Routledge.

1999 The fox's wedding. *Anthropos* 94:487–499.

2000 The origin of dragons. *Anthropos* 95:519-536.

2000 Rat ears, tree ears, ghost ears and thunder ears in Austronesian languages. *BKI* 156:687–706.

2001 Historical morphology and the spirit world: the *qali/kali- prefixes in Austronesian languages. In Joel Bradshaw and Kenneth L. Rehg, eds *Issues in Austronesian morphology: a focusschrift for Byron W. Bender*, 15–73. Canberra: Pacific Linguistics.

2002 The history of faunal terms in Austronesian languages. *OL* 41:89–139.

2005 Borneo and iron: Dempwolff's *besi revisited. *Bulletin of the Indo-Pacific Prehistory Association* 25:31–40.

2006 Whence the Malays? In James T. Collins and Awang Sariyan, eds *Borneo and the homeland of the Malays: four essays*, 64–88. Kuala Lumpur: Dewan Bahasa.

2008 Remote Melanesia: one prehistory or two? An addendum to Donohue and Denham. *Oceanic Linguistics* 47.2:445–459**.**

To appear On datus, ancient and modern. In Loren Billings and Nelleke Goudswaard, eds *Piakandata ami Dr. Howard P. McKaughan*. Manila: Linguistic Society of the Philippines and SIL Philippines. MS. 18pp.

To appear The Austronesian diaspora. To appear in the Proceedings of the Austronesian Forum, sponsored by the Council of Indigenous Peoples, Taiwan, and held in Koror, Palau, April 9 2008. MS. 16pp.

6 Principles of methodology and language change

1989 A note on semantic cycles in historical change. *Diachronica* 6:297–300.

1990 Patterns of sound change in the Austronesian languages. In Philip Baldi, ed. *Linguistic change and reconstruction methodology*, 231–267. Berlin, Mouton de Gruyter.

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.

1994 Obstruent epenthesis and the unity of phonological features. *Lingua* 93:111–139.

1995 Sibilant assimilation in Formosan languages and the Proto-Austronesian word for 'nine': a discourse on method. *OL* 34:443–453.

1996 The Neogrammarian hypothesis and pandemic irregularity. In Mark Durie and M.D. Ross, eds *The comparative method reviewed: regularity and irregularity in language change*, 135–156. New York: Oxford University Press.

1999 Subgrouping, circularity and extinction: some issues in Austronesian comparative linguistics. In Elizabeth Zeitoun and Paul Jen-kuei Li, eds *Selected papers from the Eighth International Conference on Austronesian Linguistics*, 31–94. Symposium Series of the Institute of Linguistics (Preparatory Office), Academia Sinica, No.1. Taipei: Academia Sinica.

2000 Why lexicostatistics doesn't work. In Colin Renfrew, April McMahon and Larry Trask, eds *Time depth in historical linguistics*, 311–331. Cambridge, England: The McDonald Institute for Archaeological Research.

2001 The Comparative Method applied to Austronesian languages. In Sylvain Auroux, et al., eds *History of the Language Sciences*, 1374–1383. Berlin, Walter de Gruyter.

2005 Must sound change be linguistically motivated? *Diachronica* 22:219–269.

To appear Historical reconstruction. In Patrick Hogan, ed. Cambridge Encyclopaedia of Linguistic Science.

To appear On the explanation of similarity: perspectives on the evolution of languages, cultures and species. To appear in a volume edited by William McGregor. MS. 46pp.

To appear Complex sound changes. Submitted to *Diachronica*. MS. 16pp.

7 Descriptive works (grammars, dictionaries, vocabularies etc.)

1974 A Murik vocabulary, with a note on the linguistic position of Murik. *SMJ* 22.43(NS):153–189.

1977 Sketches of the morphology and phonology of Bornean languages 1: Uma Juman (Kayan). *Papers in Bornean and Western Austronesian languages*, No.2, 7–122. Canberra: Pacific Linguistics.

1984 The Tring dialect of Long Terawan. *SMJ* 33.54 (NS):101–135.

1984 A Mussau vocabulary, with phonological notes. *Papers in New Guinea Linguistics*, No.23, 159–208. Canberra: Pacific Linguistics.

1988 Sketches of the morphology and phonology of Bornean languages, 2: Mukah Melanau. In H. Steinhauer, ed. *Papers in Western Austronesian linguistics*, No.3, 151–216. Canberra: Pacific Linguistics.

1992 On speech strata in Tiruray. In Malcolm D. Ross, ed. *Papers in Austronesian linguistics*, No.2, 1–52. Canberra: Pacific Linguistics.

1992 The long consonants of Long Terawan. *BKI* 148:409–427.

1993 Kelabit-English vocabulary. *SMJ* 44.65 (NS):141–226.

1995 Notes on Berawan consonant gemination. *OL* 34:123–138.

1998 A Lou vocabulary with phonological notes. In Darrell Tryon, ed. *Papers in Austronesian linguistics*, No.5, 35–99. Canberra: Pacific Linguistics.

1999 Notes on Pazeh phonology and morphology. *OL* 38:321–365.

2003 *Thao dictionary*. Language and Linguistics Monograph Series A5. Institute of Linguistics (Preparatory Office), Academia Sinica. Taipei: Academia Sinica. 1106pp.

2003 *A short morphology, phonology and vocabulary of Kiput, Sarawak*. Canberra: Pacific Linguistics. 102pp.

8 Surveys of Austronesian or its regions

1984 Indonesia as a 'Field of Linguistic Study'. In P.E. de Josselin de Jong, ed. *Unity in diversity:Indonesia as a field of anthropological study*:21–37. *VKI* 103. Dordrecht, Holland, Foris Publications.

1987 The linguistic study of Indonesia. *Archipel* 34:27–47.

1988 Malay historical linguistics: a progress report. In Mohd. Thani Ahmad and Zaini Mohamed Zain, eds *Rekonstruksi dan cabang-cabang bahasa Melayu induk*: 1–33. Seri Monograf Sejarah Bahasa Melayu. Kuala Lumpur, Dewan Bahasa dan Pustaka.

1995 Austronesian languages. *Microsoft Encarta Encyclopedia*. ©1995 Microsoft Corporation. 12pp. (electronic version only).

1998 Austronesian languages. *Encyclopaedia Britannica*, 15th edition, 736–745.

1998 Various small pieces. *Encyclopaedia Britannica*, 15th edition.

2009 *The Austronesian languages*. Canberra: Pacific Linguistics. c 820pp.

9 Festchrifts and other edited books

1981 *Historical linguistics in Indonesia*. NUSA 10.

1991 *Currents in Pacific linguistics: papers on Austronesian languages and ethnolinguistics in honour of George W. Grace* (editor). Canberra: Pacific Linguistics.

2002 *Between worlds: linguistic papers in memory of David John Prentice.* K. Alexander Adelaar and Robert A. Blust, eds. Canberra: Pacific Linguistics.

10 Other articles and memoirs

1972 Report on linguistic fieldwork undertaken in Sarawak. *BRB* 4.1:12–14.

1988 Dempwolff's contributions to Austronesian linguistics. *AU* 71.2:90–96.

1995 An Austronesianist looks at Sino-Austronesian. In William S-Y Wang, ed. *The ancestry of the Chinese language*, 283–298. JCL Monograph Series, 8. Berkeley, California.

1987 'Introduction' in J.P.B. de Josselin de Jong, *Wetan Fieldnotes*, vii–ix. *VKI* 130. Dordrecht, Foris Publications.

1992 Don Laycock: a personal reminiscence. In Tom Dutton, Malcolm Ross and Darrell Tryon, eds *The language game: papers in memory of Donald C. Laycock*, 19–20. Canberra: Pacific Linguistics.

1993 Position statement. *MT* 19:9–11.

1998 In Memoriam: Paul King Benedict (1912-97). *OL* 37:1–11.

2002 Preface: Jack Prentice, 1942–95. In Adelaar and Blust, eds x–xi.

11 Reviews and comments

1976 Review of O.C. Dahl, *Proto-Austronesian*. *Lg.* 52:221–237.

1978 Review of Thomas A. Sebeok, ed. *Current Trends Linguistics*, vol. 8: *Linguistics in Oceania*. *Lg.* 54:467–480.

1980 Review of Roger Keesing, *Kwaio dictionary*. *JPS* 89:267–271.

1980 Reply to Goda, Keesing. *CA* 21:417–419.

1981 Review of Shigeru Tsuchida, *Reconstruction of Proto-Tsouic phonology*. *Lg.* 57:205–211.

1981 The reconstruction of Proto-Malayo-Javanic: an appreciation. *BKI* 137:456–469. (Trans. by James T. Collins as 'The reconstruction of Proto-Malayo-Javanic: sebuah penilaian' in *DB* 26.1:48–63 (January 1982).)

1981 Reply to Chowning. *CA* 22:298–300.

1982 Review of J.N. Sneddon, *Proto-Minahasan: phonology, morphology and wordlist*. *Lg*. 58:921–926.

1984 Comment on Marshall, 'Structural patterns of sibling classification in island Oceania.' *CA* 25:626–628.

1985 Review of Paul A. Geraghty, *The history of the Fijian languages*. *Anthropos* 79:684-686.

1985 Review of Cecil H. Brown, *Language and living things*. *Lg.* 61:891–897.

1986 Book notice for James T. Collins, *The historical relationships of the languages of central Maluku, Indonesia*. *Lg.* 62:471–472.

1988 Review of Mark Durie, *A grammar of Acehnese based on a dialect of North Aceh*. *BSOAS* 51.1:178–180.

1988 Book notice for Malcolm Warren Mintz and José del Rosario Britanico, *Bikol–English dictionary*. *Lg.* 64:214–215.

1989 Review of Hans Kähler, *Enggano-Deutsches Wörterbuch*. *AU* 72.1:155–158.

1989 Comment on Headland and Reid, 'Hunter-gatherers and their neighbors from prehistory to the present.' *CA* 30:53–55.

1990 Book notice for Claire Moyse-Faurie and Marie-Adèle Néchérö-Jorédié, *Dictionnaire Xârâcùù-Français*. *Lg.* 66:201–202.

1993 Review of M.D. Ross, *Proto Oceanic and the Austronesian languages of western Melanesia*. *Lg.* 69.1:186–192.

1995 Review of W. Wilfried Schuhmacher, Juan R. Francisco and F. Seto, *The emergence of Homo sapiens and his languages in tropical Asia*. *MT* 1:217–218.

1996 Review of L. Smits and C.L. Voorhoeve, eds 1992. *Irian Jaya source materials, No.5, Series B-No. 2: The J.C. Anceaux collection of wordlists of Irian Jaya languages, A: Austronesian languages (Part II).* Leiden/Jakarta, Koninklijk Instituut voor Taal-, Land- en Volkenkunde. *BKI* 152.1:154–155.

1996 Review of Winfred P. Lehmann, *Historical linguistics: an introduction*, 2nd ed. *OL* 35:324–328.

1996 Review of Terry Crowley, *An introduction to historical linguistics*, 2nd ed. *OL* 35:328–335.

1996 Review of Peter Bellwood, James J. Fox and Darrell Tryon, eds *The Austronesians: historical and comparative perspectives*. *AF* 7:468–471.

1997 Review of René van den Berg, *Muna–English dictionary*. *OL* 36:194–198.

1997 Review of James J. Fox and Clifford Sather, eds *Origins, ancestry and alliance: explorations in Austronesian ethnography*. *JSEAS* 28:413–415.

1997 Review of Darrell T. Tryon, ed. *Comparative Austronesian dictionary: an introduction to Austronesian studies*. *OL* 36:404–419.

1999 Review of Lynch, *Pacific languages: an introduction*. *OL* 38:178–182.

1999 Comments on LaVaughn H. Hayes, 'The Austric denti-alveolar sibilants.' *MT* 5:19–21.

2000 Review of Malcolm Ross, Andrew Pawley and Meredith Osmond, *The lexicon of Proto Oceanic: the culture and environment of ancestral Oceanic society*. *Diachronica* 13:182–191.

2000 Review of Graham Thurgood, *From ancient Cham to modern dialects: two thousand years of language contact and change*. *OL* 39:435–445.

2002 Review of Jeff Marck, *Topics in Polynesian language and culture history*. JPS 111:396–398.

2003 Review of Paul Jen-kuei Li and Shigeru Tsuchida, *Pazih dictionary*. *OL* 42:531–541.

2004 Review of Nikolaus P. Himmelmann, *Sourcebook on Tomini-Tolitoli languages*. *OL* 43:525–530.

2005 Review of John Lynch, Malcolm Ross and Terry Crowley, *The Oceanic languages*. *OL* 44:544–558.

2005 Review of Aone van Engelenhoven, *Leti: a language of southwest Maluku*. *AL* 47.3:365–368.

12 Translated collections

1985 *Telaah komparatif bahasa Nusantara Barat: kumpulan karya Robert A. Blust* (*Comparative studies of Western Indonesian languages: collected works of Robert A. Blust*). Trans. by Bambang Kaswanti Purwo and James T. Collins. Seri ILDEP. Jakarta, Djambatan. 247pp. (Indonesian translation of a collection of articles.)

1986 *Penyusunan salasilah bahasa di Malaysia: kumpulan karya Robert A. Blust* (*Compilation of the genealogy of languages in Malaysia: collected works of Robert A. Blust*). Compiled and translated by James T. Collins. Bangi, Selangor: National University of Malaysia. 138pp. (Malay translation of a collection of articles.)

Part 2

Sound change

5 *Structure-preserving sound change: a look at unstressed vowel syncope in Austronesian*

JULIETTE BLEVINS

1 21st century landscapes[1]

Over the course of the past several hundred years, advances in our understanding of sound change along with 20th century advances in phonetic science and phonological typology have given rise to a new landscape of sound patterns. A particular sound change in a particular language forms part of a population of similar sound changes with similar phonetic bases. Looking at this population, we can explore evidence for phonetic and non-phonetic factors in shaping the topography. Are particular sound changes rarer or more frequent than expected on purely phonetic grounds? And, if so, are they limited to languages or language families with particular structural features? While there has been much progress in explaining the phonetic bases of regular sound change, and simulating change in the laboratory (e.g. Ohala 1974, 1981, 1990; Guion 1998; Hardcastle and Hewitt 1999; Myers and Hanson 2005, 2007), a remaining challenge for any comprehensive theory of sound change is to identify more precisely language-specific structural pressures which may play a role, and to test hypotheses with real language data.

The Austronesian language family is fertile testing ground for the interaction of phonetic and structural conditions on sound change due to its size, its well established structural features, and the numerous recurrent sound changes documented. The great majority of regular sound changes within the family have clear phonetic motivations in misperception, coarticulation, aerodynamics, and/or articulatory weakening and strengthening and are

[1] It is an honor and pleasure to dedicate this paper to Bob Blust, a dear friend and colleague, whose work continues to challenge and inspire me. The leading idea in this paper was first presented in 2003, at a meeting of the Austronesian Circle in Honolulu, where Bob was present. Seeking expertise on a range of issues, I suggested we co-author a paper on Austronesian syncope, and a manuscript, Blevins and Blust (2003), came into existence. However, in the end, the work split seamlessly into two distinct studies: Bob's scholarly and insightful paper on vowel loss between identical consonants (Blust 2007), and this preliminary account of potential asymmetries in the distribution of medial unstressed vowel syncope across Austronesian languages.

Alexander Adelaar and Andrew Pawley, eds
Austronesian historical linguistics and culture history: a festschrift for Robert Blust, 39–55.
Pacific Linguistics, 2009.
© This edition vested with Pacific Linguistics.

common outside the family (Blust to appear, *pace* Blust 2005; Blevins 2004a; Blevins 2006). Others, like final consonant-loss, appear to have primary structural motivations, and are rare outside of Austronesian (Blevins 2004b). A third class of recurrent sound changes in Austronesian have clear phonetic and structural conditioning. The most striking structural conditioning factor discovered to date is the disyllabic output constraint proposed in Blust (2007). As Blust shows, three independent and recurrent sound changes (initial vowel epenthesis, laryngeal loss, and unstressed vowel loss between identical consonants) occur precisely when their output is disyllabic. This distribution is attributed to a structural feature of the lexicon: over 90% of all reconstructed lexical bases in Proto Austronesian and other early Austronesian protolanguages were disyllables. In the course of language acquisition, high frequency disyllables act as 'attractors' drawing phonetically ambiguous tokens their way. When this current is multiplied over speakers and generations, sound changes appear to conspire to disyllabic outputs.

The existence of language-specific structural pressures has been hypothesized for some time. Within the historical literature, recurrent, or parallel changes in related languages which can not be attributed to chance, universals or diffusion, have been categorized as instances of 'drift' (Sapir 1921; Blust 1978; Blust 1990; Andersen 1990). Under drift, languages which are no longer in contact are believed to move in similar directions due to the continued, independent operation of inherited structural pressures. Though many sound changes can be viewed as having a basis in 'the tendency to increased ease of articulation' or 'the cumulative result of faulty perception', the operation of such forces cannot explain why 'one language encourages a phonetic drift that another does everything to fight' (Sapir 1921:196). Language-specific priming effects have also figured prominently in phonological analyses (e.g. De Chene and Anderson 1979; Kiparsky 1995). Though this type of priming has been looked at from many different perspectives over the 20th century, experimental paradigms of the 21st century provide new empirical support for the existence of language-specific structural pressures.

In a range of experiments, it has been shown that native language biases in speech perception emerge at an early age (e.g. Werker and Tees 2002; Best and McRoberts 2003). Additional studies demonstrate that infants as young as 6 months show sensitivity to distributional information in sound patterns based on previous exposure (e.g. Maye et al. 2002; Saffran and Thiessen 2003). At the same time, agent-based simulations using a production-perception feedback loop demonstrate 'attractor' effects, and are able to model emergent regularities over the lexicon (Wedel 2006, 2007). Blust's discovery of disyllabic attractors in Austronesian sound change noted above, can now be firmly grounded in models of acquisition where infants exposed to a biased lexicon will showed biased perceptions from an early age, and where biases in part of the lexicon will tend to be regularized over time.

2 Structure preserving sound change

Within this new landscape, a general structural principle has been suggested by Blevins (2004a:154) to account for the strong tendency of specific types of sound changes to be structure-preserving. The principle, referred to as 'Structural Analogy', is stated in (1).

(1) Structural Analogy (Blevins 2004a:154)

> In the course of language acquisition, the existence of a (non-ambiguous) phonological contrast between A and B will result in more instances of sound change involving shifts of *ambiguous elements* to A or B than if no contrast between A and B existed.

This principle is relatively simple. If two (or more) phonological segments or prosodic categories are learned quickly due to their unambiguous categorial status, then the existence of those categories in the evolving grammar will attract incoming ambiguous tokens to the pre-established categories. The overall consequence of (1) on historical grammars is 'structure-preserving sound change': the output of a sound change is a category or structure which existed prior to the sound change itself.

Strong typological evidence for structure-preserving sound change is found in surveys of compensatory lengthening (de Chene and Anderson 1979, Kavitskaya 2002) and metathesis (Blevins and Garrett 1998, 2004; Hume 2004). Kavitskaya (2002) argues that compensatory lengthening sound changes result from phonologization of pre-existing differences in phonetic vowel duration. Phonetic factors leading to longer vowel durations include longer V–C transitions, longer vowels before particular consonants, and open-syllable lengthening. In her survey of 80 languages with historical compensatory lengthening sound changes, 72 or 90% occur in languages with pre-existing long/short vowel contrasts, while only 8 or 10% are found in languages without a pre-existing vowel length contrast. The principle in (1) accounts for this tendency by mapping vowels of ambiguous length more often to a long vowel category when this category is independently established by non-ambiguous short versus long vowel contrasts inherited from a protolanguage. In metathesis, the inversion of segment order has a strong tendency to result in phonotactics which pre-exist in the language at large. For example, 23 of 24 cases of perceptual metathesis, where a long-domain feature is reinterpreted in a non-historical position, involve cases where the output of metathesis yields a pre-existing phonotactic (Blevins and Garrett 1998; 2004). A similar generalization appears to hold of regular CC metathesis; all 11 cases of regular sibilant-stop metathesis surveyed in Blevins and Garrett (2004) and Blevins (2009), and all four regular cases of CC metathesis in Hume (2004) result in consonant clusters which are characteristic of the pre-sound change stage of the language.

In this study, I would like to explore predictions of Structural Analogy (1) where unstressed vowel syncope is involved. By unstressed vowel syncope, I mean the historical loss of an unstressed vowel in the environment shown in (2), where superscripted 'V' is an unstressed vowel, and periods mark syllable boundaries. (Unstressed vowel syncope can also result in V.CCV and, rarely, VCC.V syllabifications, though these will not be discussed in this paper.)

(2) Unstressed vowel syncope as sound change

$VC^{V}CV > VC.CV$

The phonetic basis of unstressed vowel syncope is relatively uncontroversial. A historically short unstressed vowel when hypoarticulated may be reduced to a point where its vocalic status is ambiguous. The language learner is faced with an analytical problem: is the fleeting vowel a true phonological vowel, or a phonetic transition between neighbouring consonants? If a learner decides there is no medial vowel, a syncopating sound change has occurred at the level of the individual. Intrinsic properties of the medial vowel in $VC^{V}CV$ which facilitate syncope are shorter duration and predictable quality. The

shorter the vowel is, the shorter it will be under gestural reduction and compression which may accompany hypoarticulation. Short unstressed vowels are the canonical targets of syncope because, generally, unstressed vowels are shorter than stressed vowels, and short unstressed vowels are shorter than long unstressed vowels. Cross-linguistic studies of intrinsic vowel duration show that high vowels are typically shorter than low vowels (Lehiste 1970), so it is not surprising that in some languages, a high vowel is the sole target of syncope. In other languages a central vowel like [ə] may be shorter than others (Gordon 1999), and a unique target of syncope. Another contributing factor to syncopating sound change is the extent to which vowel quality is predictable from local context. When vowel quality is predictable, it can be attributed by the language learner to the phonetics, and need not be lexically represented. This can occur when the unstressed short vowel has a consistent quality, or when the vowel varies according to context, but variation is predictable. In this second case, variable vowel quality may be a consequence of phonetic coarticulation, phonological vowel-copy, or even morphological patterns, like reduplication (Blevins 2005).

The question addressed here is whether, in addition to phonetic factors, language-specific structural properties can inhibit or facilitate syncopating sound change. Structural Analogy (1) predicts that they should. In particular, languages with an unambiguous contrast between closed and open syllables will have a stronger tendency to undergo historical syncope than languages with only open syllables. This prediction appears to hold for many well described cases of historical syncope in the literature (e.g. Old Irish (Thurneysen 1980); Yupik (Jacobson 1984); Chamorro (Blust 2000)). However, the prediction has yet to be evaluated in any systematic way. In the remainder of this paper, I explore the predictions of (1) for unstressed vowel syncope in the Austronesian language family. As will be seen, Structural Analogy is supported by a clear empirical generalization: the most significant structural feature in predicting the occurrence of syncope as sound change in Austronesian languages is the pre-existence of closed syllables.

Before turning to the Austronesian facts, it is important to highlight ways that syncope may differ from two other types of unstressed vowel loss that are common within Austronesian, and especially within Oceanic: final voiceless vowel loss; and *mu > m sound changes (Blevins 2004a:162–164). In these cases, a working hypothesis is that vowels are lost because the language learner typically fails to perceive them. An input string with final …CV̥ presents no ambiguity, since it is (mis)perceived as C-final (cf. Myers and Hanson 2007). No canonical ambiguity is involved, and Structural Analogy is not implicated. If this working hypothesis turns out to be correct, it highlights the importance of distinguishing sound changes based primarily in misperception from others due to ambiguous cue-localization, or ambiguity due to articulatory variability. Certain types of sound change may be greatly influenced by ambient sound patterns, and show strong structure-preservation effects, while others may not.

3 Unstressed vowel syncope in Austronesian

The Austronesian language family is one of the biggest in the world, including over 1000 distinct languages, and covering a vast geographical region from the Indian Ocean to the eastern Pacific. Despite the size of this family, and the diversity among subgroups, there is perhaps no language family, apart from Indo-European, whose comparative phonology is better studied, and an extensive PAn vocabulary has been reconstructed

(Blust 1995, to appear). All of these features make the Austronesian language fertile testing ground for the interaction of phonetic and structural conditions on sound change.

There are several general features of Proto Austronesian (PAn) sound patterns which are relevant to the discussion which follows. First, the vowel system of Proto Austronesian contained three short vowels /i a u/ and one extra-short vowel /ə/ (often written as *e* in PAn reconstructions). Since the extra-short vowel was shorter than the other vowels, it is not surprising that reflexes of Proto Austronesian *ə are targets of syncope in many daughter languages.

Another property of Proto Austronesian relevant to syncope are the syllable types reconstructed. (Here and throughout, reconstructions are from Blust 1995 unless noted otherwise). PAn syllables can be open (*qu.lu 'head', *su.su 'breast') or closed (*nəm.nəm 'think', *səpsəp 'suck'). Closed syllables occur word-finally and in reduplicated monosyllables, and syllable codas include oral stops, fricatives, nasals liquids and glides (*likud 'back', *bukəS 'hair', *bulan 'moon', *balbal 'hit, beat', *Sapuy 'fire'). Stress, of course, is also relevant. However, few protoforms are reconstructed with stress, though most believe it was lexically distinctive.

Finally, it is worth noting that, though the majority of lexemes in Proto Austronesian were disyllabic (Blust 2007), affixed words were often longer, and trisyllables did occur, e.g. *bituqən 'star', *qabaRa 'shoulder', *Cinaqi 'guts', *CaliNa 'ear'. Given the unstressed, reduced nature of PAn *ə, the Proto Austronesian words in table 1 are more likely to undergo unstressed vowel syncope (2) in daughter languages than words with full vowels in the same positions. Indeed, there is evidence for syncope of these unstressed medial vowels in many Austronesian languages, as discussed below.

Table 1: Proto Austronesian unstressed /ə/ in potential syncope environments

a.	*baqəRuh	'new'	f.	*qaNəliC	'smell of burnt rice, burning hair, etc.'
b.	*qiCəluR	'egg'	g.	*binəSiq	'seed rice'
c.	*qaləjaw	'day'	h.	*qatiməla	'flea'
d.	*timəRaq	'tin'	i.	*paləCuk	'shoot'
e.	*tuqəlaN	'bone'	j.	*bagəqaŋ	'molar, tooth'

3.2 Formosan languages

The Formosan languages form at least nine primary subgroups within Austronesian (Blust 1999). While data from some of these subgroups is scanty, there is evidence from at least seven subgroups for unstressed vowel syncope. Some representative developments are shown in Table 2, with numbered subgroups corresponding to those in Blust (1999:45). In this and subsequent tables of this type, the final column illustrates maintenance of PAn final consonants. Wherever possible, examples have been used in which the derived coda under syncope is an exact or near segmental match to a pre-existing coda consonant. However, within the proposed model no such exact matching is assumed; in principle, any closed syllable can serve as a potential matching target for an ambiguous …$VC^{V}CV$… string.

Table 2: Syncope and closed syllables in some Formosan languages

Subgroup	Language	Syncope	PAn final *C maintained
2	Amis	*baqəRuh ‘new’ > *faʔloh*	*baRaq ‘lung’ > *falaʔ*
3	Puyuma	*paləCuk ‘shoot’ > *palTuk*	*biRbiR ‘lips’> *birbir*
4	Kavalan	*baŋəliS ‘tusk’ > *baŋris*	*biraŋ ‘to count’ > *bilaŋ*
6	Tsou	*qatiməla ‘flea’ > *timro*	*ənəm ‘six’ > *nomə*
7	Bunun	*baqəRuh ‘new’ > *baqlu*	*baRaq ‘lung’ > *bahaq*
8	Thao	*baqəRuh ‘new’ > *faqlu*	*aN-aNak ‘child’ > *al-alak*
9	Saisiyat	*binəSiq ‘seed rice’ > *binSiʔ*	*SimaR ‘grease’ > *ʃimaɾ*

In Tsouic languages (Tsou, Kanakanavu, and Saaroa), where reflexes of PAn final consonants are followed by historically excrescent vowels, Proto Tsouic lacks these final vowels, but does show evidence of medial unstressed vowel syncope, as in PAn *LiSəpis ‘thin’, Proto Tsouic *Lipis, Tsou *hipsi*. This syncope rule is distinct from later unstressed vowel loss, where pre- and post-tonic vowels are lost (Tsuchida 1976:210–211). Note that there is no possibility that these instances of syncope are inherited, since PAn is reconstructed with the syncopating vowel intact. While these facts might seem unremarkable, they are consistent with the predictions of (1): syncope is associated with the pre-existence of unambiguous closed syllables in all Formosan subgroups where it is in evidence.

3.3 Western-Malayo-Polynesian languages

There is great debate as to how the hundreds of Western-Malayo-Polynesian languages subgroup. Here I focus on points of agreement, and demonstrate that there is evidence for parallel independent sound changes of unstressed vowel syncope where PAn final consonants are maintained. The general picture is outlined in Table 3: of the fourteen subgroups of Western Malayo-Polynesian, all of which inherited WMP final consonants, there is evidence for syncope in all but five groups: Batak, Sangiric, South Sulawesi, West Flores, and in Palauan. In at least two of these groups, South Sulawesi and Sangiric, inherited final consonants are highly limited (Sneddon 1984, 1993) further supporting an association between syncope and unambiguous closed syllables. Within all other groups, historical syncope is attested in either all languages, or in some languages, but not others, suggesting independent parallel developments. Note that many Philippine languages like Tagalog show syncope of suffixed forms of **qatəp* ‘roof; thatch’ (*atíp* ‘roof’, *aptán* ‘thatch a roof), while others, like Bikol, do not (*atóp* ‘roof’, *atopán* ‘to roof, to thatch’). This shows either that syncope followed the breakup of Proto Philippines, and many of its descendant protolanguages, or that vowels have been restored by analogy with unsuffixed forms.

Table 3: Historical syncope in Western Malayo-Polynesian languages

Subgroup/language	Syncope?	PAn final *C maintained
1. Central Cordilleran	yes	yes
2. Central Philippines	yes	yes
3. Manobo	yes/no**	yes
4. Samalan	yes/no**	yes
5. Chamorro	yes	yes
6. Palauan	no	yes
7. North Sarawak	yes	yes
8. Malayo-Chamic	yes	yes
9. Barito	some languages	yes
10. Batak	no	yes
11. Sangiric	no	rarely
12. South Sulawesi	no	rarely
13. West Flores	no	yes
14. Watubela	yes	yes

**Contradictory evidence: some reflexes of *…VCəCV… show syncope while others do not

For particular cognate sets, it is difficult to find Western Malayo-Polynesian languages which maintain reflexes of initial and final CV in PAn *CVCəCV but do not show regular syncope of the medial vowel. For example, for **baqəRuh* 'new', of the 47 WMP languages listed in Blust (1995) and the 43 minor Philippine languages surveyed in Reid (1971), only two show unambiguous trisyllabic cognates without syncope: Sundanese *bahayu* 'recently', and Samal *bahaqu* 'new'. Blust (1995) also shows Malay variants *baharu, bahru, baru* 'new; fresh; now at last', with two variants reflecting syncope. Table 4 gives a representative subset of WMP languages where syncope and maintenance of PAn final consonants are both in evidence.

Table 4: Syncope and closed syllables in some WMP languages

Language	Syncope	PAn final *C maintained
Ilokano	*qaləjaw 'day' > *aldaw*	*biRbiR 'lips' > *birbir* 'rim'
Isneg	*qaləjaw 'day' > *alxaw*	*bituqən 'star' > *bittuwan*
Bikol	*qaləjaw 'day' > *aldaw*	*sakal 'muzzle, yoke' > *sakal*
Chamorro	*qaləjaw 'day' > *atdaw* 'sun'	*qipil 'k.o. tree, *Intsia bijuga*' > *ifet*

3.4 Central Malayo-Polynesian languages

Blust (1993) demonstrates that PAn final consonants are retained in Proto Central Eastern Malayo-Polynesian and in Proto Central Malayo-Polynesian (PCMP). Reconstructions in Table 5 from Blust (1993) show that word-medial consonant clusters were simplified to single consonants (5i), while word-final consonants were typically maintained (5i-ii). (Two exceptions to this are regular loss of PAn final **h* (**talih* > *tali*) and irregular loss of final **k* in **qabu* 'dust' < **qabuk*.)

Table 5: PAn coda consonants lost medially, retained finally in CEMP (Blust 1993)

	PMP	PCEMP	PCMP	Gloss
i.	*baqbaq	*babaq	*babaq	‘mouth’
	*səpsəp	*səsəp		‘suck’
	*gurgur	*gugur	*gugur	‘thunder’
	*təktək	*tətək	*tətək	‘cut wood’
ii.	*qatəp	*qatəp	*qatəp	‘thatch’
	*kulit	*kulit	*kulit	‘skin’
	*taŋis	*taŋis	*taŋis	‘cry’
	*inum	*inum	* inum	‘drink’

Given the suggested correlation between pre-existing closed syllables and unstressed vowel syncope, syncope is expected in CMP languages prior to subsequent developments involving final consonant loss and/or final vowel accretion. However, this hypothesis is very difficult to evaluate in many CMP languages due to independent sound changes. For example, the loss of prepenultimate *{h,q}Vs (Blust 1993:263–264) eliminates certain trisyllables (e.g. continuations of b, c, f, h in Table 1) which might otherwise be subject to syncope. For example, reflexes of PAn **qatəluR* ‘egg’ are disyllabic: Bima *dolu*; Manggarai, Ngadha, Lio, Sika *telo*; Lamaholot *telu-k*; Kedan *tolor*; Kambera *tilu*; Savu *delu*; Roti *tolo*; Tetun *tollu-n*; Kemak *telo*; Mambai *telo-n*; Kisar *keru-nne*; Leti *ternu* (< **terunu*); Selaru *tesu*; Ujir *tuli*; Ngaibor *tulir*; Kei *tolur*, Elat *tulur*; Geser *tolu*; Bonfia *toli-n*; Nuaulu *tou-ne*; Paulohi *terur*; Buru *telu-n*; Soboyo *tolu*. As disyllables, these words are no longer targets of the sound change in (2), whose input is trisyllabic. However, in at least one CMP language Watubela, a change of **q* > *k* appears to have saved the initial syllable, with subsequent syncope giving rise to *katlu* ‘egg’.

Another irregular sound change applies to **bəqəRu* ‘new’ (Blust 1993:266). Perhaps due to the sequence of schwas, there is irregular coalescence of the first two syllables, bleeding syncope: PAn *baqəRuh ‘new’; PCEMP **bəqəRu*; Bima *ʔbou*; Manggarai *weru*; Sika *weru-ŋ*; Roti *beu-k*; Atoni *feʔu*; Tetun *foo-n*; Mambai *heu*; Kisar *woru-woru*. It appears then that in the majority of CMP languages, *…VCəCV strings have been simplified to disyllabic strings by independent sound changes, eliminating many potential targets of syncope. Where trisyllabic strings like PAn **qateluR* ‘egg’ have not been reduced to disyllables by other means, syncope is in evidence, as in Watubela *katlu*.

3.5 South-Halmahera-West New Guinea languages

The South Halmahera-West New Guinea group includes about 40–50 languages. Blust (1993) provides comparative data for Buli and Numfor which make it clear that Proto SHWNG also maintained inherited final consonants: compare PCEMP *laman ‘deep’, Buli *m-laman*, Numfor *ramen*; PCEMP *malip ‘laugh’, Buli *a-mlif*, Numfor *mbrif*, etc. Blust (1978) reports on one of the distinctive sound changes evidenced in many SHWNG languages: post-nasal syncope giving rise to word-initial NC… sequences. For example, in the Waikyon dialect of Taba (Bowden 2001:50), we find *mto* ‘eye’ < PSH **mta* < PAn **maCa*. The same syncope is in evidence in Buli *-mlif* and Numfor *mbrif* ‘laugh’ from PCEMP *malip ‘laugh’. Due to this sound change as well as others, there are no cases where an inherited *…VCəCV from PAn, PMP, or PCEMP is in evidence. In some

instances, as in CMP languages, *…VCəCV is reduced by other means to a disyllable. In other cases, penultimate stress (with vowel quality change) blocks vowel loss.

However, the synchronic phonologies of SHWNG languages for which detailed descriptions are available provide evidence for historical syncope. One of these languages is Taba (Bowden 2001). Taba, like Buli and Numfor, shows retention of inherited final consonants: *mloŋan* 'depth' < *ma* + PCEMP **laman*, *-amlih* 'laugh' < PCEMP **malip*, *mlút* 'be soft' < *ma* + PCEMP **lut*), etc. In Taba, stress is regularly penultimate (Bowden 2001:51–53). Synchronic alternations in (3) appear to reflect historical syncope of an unstressed post-tonic vowel (Bowden 2001:64–70). The syncope in (3) only occurs with applicatives *-o* and *-Vk*. The fact that regular penultimate stress is not found with applicative *-o* and *-Vk* suggests that these affixes predate the regular penultimate stress rule, and hence reflect an older stage of the language when *VC^{V}CV > *VCCV.

(3) Synchronic syncope in Taba, a SHWNG language (Bowden 2001:64–70)

Verb	Applicative verb
lékat	*lékto* 'be bad'
báliŋ	*bálŋik* 'wrap up'
búlaj	*búljak* 'twist'
téden	*tédnek* 'stack'

Another SHWNG language with limited synchronic syncope alternations is Sawai (Whisler and Whisler 1994). With the location focus suffix which may be cognate with Taba *-o*, we find alternations like: *n-ɔbɛn, wɔbn-o* 'new', *tapɛn, n-tapn-o* 'shoot'.

3.6 Oceanic subgroups with inherited final consonants

Within the Oceanic subgroup inherited final consonants are maintained only in Admiralties, Western Oceanic (North New Guinea Cluster, Papuan Tip Cluster, Meso Melanesian Cluster), New Caledonia, and Southern Vanuatu. Other subgroups (South-East Solomons, Micronesian, North Central Vanuatu and Central Pacific) are characterized by loss of inherited final consonants. Where this loss has resulted in languages with only open syllables, e.g. in Central Pacific, syncope is unexpected under the present account. However, in languages which maintain final consonants, historical syncope is expected.

Within the Oceanic group, however, additional sound changes, including shifts of stress to the penultimate syllable, have eliminated many contexts where syncope could apply. This appears to be the case, for example in the Huon Gulf languages of the North New Guinea Cluster within Western Oceanic. Huon Gulf languages maintain inherited final consonants (Table 6), though the same consonants were lost in the sister Schouten Chain group (Ross 1988:124).

Table 6: Retention of POC final consonants in Proto Huon Gulf

POC	PHG	Wampur	Hote	Mapos Buang	Gloss
*banic	*banic	*banit*	*banik*	*banis*	'wing'
*kupit	*kupic	*ubit*	*kupik*		'bark'
*manuk	*manuk		*menak*	*mank*	'bird'

However, regular penultimate stress blocks potential syncope in reflexes of PHG **tavuRi* 'Triton shell', PHG **ʁalivan* 'centipede', PHG **maliboŋ* 'flying fox', and PHG *vituʁun* 'star'. Instead, some Huon Gulf languages show reduction (or deletion) of the post-tonic or pre-tonic vowel as in, e.g. Mapos Buang, Kaiwa *mank* < **manuk*, Mapos Buang *btuʁŋ* < PHG **vituʁun* (Ross 1988:71), while others maintain trisyllabic CVCVCV sequences with penultimate stress, e.g. Kove *pitoho*, Tuam *pitola* < Proto Bariai **pitoRo* < POC **pitolo* 'hungry' (Ross 1988:176).

Within Remote Oceanic, where subgrouping is better established, and where historical phonology is better documented, we can see clear cases where languages with inherited final consonants show syncope. This is true of Southern Vanuatu languages, as detailed by Lynch (2001). In Table 7 Southern Vanuatu languages show retention of word-final Proto Oceanic consonants.

Table 7: Retention of Proto Oceanic final consonants in Southern Vanuatu languages (Lynch 2001:102)

POc	Sye	Lenakel	Anejom̃	Gloss
*kurat	*no/ɣrat*	*na/uias*	*no/uras*	*Morinda citrifolia*
*tanum	*tenəm*	*renəm*	*a/tenom*	'bury'
*saqat	*sat*	*taat*	*has*	'bad'
*manuk	*menuɣ*	*menuk*	*n/man*	'bird'
*rarap	*n/arap*	*n/aiəv*	*n/ara*	Erythrina sp.
*quloc	*n/ilah*	*n/ilah (S)*	*n/ija*	'maggot'
*pekas	*e/vɣah*	*a/vhe*		'defecate'
*tuqur	*e/tur*	*a/lel (S)*		'stand'
*(ŋ)awaŋ	*ovaŋ*	*owaŋ*		'be open'

In Table 8 Southern Vanuatu languages show evidence of unstressed vowel syncope. In these examples medial pre-tonic vowels were regularly deleted. Stress in these languages is on a final CVC syllable, otherwise on the penultimate syllable.

Table 8: Southern Vanuatu syncopating sound change in verbs (Lynch 2001:115–117)

Proto Oceanic	Sye	Lenakel	Anejom̃	Gloss
*a-bulut-i	*amplehi*			'stick to'
*a-paŋan-i	*avŋoni*			'feed'
*a-panako		*əvnak*		'steal'
*a-punuq-i			*opra*	'long'
*a-bulut	*amplet*	*apwiit*	*apwol*	'sticky, stick to'
*a-likos	*elki*	*əlkəs*	*ajɣei*	'hang, tie up'
*a-mataq	*emte*	*amra*	*mat*	'raw, new'
*a-labwat		*ipwər*	*alpwas*	'large'

In Southern Vanuatu, and the majority of Austronesian languages which have inherited closed syllables from Proto Austronesian, a sound change of unstressed vowel syncope has occurred. The same is not true of Austronesian languages which have evolved to have only open syllables, as I now demonstrate.

4 The absence of syncope in languages without closed syllables

4.1 Central Pacific

The loss of final consonants in the history of Oceanic has given rise to subgroups which, historically, lacked closed syllables altogether. This is the case for Central Pacific, which includes Rotuman, the Fijian languages, and the Polynesian languages. An interesting aspect of the sound patterns of these languages is not only that none appear to have synchronic syncope alternations, but also that in the several thousand years of independent development, none has innovated a regular (medial) syncopating sound change. (For a discussion of vowel loss in word-initial [CVC… sequences, see Blust (2007).)

At the same time, the absence of historical syncope cannot be attributed to the failure of unstressed vowels to reduce. Phonetic studies of Fijian fast speech show significant reduction of unstressed vowels (Tamata 1994; Erickson 1996), and in the history of Rotuman, final unstressed vowels have arguably become voiceless, with subsequent loss giving rise to the so-called 'incomplete phase', yielding synchronic closed syllables (Blevins and Garrett 1998). An explanation for the failure of syncope as sound change then cannot be found in the phonetics of the language. The Austronesian languages that have undergone syncope, as well as those that have not, all show $V.C^{V\cdot}CV$ sequences, but syncopating sound change is common where closed syllables pre-exist, and rare where they do not.

4.2 Muna, a Western Malayo-Polynesian language

Other languages with only open syllables show the same general resistance to syncopating sound change. While general final C-loss is rare outside of the Oceanic languages, it has occurred in many Western Malayo-Polynesian languages of eastern and southeastern Sulawesi and neighbouring islands, including Muna (Van den Berg 1989, 1991, 1996). In Muna, as in the Central Pacific languages, stem-final consonants are lost finally, but may be preserved under suffixation. Compare, for example Muna *kuli* 'skin' and *kulusi* 'peel', both from PMP **kulit* skin'. Unlike many Western Malayo-Polynesian languages which show synchronic reflexes of earlier historical syncope, there is no evidence of syncope in Muna. Compare PAn *qatəluR 'egg', Muna *ʀunteli*; PAn baqəRu 'new', Muna *buʀou*, *qapəju 'gall', Muna *ʀufei*: in all cases, trisyllables are maintained.

4.3 Malagasy

Malagasy is thought to be most closely related to the Southeast Barito languages of South Kalimantan (Blust to appear). While the loss of final closed syllables via final vowel paragoge is an areal feature of many Sulawesi languages, the majority of Southeast Barito languages have inherited WMP final consonants, and maintain final closed syllables. The majority of these languages also show evidence of historical unstressed vowel syncope.

However, modern Malagasy has no closed syllables at all. Medial clusters resulting from earlier syncopes have been eliminated, and all historical final consonants are either lost, or followed by an epenthetic vowel as illustrated in (4) for the Merina and Sakalava dialects.

(4) Malagasy final C-loss and V-insertion

a. *taŋan 'hand' > Merina *tánană,* Sakalava *táŋa*
b. *kulit 'skin' > Merina *húditra,* Sakalava *húlitse*
c. *putiq 'white' > Merina *futsi,* Sakalava *futi*

Whether the shift to open syllables is due to contact with Bantu languages (Dahl 1988) or not, the only evidence for syncope is that which occurred in the mother language of Malagasy and Kelabit, before the sound changes in (4). An example showing this early syncope is PAn *bekelaj 'spread out, unroll', Kelabit *beraʔaŋ,* Merina Malagasy *velatra*. There is no evidence for syncope in Malagasy post-dating its shift to an open-syllable-only language: compare **enem* 'six' Merina *énina*; **eseŋ* 'blow the nose', Merina *ésina,* etc. This is true even though the sound change of *ĕ > *i* makes unstressed *i* an expected syncope target in these word types.

4.4 Early versus Late Micronesian developments

Proto Micronesian lost inherited Proto Oceanic final consonants. As in other Oceanic languages, these consonants were maintained in suffixed forms, including transitive verbs and possessed nouns. After loss of these final consonants, Proto Micronesian shows only open CV syllables. Consistent with (1), there is no evidence of regular syncope in Micronesian languages until closed syllables re-evolve by a later sound change involving loss of word-final voiceless vowels (Rehg 1991). Within Micronesian, syncope is evident at this later stage for several languages, including Mokilese (MOK): PCEMP **saku layaR* 'sailfish, swordfish > MOK *daklar*; POC **qapaRa* 'shoulder' MOK *aprɔ* 'his shoulder'; POC **takuRu* 'back' > MOK *jarkin* 'his back'; MOK *pwalik/pwɔlkɔ* 'foot/3sg'. Note that MOK *pwirej* 'dirt' < **pwire-/jV* < PMP **budiN* 'charcoal, carbon, soot' supports the loss of final Vs before syncope, though the etymology differs from Harrison (1984), who suggests MOK *pwije* 'excrement' from POC *mpu(dr)i(t), and *pwirej* 'dirt' < *mpu(dr)i(t)-V.

4.5 A Polynesian exception that proves the rule?

The account above associates the absence of syncopating sound changes in Central Pacific with the open-syllable template of these languages. The absence of ambient closed syllables means that language learners are less likely to reinterpret $VC^{V}CV$ as VC.CV; where closed syllables have already been experienced and categorized as such, they can serve as templates to which ambiguous $VC^{V}CV$ strings can be matched. Among the Central Pacific languages, there are few exceptions to this association. However, at least one Polynesian Outlier, Mele-Fila of Vanuatu, is described with medial vowel syncope (Elbert 1965). Of particular interest is that this language has borrowed a substantial proportion of its vocabulary from neighbouring Efate dialects, which do have closed syllables and consonant clusters. Consonant clusters in borrowed words include /tl, np, nt, nf, ns, nm, nl, nr, rp, rk, rs, lt, ls, lm, st, sm, ft, fk, fm, mk, kt, km/, while those resulting from syncope include /tp, tf, tn, tl, tm, tv, sk, fr, pl/. In this case, contact-induced change

has increased the phonotactic complexity of Mele-Fila, endowing it with consonant clusters and closed syllables. This change, in turn, appears to be the trigger of subsequent syncope.

5 Discussion

While the facts surveyed above are generally consistent with Structural Analogy (1), and might be used to support non-phonetic structural influences on sound change, alternative explanations for syncope resistance clearly exist. Languages with only open syllables may be 'syllable-timed', with less vowel reduction than 'stressed-timed' languages (see Dauer 1983 for refinement and decomposition of these terms). With less vowel reduction, syncope may be less likely for purely phonetic reasons. An additional factor concerns the positioning of stress: perhaps vowels that syncopate in one language are stressed in another, due to stress shift. In the most comprehensive discussion of Proto Oceanic stress to date, Lynch (2000) documents shifts from the Proto Oceanic system which stressed final closed syllables, otherwise penult. In all cases, medial unstressed vowels are present once affixed forms are taken into account. Though it is true that languages with regular penultimate stress should only show syncope (2) in words of four or more syllables, words of this type are attested in the Oceanic languages under study, but do not show syncope.

An interesting question is whether the type of structural priming argued for in this study can be demonstrated in the laboratory. Experimental paradigms for testing phonological priming exist, though to date, they have been used primarily to explore the nature of phonemic categories. Mielke (2003) demonstrates the role of language-specific knowledge in the perception of /h/, including distributional properties, while Hallé and Best (2007) demonstrate significant language-specific phonotactic effects on the perception of coronal-lateral consonant clusters. Cutler et al. (2005) summarize a range of work on phonemic category plasticity: when listeners are exposed to phonetically shifted categories in lexical decision tasks (e.g. /f/s shifted phonetically towards /s/, or vice versa), subsequent categorization tasks show that (i) subjects have more inclusive phonemic categories depending on the shifts they are exposed to; and (ii) they extend this shift to other words. Could experience with open-syllable-only words induce listeners to categorize $VC^{V}CV$ sequences as VCVCV, while others, exposed only to VCCV shift $VC^{V}CV$ tokens to instances of VCCV? It is hoped that this brief survey of unstressed vowel syncope in Austronesian will stimulate experiments of this sort, as well as further cross-linguistic studies of syncopating sound change. Does unstressed vowel syncope occur significantly more often in languages with pre-existing closed syllables, as the principle of Structural Analogy suggests? For Austronesian languages, the answer appears to be yes. However, closer examination of these cases along with broader cross-linguistic surveys are clearly necessary before one can answer this question with any degree of certainty.

References

Andersen, Henning. 1990. The structure of drift. In H. Andersen and K. Koerner, eds *Historical linguistics 1987*, 1–20. Amsterdam: John Benjamins.

Best, Catherine C. and Gerald W. McRoberts. 2003. Infant perception of non-native consonant contrasts that adults assimilate in different ways. *Language and speech* 46:183–216.

Blevins, Juliette. 2004a. *Evolutionary phonology: the emergence of sound patterns.* Cambridge: Cambridge University Press.

—— 2004b. The mystery of Austronesian final consonant loss. *Oceanic Linguistics* 43:208–213.

—— 2005. The role of predictability in sound change: privileged reduction in Oceanic reduplicated substrings. *Oceanic Linguistics* 44.2:517–526.

—— 2006. A theoretical synopsis of evolutionary phonology. *Theoretical Linguistics* 32:117–166.

—— 2009. Phonetic explanations for recurrent sound patterns: diachronic or synchronic? In Eric Raimy and Charles E. Cairns, eds *Contemporary views on architecture and representations in phonology*. Cambridge, MA: MIT Press.]

Blevins, Juliette and Robert Blust. 2003. Towards a theory of drift: patterns of Austronesian syncope. MS, University of California, Berkeley.

Blevins, Juliette and Andrew Garrett. 1998. The origins of consonant–vowel metathesis. *Language* 74:508–556.

—— 2004. The evolution of metathesis. In Bruce Hayes, Robert Kirchner and Donca Steriade, eds *Phonetically based phonology*, 117–156. Cambridge: Cambridge University Press.

Blust, Robert. 1978. *The Proto-Oceanic palatals*. Wellington: The Polynesian Society.

—— 1990. Three recurrent changes in Oceanic languages. In J.H.C.S. Davidson, ed. *Pacific island languages: essays in honour of G.B. Milner*, 7–28. London: University of London, School of Oriental and African Studies.

—— 1993. Central and Central-Eastern Malayo-Polynesian. *Oceanic Linguistics* 32:241–293.

—— 1995. Austronesian comparative dictionary. MS, University of Hawai'i.

—— 1999. Subgrouping, circularity, and extinction: some issues in Austronesian comparative linguistics. In Elizabeth Zeitoun and Paul Jen-kuei Li, eds *Selected papers from the Eighth International Conference on Austronesian Linguistics*. Symposium Series of the Institute of Linguistics (Preparatory Office), 31–94. Academia Sinica No.1. Taipei: Academia Sinica.

—— 2000. Chamorro historical phonology. *Oceanic Linguistics* 39:83–122.

—— 2005. Must sound change be linguistically motivated? *Diachronica* 22:219–269.

—— 2007. Disyllabic attractors and antiantigemination in Austronesian sound change. *Phonology* 42:1–36.

—— 2009. *Austronesian languages*. Canberra: Pacific Linguistics.

Bowden, John. 2001. *Taba: description of a South Halmahera language*. Canberra: Pacific Linguistics.

Cutler, Anne, J.M. McQueen, J.M. and D. Norris. 2005. The lexical utility of phoneme–category plasticity. *Proceedings of the ISCA Workshop on Plasticity in Speech Perception*, 103–107. London.

Dahl, Otto Christian. 1988. Bantu substratum in Malagasy. *Études Océan Indien* 9:91–132.

Dauer, R.M. 1983. Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics* 11:51–62.

De Chene, Brent E. and Stephen R. Anderson. 1979. Compensatory lengthening. *Language* 55:505–535.

Elbert, Samuel H. 1965. Phonological expansions in Outlier Polynesia. *Lingua* 14:431–442.

Erickson, Blaine. 1996. [1998] Issues in Fijian phonology. *Hawaii Working Papers in Linguistics* 28:13–44.

Gordon, Matthew. 1999. Syllable weight: phonetics, phonology and typology. PhD dissertation. Los Angeles: UCLA.

Guion, Susan G. 1998.The role of perception in the sound change of velar palatalization. *Phonetica* 55:18–52.

Hallé, Pierre A. and Catherine T. Best. 2007. Dental-to-velar perceptual assimilation: a cross-linguistic study of the perception of dental stop + /l/ clusters. *Journal of the Acoustic Society of America* 121:2899–2914.

Hardcastle, William J. and Nigel Hewlett, eds. 1999. *Coarticulation: theory, data, and techniques*. Cambridge: Cambridge University Press.

Harrison, Sheldon P. 1984. Mokilese reflexes of Proto-Oceanic. In Byron W. Bender, ed. *Studies in Micronesian Linguistics*, 339–373. Canberra: Pacific Linguistics.

Hume, Elizabeth. 2004. The indeterminacy/attestation model of metathesis. *Language* 80:203–237.

Jacobson, Steven A. 1984. *Yup'ik Eskimo dictionary*. Fairbanks: Alaska Native Language Center.

Kavitskaya, Darya. 2002. *Compensatory lengthening: phonetics, phonology, diachrony*. New York: Garland.

Kiparsky, Paul. 1995. The phonological basis of sound change. In John A. Goldsmith, ed. *The handbook of phonological theory*, 640–670. Cambridge, MA: Blackwell.

Lehiste, Ilse. 1970. *Suprasegmentals*. Cambridge, MA: MIT Press.

Lynch, John. 2000. Reconstructing Proto-Oceanic stress. *Oceanic Linguistics* 39:53–82.

—— 2001. *The linguistic history of southern Vanuatu.* Canberra: Pacific Linguistics.

Maye, Jessica, Janet F. Werker and LouAnn Gerken. 2002. Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition* 82:B101–B111.

Mielke, Jeff. 2003. The interplay of speech perception and phonology: Experimental evidence from Turkish. *Phonetica* 60:208–229.

Myers, Scott and Benjamin B. Hansen. 2005. The origin of vowel length neutralisation in vocoid sequences: evidence from Finnish speakers. *Phonology* 22:317–344.

—— 2007. The origin of vowel length neutralization in final position: evidence from Finnish speakers. *Natural Language and Linguistic Theory* 25:157–193.

Ohala, John J. 1974. Phonetic explanation in phonology. In A. Bruck, R.A. Fox and M.W. LaGaly, eds *Papers from the Parasession on Natural Phonology*, 251–274.Chicago: Chicago Linguistic Society.

—— 1981. The listener as a source of sound change. In Carrie S. Masek, Robert A. Hendrick and Mary Frances Miller, eds *Papers from the parasession on language and behavior*, 178–203. Chicago Linguistic Society, Chicago, Il.

—— 1990. The phonetics and phonology of aspects of assimilation. In J. Kingston and M. Beckman, eds *Papers in laboratory phonology I*, 258–275. Cambridge: Cambridge University Press.

Rehg, Kenneth. 1991. Final vowel deletion in Micronesian languages: an exploration of the dynamics of drift. In Robert Blust, ed. *Currents in Pacific Linguistics: papers on Austronesian languages and linguistics in honour of George W. Grace*, 383–401. Canberra: Pacific Linguistics.

Reid, Lawrence. 1971. *Philippine minor languages: Word lists and phonologies. Oceanic Linguistics, Special Publication*, No. 8. Honolulu: University of Hawaii Press.

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

Saffran, Jenny R. and Erik D. Thiessen. 2003. Pattern induction by infant language learners. *Developmental Psychology* 39:484–494.

Sapir, Edward. 1921. *Language*. New York: Harcourt–Brace.

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

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

Tamata, Apolonia N. 1994. Phonological change in standard Fijian fast speech. Honolulu: University of Hawaii MA thesis.

Thurneysen, R. 1980. *A grammar of Old Irish*. Dublin: The Dublin Institute for Advanced Studies.

Tsuchida, Shigeru. 1976. *Reconstruction of Proto-Tsouic phonology.* Study of Languages and Cultures of Asia and Africa, Monograph Series, No.5. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa.

Van den Berg, Rene. 1989. *A grammar of the Muna language*. Dordrecht: Foris.

—— 1991. Muna historical phonology. In James N. Sneddon, ed. *Studies in Sulawesi Linguistics, part II. NUSA, Linguistic Studies of Indonesian and other languages in Indonesia 33*, 3–28.

—— 1996. *Muna–English dictionary*. Leiden: KITLV Press.

Wedel, Andrew. 2006. Exemplar models, evolution, and language change. *The Linguistic Review* 23:247–274.

—— 2007. Feedback and regularity in the lexicon. *Phonology* 24:147–185.

Werker, Janet F. and Richard C. Tees. 2002. Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant behavior and development* 25:121–133.

Whisler, Ronald and Jacqui Whisler. 1994. Sawai. In Darell T. Tryon, ed. *Comparative Austronesian dictionary*, 659–665. Berlin: Mouton de Gruyter.

6 *Irregular sound change and the post-velars in some Malakula languages*

JOHN LYNCH

1 Introduction

There are a number of thorny little problem areas in Austronesian historical phonology, most of which have been investigated, at one time or another, by Bob Blust. In this volume in his honour, I would like to raise—or rather, re-raise—another thorny little problem, similar to one raised in Blust (1996).[1]

I will be examining the reflexes of just two protophonemes in just a handful of relatively closely related Oceanic languages, but the issue with which I am concerned is a considerably broader one: the fact that, despite the general principle that sound change is regular (or regular with certain definable exceptions), there are cases where this principle seems not to apply, and where a particular sound change does seem to be irregular.

The two protophonemes I will be concerned with here are the Proto Oceanic (POc) post-velars: the stop *q and the trill *R. Both of these have caused historical linguists considerable problems of one sort or another—in terms of what their position of articulation was, in terms of their manner of articulation (at least for *R), and in terms of their widely varied reflexes—not least of which is the fact that they are probably lost more often than any other POc phonemes. I will concentrate on their reflexes in a number of Malakula languages for which we have reasonable amounts of data, most of which either have been only very recently published or are still unpublished.

I will be dealing in this paper with the reflexes of these two protophonemes in eleven languages spoken on the island of Malakula. The thirty or so actively spoken and moribund Malakula languages probably belong to a single sub-linkage within the Central Vanuatu

[1] Bob Blust and I were graduate students together at the University of Hawai‘i in the late 1960s and early 1970s. Although our paths have not physically crossed many times since then, I have always valued his comments on drafts of papers I have written, and have enjoyed reading his numerous and diverse contributions to Austronesian historical linguistics. I would also like to thank Andrew Pawley and an anonymous reviewer for comments on an earlier draft of this paper.

Alexander Adelaar and Andrew Pawley, eds
Austronesian historical linguistics and culture history: a festschrift for Robert Blust, 57–72.
Pacific Linguistics, 2009.
© This edition vested with Pacific Linguistics.

linkage of the Southern Oceanic group, whose other members contain the non-Polynesian languages of the rest of Vanuatu and New Caledonia. There are probably two major subgroups of Malakula languages, though at this stage of research I cannot be definite about this, since there are a few languages which I cannot yet neatly classify. The languages I am dealing with here, together with their subgroup affiliation and the main sources of data, are as follows (within each subgroup I list languages roughly from north to south, and within examples I list languages in the order below):[2]

Eastern: Nese (Crowley 2006c)
Tirax (Amanda Brotchie, dictionary file)
Avava (Crowley 2006a)
Uripiv dialect of Northeast Malakula (Ross McKerras, dictionary file)
Unua (Elizabeth Pearce, dictionary file)

Western: V'ënen Taut (Fox 1979; Greg Fox, dictionary file
Tape (Crowley 2006d)
Naman (Crowley 2006b)
Neverver (Julie Barbour, dictionary file)
Neve'ei (Musgrave 2001; Jill Musgrave and Terry Crowley, dictionary file)
Nāti (Crowley 1998)

2 Proto Oceanic *R

Lynch, Ross and Crowley (2002:64) state that POc *R 'was probably a uvular trill, which is frequently lost or merged with a liquid (*r or *l) in daughter languages'. Wolff (2003:7), on the other hand, states that the ancestral phoneme in Proto Austronesian was 'a voiced back spirant or possibly a back stop', which he writes as *ɣ rather than *R and which he treats as the voiced equivalent of *q. Given its reflexes in Oceanic languages—often a liquid or a fricative, and (almost?) never a uvular trill—it is possible that Wolff may be closer to the mark in assessing the phonetic nature of *R.

Geraghty (1990:51) says that, 'in the historical phonology and classification of Oceanic languages, probably no phoneme has been more extensively studied and used than *R'. His very thorough study of reflexes of Proto Eastern Oceanic (PEOc) *R shows that this protophoneme was definitely retained in all Vanuatu languages in some lexical items; where it is retained, it appears to merge with *r. He postulated that '*R is lost in proportion to distance from Western Oceanic, beginning in the Southeast Solomons' (1990:90). Based on whether *R is retained or lost in particular lexical items, he suggested three major 'boundaries' or isogloss bundles in Vanuatu: one between Mota and Raga, a second between Paama and Namakir, and a third between central Vanuatu and Erromango; this would divide the languages of Vanuatu into four groups: far north, north-central, Epi-Efate and southern.

---

[2] It goes without saying that I am extremely grateful to Julie Barbour, Amanda Brotchie, Greg Fox, Ross McKerras, Jill Musgrave and Elizabeth Pearce for so freely making available their unpublished data, and to the late Terry Crowley, who first got me interested in comparative Malakula linguistics.

Clark's (to appear) Proto North-Central Vanuatu (PNCV) reconstructions illustrate the nature of the general problem.[3] Examine the following:

(1)	POc/PEOc	PNCV	
a.	*bakuRa	*bakura	'*Calophyllum inophyllum*'
	*Ropok	*rovo	'to fly'
b.	*cakaRu	*sakaRu	'coral reef'
	*draRaq	*daRa	'blood'
c.	*baReko	*baeko	'breadfruit'
	*Raka	*aka	'k.o. vine, *Pueraria*'

Cases like those in (1a) show *R merging with *r and being reflected by the reflex of *r in all or most of the languages which he considered. With those in (1b), on the other hand, *R merges with *r only in a few languages—usually Mota in Clark's sample, and further investigation suggests that it is only the northernmost languages (Torres and Banks Is.) which retain it—but is lost in the remainder. Finally, cases like (1c) illustrate those where *R is apparently universally lost in PNCV.

However, there is some counter-evidence to Geraghty's hypothesis—cases where *R is retained further south but apparently lost further north. Just a few examples are listed below:

(2)	POC/PEOC		*R lost further north	*R retained further south
	*cakaRu	'coral reef'	NE Ambae sakau, S Efate n/skau	Sye i/kri
	*cuRi-	'bone'	Raga hui-, Paamese sī-	Sye no/ura-
	*Ruap	'high tide'	Tamambo ua, Paamese ue	Lenakel e/lu-elu, Kwamera a/rə-ruk^{w}

Let us now turn to the Malakula data to see what patterns can be identified in a lower-level subgroup of NCV with a large number of members. The first thing that needs to be said is that, in general terms, the retention or loss of *R is fairly consistent across the languages I have been working with. That is to say, if *R is retained in one of these languages in a particular lexical item, it is likely to be retained in all of them. Below are a few examples which illustrate this.[4]

---

[3] Clark in various publications has proposed that there is a North-Central Vanuatu (NCV) subgroup or linkage, and has made a large number of lexical reconstructions (Clark to appear). While the integrity of this grouping is still under discussion within the wider context of the Southern Oceanic hypothesis (Lynch 2000), the reconstructions nevertheless have considerable validity.

[4] Reconstructions are POc unless preceded by N or C, indicating that they are attributed respectively to PNCV or PCV only. Blanks indicate no cognate or no data, while italics indicate that the form does not follow the general trend. V'ënen Taut and Nese have apicolabial consonants, written as *m̪*, *b̪*, etc. In Tirax, these apicolabials have changed further to alveolars: note *m > *n* in *naxnal* < *na-kamaliR 'men's house'. A number of languages have a prenasalised bilabial trill, written *ʙ*.

Table 1: Regular retention of *R

	*Rapi ‘evening’	N *vaRa ‘hand’	*kaRi(a) ‘*Cordyline*’	*maRaŋo ‘dry (coconut)’	*ma-wiRi ‘left (side)
Nese	revrav	na/vara-		naraŋ	
Tirax	rɛvrɛv	vra-	karɛ	mraŋ	maɛr
Avava	ki/drap	vara-	a/ari		m^{w}iir
Uripiv	rivriv	ne/vre-	gari	raŋraŋ†	mair
Unua	ɾevɾev	veɾe-		mɾaŋ	ɣi/maiɾ
V’ënen Taut	kəna/ray̰	*va-*		m̰əran	m̰ir
Tape	rivrip		na/arə/s	məraŋ	mor/ne-
Naman	revrev	nə/verə-	na/ɣari	meraŋ	ɣə/mir
Neverver‡	livrav	ne/vra-	na/xari	mallaŋ	mer
Neve’ei	rivirav	ne/vera-	na/ʔari	meraŋ	m^{w}iyir
Nāti	revrev	ni/vara-	na/ʔari		

† Uripiv unexpectedly loses initial *ma- in this form.
‡ The occasional *l* and *ll* reflexes in Neverver are unexplained.

Similarly, if *R is lost in one of these languages in a particular lexical item it is likely to be lost in all of them:

Table 2: Regular loss of *R

	*Rum^{w}aq ‘house’	*kuRita ‘octopus’	*tapuRiq ‘conch’	*tuqaRi ‘long ago’	*kamaliR ‘men’s house’
Nese	na/ine, n/em-	ne/ɣte	tavu	tua	na/ɣm̰al
Tirax	na/in		n/tav	tuɛ	na/xnal
Avava	i/im	koit	a/taap	tua	amal
Uripiv	na/im	na/it	davö	tuwi	loln/amel
Unua	na/im	ɣuti/nbon	davu	tue ‘forever’	ɣemeɾ
V’ënen Taut	nə/maɣ	ɣut	na/tav	ti/tuei	n/am̰el
Tape	nə/maɣ		tivwi	te/two	n/imel
Naman	ne/maɣ	ni/ɣət		toɣe	na/amil
Neverver	a/iem	no/xoit	ne/tav	tue	na/xamal
Neve’ei	ni/yim	no/ɣoit	ne/tavu	tuɣoi	na/ʔamal
Nāti		no/ʔoiyit	tāvu		na/ʔamel

These languages exhibit an overall consistency in their treatment of *R—whether it is retained or lost—and there are only a handful of cases which show considerable inconsistency; these are illustrated in the table below, where retentions are in regular font and losses in italics:

Table 3: Irregular treatment of *R

	N *[vo]m^{w}aRaki ‘ground dove’	*paRage ‘*Pangium edule*’	*takuRu- ‘back’	*suRuq ‘juice, liquid’
Nese	*no/vomaɣ*			ne/jira-, *nu/suwu-*
Tirax	*vɛmɛx*		*taxu-*	
Avava	a/pmar	b^{w}iki/var	u/dru-	a/sur, *e/s-*
Uripiv	wum^{w}er	bik/wer	*vitu*	*suwe-*
Unua	na/maɾ		dure-	*sue-*
V’ënen Taut	na/vimar	*na/ɣ̰ak*	*du-*	*ui-*
Tape	na/v^{w}imar	*nə/vək*	*e/taɣ* ‘behind’	jəre-
Naman	vomar	big/var	ne/tre-	*ne/swe-*
Neverver		nibig/var		
Neve’ei	ne/vimar	nabugu/var	*ne/taʔa-*	*na/s-*
Nāti	ne/vimar	*ne/vaŋk*	*ni/taʔu-*	

With a few exceptions, then, *R behaved fairly consistently in these languages (and indeed in PNCV), in the sense that it was retained in all languages in certain lexical items and lost in all languages in certain others. But is there any consistency in the patterns of loss and retention?

One general statement that can be made is that *R was lost in absolute final position. There is one possible case of retention of root-final *R when followed by possessive suffixes, but note that the reflexes show an additional vowel following the reflex of *R:

(3) *ikuR- ‘tail’ > Uripiv n/erure-, Unua ɣoɣoɾe-, Neverver no/xore-, Naman no/ɣoɣorə-

With the form *maluR ‘shade, shadow’, final *R seems to have been retained in Avava *milier* and Neve’ei *ni/milier*, but lost in Uripiv *na/mol*, Neverver *ni/milu-*, V’ënen Taut *m̰əle/ka-* and Tape *məlmilə-*. There is no evidence of final *R being retained in any of these languages in the following etyma (though of course we do not have reflexes for every etymon in every language, either because the data are inadequate or the etymon was lost):

(4)

*lasoR	‘testicles’	*qatoluR	‘egg’	*maturuR	‘sleep’
*qipaR	‘in-law’	*waiR	‘water’	*kamaliR	‘men’s house’
*roŋoR	‘hear’	*saliR	‘float’	*madraR	‘fermented breadfruit’
*toŋoR	‘mangrove’	*niuR	‘coconut’	*pusuR	‘bow and arrow’
*sinaR	‘shine’	*rapu(R)	‘ashes’		

However, as Tables 1–3 show, there are cases of both retention and loss of both root-initial and root-medial *R.[5]

Geraghty (1990:85) suggested that there may have been some phonological conditioning: e.g., initial *R was likely to be lost before *u but retained before *a, medial *R was more likely to be retained between identical vowels, etc. However, there does not seem to be any compelling evidence for this in the languages I have been examining, although there are a few trends. Table 5 summarises retention and loss of non-final *R in all vocalic environments.

[5] There are virtually no cases of POc initial *R occurring word-initially in these languages, since nouns generally have a fused article and verbs normally take one or more prefixes.

Table 4: Retention / loss of *R in vocalic environments

V1↓ V2→	a	e	i	o	u	TOTAL
a	12 / 1		3 / 5	4 / 0	4 / 5	**23 / 11**
e	1 / 0		0 / 1		1 / 0	**2 / 1**
i	2 / 1		2 / 0			**4 / 1**
o	2 / 2	1 / 0		1 / 1	0 / 1	**4 / 4**
u	3 / 0		1 / 5		2 / 0	**6 / 5**
TOTAL	**20 / 4**	**1 / 0**	**6 / 11**	**5 / 1**	**7 / 6**	**39 / 22**

No clear-cut patterns emerge from the figures in Table 4. There are a couple of apparent tendencies, but none of these is exceptionless:

- there is a strong tendency for *R to be retained before *a and *o;
- there is, as Geraghty suggested, a tendency for *R to be retained between identical vowels (the figures from Table 4 are 17 retentions, 2 losses); and
- there is a tendency for *R to be lost when before *i (except / *i __ i).

However, there appear to be no particular patterns in relation to other vocalic environments; and, indeed, there are apparent 'minimal pairs', where *R is retained in one item and lost in exactly the same vocalic environment in another item. For example, *R is lost in the environment *u __ i in reflexes of *tapuRiq and *kuRita in Table 2, but retained in the following:

(5) *tuRi 'sew' > Nese rur, Tirax drur, Avava tur, Uripiv o/tri, Neve'ei dur, Nāti tur

Similarly, *R in the environment *a__u is retained in (6a) below but lost in (6b):

(6) a. *yaRu 'casuarina' > Nese n/iar, Avava iar, Uripiv n/ur, Vënen Taut ne/ier, Tape n/iar, Neve'ei n/iar, Nāti n/iar

b. *paRu '*Hibiscus tiliaceus*' > Tirax na/ve, Uripiv vava, V'ënen Taut v̥iv̥ei, Tape vive, Naman nə/veve

Apart, then, from the almost universal loss of POc final *R, it is not possible to define phonologically with any exactitude the conditions under which *R is retained or lost. Its retention or loss appears on the surface to be quite random: it is retained in some lexical items, but lost in others.

3 Proto Oceanic *q

When we examine the reflexes of *q (apparently a uvular or back velar stop in POc) in these languages, we find a far greater degree of inconsistency than we do with *R.

The only Vanuatu language which regularly retains *q is Namakir, in which the reflex is the glottal stop (Sperlich 1989):

(7)

POc		Namakir	POc		Namakir
*qasu	'smoke'	ʔah	*toqa	'fowl'	toʔ
*qusan	'rain'	ʔih	*punuq	'killed'	biniʔ
*paqoRu	'new'	boʔo	*daRaq	'blood'	daʔ
*maqetom	'dark'	maʔet	*taRaq	'cut'	daʔ

Recently, however, Lynch and Crowley (2003) pointed to occasional retentions of *q in a few Malakula languages, and Lynch (2004) also noted occasional retention of *q in Southern Vanuatu languages. In this paper, I expand considerably on Lynch and Crowley's discussion.

3.1 The overt reflexes of *q

POc *q is lost far more often than it is retained. In this section I discuss the reflexes of *q when it is not lost, without paying any attention to patterns of loss and retention, which I will discuss in §3.3. These reflexes are listed in Table 5, along with the reflexes of *k for comparison; parentheses indicate a conditioned reflex, while the notation *i-i-* is shorthand for *i-*, *-i-* (i.e. *i* initially and medially).

Table 5: Overt reflexes of POc *q and *k in eleven Malakula languages

	*q-	*-q-	*-q	*k
Nese	-i-	-i-	∅	ɣ
Tirax	-i-		-x	x
Avava	i-, y-; -k-		-k-k	k (∅)
Uripiv	-i-		-i-	∅
Unua	ɣ-ɣ-		-ŋ-ʔ; -gʔ	ɣ
V'ënen Taut	i-i-; -ɣ-	-ɣ	-ɣ-ɣ	ɣ (∅)
Tape	i-i-; w-w-	-w-; -ɣ-ɣ	-ɣ-ɣ	ɣ (∅)
Naman	i-i-; ɣ-ɣ-	-ɣ-	-ɣ-ɣ	ɣ (∅)
Neverver	x-x-; -i-	-i-	-x-x; (-k)	x
Neve'ei	ʔ-ʔ-; -i-; -ɣ-; w-w-	-ɣ-	-ɣ-ɣ; (-ʔ-)	ɣ (ʔ)
Nāti	-w-	-w-, -ʔ	-ʔ	ʔ

Final *q seems to have been retained in all of the eleven languages except Nese in at least one etymon, though in three of these (including the problematical Unua reflex of *tob^{w}aq below) only one etymon seems to show retention. In all cases except Uripiv (see *tob^{w}aq below) the reflex is a velar or glottal obstruent. Examples:

(8) *Rum^{w}aq 'house' > V'ënen Taut nə/maɣ, Tape nə/maɣ, Naman ne/maɣ
*mimiq 'urinate' > Avava memek, V'ënen Taut məɣei, Tape moɣ/wo, Neverver maxmax, Neve'ei maɣmaɣ
*tob^{w}aq 'belly' > Tirax təbax, Uripiv depai-, Unua dabaŋo-, dobog, Naman dabaɣa-, Neverver ni/demxe-, Neve'ei ne/tabaʔa-
*mataq 'new, raw' > Tirax mdrax, V'ënen Taut m̥ədaɣ, Tape mədaɣ, Neverver mrex
*luaq 'vomit' > Tape luaɣ, Neverver lialuk, Neve'ei yoɣyoɣ ?

Excluding the Uripiv and Unua cases, the obstruent reflex of *q is the same as the reflex of *k in all languages (except Nese, which has no obstruent reflex). Unua seems to show voicing crossover from *k to *g.

In non-final position the reflexes of *q appear to be both a velar obstruent (or a glottal stop) and a high vowel/semivowel. Interestingly, medial *q is rarely retained; but initial *q is retained more frequently. When it occurred before a back vowel, *q is normally reflected as *w*:

(9)	*qone	‘sand’	>	Tape nun/win, Naman dabano/wen, Neve’ei ne/wenwen, Nāti nempin/wen
	*qudu	‘palolo worm’	>	Neve’ei nu/wud
	*quloc	‘maggot’	>	Tape wiləs
	*qutin	‘penis’	>	Neve’ei nu/wus-
	*maquRip	‘alive’	>	Nāti mewur

but there are a couple of cases where some other reflex is found in some languages:

(10)				*q > *w*	Other reflex
	*quluŋ-an	‘pillow’	>	Neve’ei wulwul (v.)	Avava u/kulaŋ, V’ënen Taut n/iululna-
	*qusan	‘rain’	>	Nāti nu/wuh (n.)	Tape iu (v.), Naman ius (v.)
	*qutan	‘inland’	>	Nāti ne/wut	Unua ve/ɣut, Naman ɣaut, Neve’ei ʔout
	*leqo-	‘voice’	>		Nāti ni/loʔ

When not before back vowel, *q is sometimes retained as a velar or glottal:

(11)	*qalawa-	‘sibling’s child’	>	Neverver xala- ‘nephew, uncle’
	*qaloŋo	‘*Acanthurus* sp.’	>	Unua va/ɣaro
	*qapaRa-	‘shoulder’	>	Naman no/ɣoverə- ‘wing’, Neverver na/xarevra- ‘wing’
	*qatoluR	‘egg’	>	Unua ɣori- V’ënen Taut na/ɣadrəl, Neve’ei na/ʔadle-
	*laqia	‘ginger’	>	Tape ləɣləɣ
	*tuqaRi	‘long ago’	>	Naman toɣe, Neve’ei tuɣoi, duɣoi

In other cases, the reflex is *i*. I give the data first (and see also *quluŋ-an and *qusan in (10) above), and then discuss it.

(12)	*qaŋaRi	‘*Canarium*’	>	Avava yaŋa, Uripiv n/iŋi, Tape n/ieŋe, Neverver n/iŋa, Neve’ei n/iŋi
	*qaRa(r)	‘fence’	>	Nese, Tirax, Neverver, Neve’ei n/iar
	*qasu	‘smoke’	>	Nese n/ies, V’ënen Taut ie-naɸ, Naman ies (v.), n/iisə- (n.)
	*qatop	‘sago’	>	Nese, Tirax, Uripiv, Neverver, Neve’ei n/iat, Avava iat, V’ënen Taut ne/iet; Tape, Naman n/iet

Generally, there is consistency between these languages as to whether a velar or a high vowel is the reflex (though not as to whether *q is retained or not). The one significant case of inconsistency is the following:

(13)	*qase-	‘jaw, chin’	>	**velar:** V’ënen Taut na/ɣ-, Neverver na/xas-, Neve’ei na/ɣase-
				i: Nese n/ias-, Uripiv n/ise-, Tape n/isi-

Now *i* occurs as a putative reflex of *q mainly when *q was initial and mainly in nouns, and thus it often occurred preceded by an article. The comparison *na-qatop ‘sago’ >

V'ënen Taut *ne/iet*, however, is the only case where we can be absolutely sure that *q became *i*; that is:

*	n	a	-	q	a	t	o	p
	↓	↓		↓	↓	↓		
	n	e		i	e	t		

In all other cases, we need to assume that *na-qa.. > *na-ia.. > nia..* (sometimes further > *nie..* or just *ni..*). However, there is some evidence from Avava which suggests that this is the correct interpretation. Avava reflects the POc article *a, not *na, and this often underwent vowel harmony (thus *bokasi 'pig' > a/buah, *mwata 'snake' > a/mwat, *toqa 'fowl' > o/to, *paRi 'stingray > e/ve, *mwele 'cycad' > i/mwil, etc.). Noun-initial *a* can reflect both the article (when retained) or the first vowel of the root (when the article was not fused),[6] and note that in the latter case there is no prothetic consonant (as there is with a number of languages—e.g. Fijian underlying word-initial *a* takes a prothetic *y*):

(14)

POc		Fused article	POc		Root-initial *a*
*a-baga	'banyan'	a/baŋ	*kamaliR	'men's house'	amal
*a-manuk	'bird'	a/man	*katabola	'dragon plum'	atipol
*a-niuR	'coconut'	a/ni	*kavika	'Malay apple'	avik

It will be seen from the data in the right hand column that root-initial *a* generally occurs when initial *k has been lost. In cases where this happens and the article is retained, it coalesces with root-initial *a* as a long vowel:

(15)

POc		Avava
*a-kabu	'fire'	a/aʙ
*a-karia	'cordyline'	a/ari

There are only two *ia*-initial nouns in the Avava lexicon: *iar* 'casuarina' < *yaRu and *iat* 'sago' < *qatop, but there is also *yaŋa* 'canarium' < *qaŋaRi. I suggest that in the latter two cases the *i* and *y* clearly reflect *q, and that *a-qatop > *a-iatop* > *iat*, while *yaŋa*, being bi-moraic, did not take the article. That is, there is no evidence of prothetic *i* or *y* in Avava, and there is evidence that *q > *i* or *y*.

Although we can predict that final *q will merge with *k as a velar obstruent and that *w* is the usual reflex of *q before a back vowel, it is more difficult to predict when non-final *q is reflected as a velar and when it is reflected as *i* or *y*, since both are found in similar environments (particularly *na- __ a).

3.2 The reflexes themselves

The discussion above has shown that, when retained, final *q is normally reflected as a velar or glottal obstruent, while non-final *q is reflected as *w* before *u and as *i* or the corresponding glide or else a velar elsewhere. Although not strictly germane to the overall topic of sporadic reflexes, it is worth briefly discussing why these different reflexes may have developed.

---

[6] Avava is one of a number of Malakula languages in which the article is fused if the root contains one mora but is normally not retained if the root is bi- or multi-moraic (see Lynch 2007)—though *a-karia > *a/ari* in (15) seems to be an exception.

Final *q merges with *k in all of the eleven languages which reflect it. Given that POc *q was close to *k in articulatory terms, there is nothing very unusual about it having a velar reflex.

It is unusual, however, for *q to have a high vowel/glide reflex, which is what happens with some cases of retained non-final *q. First, though, this clearly shows that *q in this environment did not merge with *k; and so we can assume that *q was retained as a distinct phoneme in any 'Proto Malakula'. I assume that *q eventually became a glide in this environment, later vocalising in some environments: a direct *q > *i* change seems highly unlikely.

There is, however, some unpredictability as to what the non-final reflex is. Some languages show only a velar (Unua), some only a high vowel or glide (Nese, Uripiv, Tirax), but others show both, and there seems to be no phonological conditioning involved. Indeed, considering *na-qa-initial forms, *q surfaces at least once as *i* and at least once as a velar in Neverver, V'ënen Taut, Naman and Neve'ei.

One possible explanation may be that *q became a fricative in this environment, which was basically intervocalic. This fricative may have been uvular or velar and, given the intervocalic environment, it may have been voiced rather than voiceless: i.e. [ʁ] or [ɣ]. (This must have predated any lenition of *k in this environment, though.) If this was the case, we would then need to assume [ɣ] > [w] before back vowels and sometimes > [y], sometimes to (or remaining) a velar before non-back vowels. Both seem to be more natural developments.

3.3 Loss and retention of *q

Having established the reflexes of *q when it is retained, I move now to look at overall patterns of retention and loss. The first thing that needs to be said is that the situation is somewhat different from that of *R: even in cases where some languages show retention of *q, others show loss in the same etymon. Table 6 illustrates this: retentions are in regular font, losses in italics.

Table 6: Irregular treatment of *q

	*qase- 'chin, jaw'	N *qaŋaRi *'Canarium indicum'*	*tuqaRi 'long ago'	*mimiq 'urinate'
Nese	n/ias-	*n/eŋa*	*tua*	
Tirax	*n/hɛ-*	*n/ɛŋa*	*tuɛ*	*muŋɛ*?
Avava	*as-*	yaŋa	*tua*	memek
Uripiv	n/ise-	n/iŋi	*tuwi*	*meme*
Unua	*n/ese-*	*n/eŋe*	*tue* 'forever'	*me, meme*
V'ënen Taut	na/ɣ-		*ti/tuei*	məɣei
Tape	n/isi	n/ieŋe	*te/two*	moɣwo
Naman	*na/ase-*	*n/eŋe*	toɣe	*mimi*
Neverver	na/xas-	n/iŋa	*tue*	maxmax
Neve'ei	na/ɣase-	n/iŋi	tuɣoi, duɣoi	maɣmaɣ
Nāti	*n/ehe-*	*n/eŋei*		*mimim*

POc *q has no overt reflex in any of these languages in around 60 etyma. Figures for its retention (excluding doubtful cases) in each of the eleven languages are as follows:

Table 7: Retention of *q

	*q-	*-q-	*-q	**TOTAL**		*q-	*-q-	*-q	**TOTAL**
Nese	4	1		**5**	V'ënen Taut	3	3	5	**11**
Tirax	2		1	**3**	Tape	9	4	5	**18**
Avava	3		2	**5**	Naman	6	1	2	**9**
Uripiv	3			**3**	Neverver	6	1	5	**12**
Unua	3		1	**4**	Neve'ei	10	1	5	**16**
					Nāti	5	3	1	**9**

It is clear from Table 7 that *q is retained in only a minority of cases. It is also clear that retention figures are higher among languages of the putative Western subgroup (on the right of the table above) than those of the Eastern subgroup (on the left).[7]

I do not need to list many examples of loss of *q. The following handful shows no retention of *q in four phonological environments in which it is retained in some items in some languages (see §3.1), showing that there seems to be no phonological conditioning involved.

Table 8: Loss of *q

	*qupi 'yam'	C *kumaqu '*Intsia bijuga*'	*suRuq 'fluid, juice'	*puaq 'fruit'
Nese			ne/jira-, nu/suwu-	no/vo-, nu/vu-
Tirax				bi/vu
Avava	o/ovi 'k. yam'		a/sur	va/na-
Uripiv	n/ov 'k. yam'		suwe-	we/ne-
Unua				ve/ne-
V'ënen Taut		ɣəmau	ui-	na/va-
Tape		nə/ɣmo	jəre-	no/vo-
Naman		no/ɣmo	ne/swe-	na/va-
Neverver		nu/xuma		
Neve'ei	n/obi 'k. yam'	nu/ɣumo	na/s-	ne/v^{w}e-
Nāti		ne/ʔumou		nö/van

To summarise: *q was normally lost, but was sometimes retained. Although we can make some predictions about what its reflex is when it is retained, we can make no prediction on any phonological basis about whether or not it is retained (a) in a particular etymon or phonological environment and (b) in a particular language—apart from just possibly suggesting that *q is more likely to be retained than lost initially before *a. (Even two such closely related languages as V'ënen Taut and Tape are inconsistent in their retention of *q in particular items.)

[7] The amount of data available varies from language to language, but the disproportionate number of retentions of *q in the west has nothing to do with that. We have, for example, more data on Uripiv than any of the others, and yet Uripiv has just about the smallest number of retentions; we have more data on Tirax, Unua and Avava than on Nāti, and yet the latter shows at least twice as many retentions.

4 Summary of Malakula data

To summarise the data on these two protophonemes in these languages:

- with *R, there is consistency among the various languages as to whether or not *R was retained in any particular lexical item;
- with *q, there is no consistency among the various languages as to whether or not *q was retained in any particular lexical item;
- with both, there is basically no consistency in terms of phonological environment as to when each was retained and when it was lost, apart from one or two minor trends; and
- with *q there is no consistency as to whether its non-final reflex is a glide or a velar obstruent.

The behaviour of *R in these language I have been looking at is not dissimilar to what happens to *R elsewhere in Vanuatu, so to that extent Malakula is a microcosm of the rest of Vanuatu. However, the behaviour of *q is quite different: *q is regularly retained in Namakir (and nowhere else in Vanuatu) and in parts of New Caledonia; elsewhere in this area it seems to be regularly lost (with the exception of the very few apparent retentions in Southern Vanuatu described in Lynch 2004). The Malakula case, then, is *not* a microcosm of the rest of this area as far as *q is concerned.

The regular retention of *q in Namakir and New Caledonia implies that *q was regularly retained in Proto Southern Oceanic and its various daughter-languages (unlike *R, which was irregularly lost). The total loss of *q in many of the descendants of Proto Southern Oceanic thus has to be seen as having occurred through a number of independent but probably identical changes. However, at least as far as Malakula languages are concerned, *R and *q are similar in that they are sometimes retained and sometimes lost.

5 Irregular sound change

One of the tenets of historical linguistics is that sound change is regular. This dates back to the latter decades of the nineteenth century, when the Neogrammarians (*Junggramatiker*) declared that sound 'laws' were without exception. Since then, the view has moderated somewhat, but it can be fairly said that most historical linguists would hold that sound changes are *generally* regular, and that while there may be exceptions all or most of these can be explained in some way.

5.1 Exceptions to regular sound change

Three of the explanations often put forward—quite legitimately—to explain exceptions are analogy, avoidance of homophony and contact between related languages. Koch (1996:220), for example, shows the operation of analogy to produce a phonological irregularity in Semitic languages. Note first the following verb inflections:

(16)

	1SG	2SG
Proto Semitic	*-ku	*-ta
Akkadian	-ku	-ta
Arabic	-tu	-ta
Ethiopic	-ku	-ka

Akkadian is regular, but Arabic and Ethiopic 'are each assumed to have altered their consonant by analogy with the other member of their respective paradigms': Arabic shows irregular *k > *t* in 1SG, Ethiopic irregular *t > *k* in 2SG.

Campbell (1996:77–78) notes how avoidance of homophony may bring about lexical replacements and phonological irregularity, and gives as one example of the latter certain German dialects in which two regular changes—loss of intervocalic *g* and unrounding of *ü*—would have meant that *liegen* 'lie (down)' and *lügen* 'lie (= tell lies)' would be homophonous. Instead, these changes do not apply in just these two words, to avoid this homophony.

There are numerous cases within Oceanic of contact between related languages producing what appear to be irregular phonological changes: indeed, this irregularity is one of the techniques for identifying borrowings. For example, in non-Polynesian languages of Southern Vanuatu POc final vowels are regularly lost: *rua 'two' > Lenakel *k/iu*, *tolu 'three' > *kə/sil*, *kani 'eat' > *kən*, etc. Cases of irregular retention of final vowels do occur: *kiajo 'outrigger boom' > *nə/kiatu*, *jila 'sail' > *tila*, etc. But it is apparent that these words have been borrowed from the nearby Polynesian Outlier language Futuna-Aniwa, which retains POc final vowels.

However, while most sound changes may well be perfectly regular, or may be almost perfectly regular but have some explainable exceptions, this is not always the case. For example, Blust (1996:137) says that:

> To a large extent the success of the Neogrammarian hypothesis has stemmed from the apparent fact that sound change *is* overwhelmingly regular. Where irregularities exist it has generally been found possible to explain them (or, all too often, to 'explain them away') as products of borrowing, analogy, or some other mechanism of secondary change. At the same time the problems associated with the Neogrammarian hypothesis stem from two apparent facts which may conflict with it: (1) The regularity of phonological change is an epiphenomenon rather than a primary datum, and (2) despite its overwhelming regularity, not *all* sound change is regular.

And again:

> Lest I be misread, let me emphasize in the strongest terms that I do not advocate a facile acceptance of irregularity in sound change. Every effort should be made to find rule-governed explanations for the primary observations. But when plausible explanations for irregularity cannot be stated it is pointless to resort to mechanical contrivances out of fear that the only alternative to such ad hoc solutions is to open a Pandora's box of methodological chaos. Irregularity is not mere chaos. Rather, … irregularity appears to be an integral part of the natural process of language change
>
> (Blust 1996:153)

A number of studies in Durie and Ross (1996) examine real or apparent irregularity in phonological development. In their introduction, the authors discuss lexical diffusion of phonological changes in a number of the case studies in the following terms:

> [A]t least in some of these cases, a lexically diffusing sound change has been halted before it completed its journey through the lexicon. … [A] speaker-oriented version of the hypothesis can be formulated. It says:
>
> (i) that each speaker who adopts a sound change does so first as part of the orderly variation of that speaker's speech;
>
> (ii) that this variation progressively shifts in favour of the 'new' sound;
>
> (iii) that (ii) applies initially only to certain items in the lexicon which contain the relevant sound in the relevant environment, then progressively on to other elements. (Durie and Ross 1996:23)

The assumption is that with a wholly regular change, process (iii) continues to apply until all appropriate items are affected for all speakers who have undergone the change. With an irregular change, however, process (iii) stops before applying to all appropriate items.

5.2 POc *R and *q in Malakula

Can the sporadic loss of *R and *q in Malakula be explained by factors like analogy, avoidance of homophony or contact with related languages? Or do these changes fit the truly irregular case? And if the latter, why?

I cannot see that analogy has any role here, since this usually occurs in cases like the Semitic one above, where paradigm sets are involved; in the Malakula cases I have been talking about, we are dealing with consonants which are root-initial, -medial and -final, and affixal morphology does not enter into it.

Avoidance of homophony also does not seem to be an issue here. Let us take the case of *R. We can make one of two assumptions:

a. The regular change was *R > *r*, but *R was lost to avoid homophonous forms with *r* from other sources (mainly *r). As an imaginary example, assume a form *maRu ‘dog’ which should become *maru*, but because there is already *maru ‘snake’ > *maru* the former loses *R to avoid this homophony and becomes *mau*.

b. Conversely, the regular change was *R > ∅, but this was blocked when a homophonous form would have been produced, and *R remained *r*. As another imaginary example, assume a form *baRi ‘tree’ which should become *pai*, but because there is already *bai ‘fruit’ > *pai* the former retains *R to avoid homophony (*baRi becoming *pari*).

This is the kind of change which might account for one or two—even perhaps half a dozen—exceptions (as in the German case above), in perhaps one or two languages. But I cannot see this happening on such a large scale as we have with reflexes of *R and, to a slightly lesser extent, *q in Malakula languages—a large scale in terms of both the number of etyma and the number of languages involved.

Contact with neighbouring and related languages was probably frequent in Malakula, given the small size of most of the language-communities there, and borrowing from one language to another undoubtedly took place. Given the situation described for *q above, it *might* be just possible to imagine that some language(s) which regularly lost *q borrowed the occasional etymon from some other neighbouring language in which *q was retained (or vice versa), even though in making this assumption we would have to allow for rather widespread borrowing of ‘basic’ vocabulary (and for the fact that many cognate morphemes are formally different in other respects as well). But this situation is just unimaginable with *R, given the regularity of its loss or retention as between different languages. That is:

- we would have to assume that in a number of languages in which *R > ∅ was regular, *the same lexical items* were borrowed from certain other languages in which *R > *r* was regular; and/or

- we would have to assume that in a number of languages in which *R > *r* was regular, *the same lexical items* were borrowed from certain other languages in which *R > ∅ was regular.

These assumptions are untenable.

I think the only conclusion that can be drawn is that we are dealing here with a true case of irregularity. In Malakula (and many other Vanuatu languages as well), *R—which must have been phonetically similar to *r—began to be lost finally and probably before high vowels. This change, however, was not completed before a second change took place: the merger of *R and *r. Thus *R is lost in some lexical items but retained as the reflex of *r in other items in the same phonological environment. Something similar happened with *q, although here the process of lenition and subsequent loss was further advanced before the remaining reflexes of *q merged with a velar in some environments and a high vowel (via a glide) in others.

Both cases which I have detailed in this paper, then, illustrate the fact that *some* sound changes can be truly irregular.

References

Blust, Robert. 1996. The Neogrammarian hypothesis and pandemic irregularity. In Durie and Ross, eds,135–156.

Campbell, Lyle. 1996. On sound change and challenges to regularity. In Durie and Ross eds, 72–89.

Clark, Ross. To appear. *North and Central Vanuatu: a comparative study*.

Crowley, Terry. 1998. A salvage sketch of Nāti (Southwest Malakula, Vanuatu). In D.T. Tryon, ed. *Papers in Austronesian linguistics No.5*, 101–148. Canberra: Pacific Linguistics.

—— 2006a. *The Avava language of Central Malakula* [ed. by John Lynch]. Canberra: Pacific Linguistics.

—— 2006b. *Naman: a vanishing language of Malakula (Vanuatu)* [ed. by John Lynch]. Canberra: Pacific Linguistics.

—— 2006c. *Nese: a diminishing speech variety of Northwest Malakula* [ed. by John Lynch]. Canberra: Pacific Linguistics.

—— 2006d. *Tape: a declining language of Malakula (Vanuatu)* [ed. by John Lynch]. Canberra: Pacific Linguistics.

Durie, Mark and Malcolm Ross. 1996. *The comparative method reviewed: regularity and irregularity in language change*. New York: Oxford University Press.

Fox, G.J. 1979. *Big Nambas grammar*. Canberra: Pacific Linguistics.

Geraghty, Paul. 1990. Proto-Eastern Oceanic *R and its reflexes. In J.H.C.S. Davidson, ed. *Pacific Island languages: Essays in honour of G.B. Milner*, 51–93. London and Honolulu: School of Oriental and African Studies and University of Hawaii Press.

Koch, Harold. 1996. Reconstruction in morphology. In Durie and Ross (eds), pp. 218–263.

Lynch, John. 2000. Linguistic subgrouping in Vanuatu and New Caledonia. In Bill Palmer and Paul Geraghty, eds *Proceedings of the Second International Conference on Oceanic Linguistics (SICOL),* vol.2: *Historical and descriptive studies*, 155–184. Canberra: Pacific Linguistics.

—— 2004. Bilabial reflexes of Proto-Oceanic *q in Southern Vanuatu languages. *Oceanic Linguistics* 43.1:214–220.

—— 2007. The Malakula linkage of Central Vanuatu. Paper presented to the Seventh International Conference on Oceanic Linguistics, Nouméa, New Caledonia, July 2007.

Lynch, John, and Terry Crowley. 2003. More reflexes of Proto-Oceanic *q in Vanuatu languages. *Oceanic Linguistics* 42.1:233–238.

Lynch, John, Malcolm Ross and Terry Crowley. 2002. *The Oceanic languages*. London: Curzon Press.

Musgrave, Jill. 2001. A grammar of Nefe'ei, Vanuatu. Unpublished MPhil thesis, University of Waikato.

Sperlich, Wolfgang B. 1989. Is the Na-makir (Central Vanuatu) glottal stop a reliable witness for POC *q? In Ray Harlow and Robin Hooper, eds *VICAL 1: Oceanic Languages (Papers from the Fifth International Conference on Austronesian Linguistics)*, 527–553. Auckland: Linguistic Society of New Zealand.

Wolff, John U. 2003. The sounds of Proto Austronesian. In John Lynch, ed. *Issues in Austronesian historical phonology*, 1–14. Canberra: Pacific Linguistics.

7 *In search of an historical Sea-People Malay dialect with* -aba-

WARUNO MAHDI

1 Introduction

One particularity in the historical phonology of Malay (Ml.), that has become relatively well-known after it was explicitly formulated by Adelaar (1992a:74–75; cf. also Nothofer 1975:82), is that the Proto Austronesian (PAn) and lower-level sequence *-aba- is reflected as Ml. *-awa-*, although *b in other positions is regularly retained as Ml. *b*. In Sanskrit (Sk.) loanwords, an *-ava-* of the precursor is regularly rendered as *-awa-* as well.

The sound shift is shared by most other Malayic isolects, including Banjarese, Minangkabau, Serawai, and Jakartanese. Various evidence suggests, however, that a relatively influential language or languages must have served as source of borrowings featuring the sequence *-aba-* in cognates of words having *-awa-* in Malay. Adelaar (ibid.) names two Malayic isolects that retain *-aba-*, Iban and Kendayan, but these are not known to have been particularly influential in historical times.

A particular interest in identifying the *-aba-* source is connected with the transmission of the oldest geopolitical name of the Malayan world, Sanskrit *Yava*[*dvīpa*], besides some occasional other borrowings. Early Chinese and Arabic renderings suggest intermediate precursor forms featuring *-*aba-* for Sanskrit *-ava-*, as also for PAn *-aba-. The particular circumstances of the borrowings, and some historiographic material, suggest that some Sea-People isolects may have been the immediate donors.

Collateral evidence of reflection of initial PAn *q as *k* suggests that one of the donor isolects may have been the precursor of present day Moken. Another isolect involved as donor seems to have undergone preplosion of final nasals.

2 Malay *-awa-* as a regular reflection

It was noted above that Sk. *-ava-* is rendered in Malay Sanskritisms as *-awa-*. This can be demonstrated in the example of the following borrowings:

Alexander Adelaar and Andrew Pawley, eds
Austronesian historical linguistics and culture history: a festschrift for Robert Blust, 73–89.
Pacific Linguistics, 2009.
© This edition vested with Pacific Linguistics.

Sk. *arthavān* > Ml. *hartawan* ‘wealthy person’;
saindhava > Ml. *sendawa* ‘saltpeter’;
vibhava > Ml. *wibawa* ‘authority’
(de Casparis 1997:19, 33, 38).

Intervocalic *v* in Sanskrit is quite generally rendered *w* in Malay (except when the preceding vowel is shifted to schwa), and not only between two *a*-s. For example:

Sk. *devī* > Ml. *déwi* ‘goddess’ (ibid.:16).

In inherited forms, Malay regularly reflects *-aba- in a protoform as *-awa-*. In view of the circumstance that Ml. *-awa-* could also reflect PAn *-awa-, I will cite a Tagalog (Tg.) and/or Toba Batak (Tb.) cognate with *b* to disambiguate the *b of the protoform in the following examples. Consequently, an etymon will be cited for Proto West-Malayo-Polynesian (PWMP), their last common parent language with Malay:[1]

PWMP *baba > Ml. *bawa* ‘carry, bring’ (Tg. *babá*);
*ba-baq > Ml. *bawah* ‘underneath, below’ (Tg. *babâ*);
*laban > Ml. *lawan* ‘adversary, opposite, oppose’ (Tg. *laban*);
*taban > Ml. *tawan* ‘capture’ (Tg. *taban*, Tb. *taban*);
*tabaR > Ml. *tawar* ‘tasteless, fresh [not salty]’ (Tg. *tabag*, Tb. *tabar*)
(cf. Dempwolff 1938:18, 88, 124).

There are a few exceptions in which Malay appears to retain *-aba-*:

PWMP *[c]abaŋ > Ml. *cabang* ‘branch’ (Tg. *sabang*, Tb. *sabang*);
*zabat > Ml. *jabat* ‘hold, grasp’ (Tb. *jabat*)
(cf. Dempwolff 1938:85, 45).

These are perhaps not inherited forms,[2] but borrowings from an as yet unidentified source, made after the sound law underlying the *-aba- > -awa-* shift ceased to be operative. The speech community that spoke the unidentified *aba*-source language or dialect must have either exercised certain influence, or at least played some relevant role in communication and interethnic contacts in the region in (proto)historical times.

3 *Yava ~ Java* as historical geopolitical name and as the name of a cereal

The oldest geopolitical name in the Malayan world, *Yavadvīpa*, is first mentioned in the Sanskrit epic *Rāmāyaṇa* of Valmiki (Kern 1869:640; Mahdi 1994:215 n.93, 2008:111). Being a composite of Sk. *yava* ‘barley’ and *dvīpa* ‘island’, the name could be interpreted literally as ‘barley island’. However, Lassen (1852:1042) already pointed out that barley is not cultivated in the Malayan Archipelago, so that *yava* in this context must have referred to another cereal, foxtail millet or sorghum, or even to cereal in general (Kern 1871:120; Mahdi 1994:431–434 and 469 n.111).

---

1 The protoforms originate in part from Proto Malayo-Polynesian (PMP) or even Proto Austronesian (PAn), but that is irrelevant to this discussion.

2 Remarkably, the reconstructed protoforms involve the protophonemes *c and *z, with regard to which Wolff (1982) suggested that they may have been not authentic at least for PAn.

By the beginning of the current era, Sanskrit was a dead language kept in use artificially as a cult language like Latin in Europe. The contemporaneous 'living' Prakrit (Pk.) reflex *java* 'barley' had initial *j-* in place of *y-* (see Turner 1966:603 #1043).[3] Both shifts, semantic and phonological, are apparent in borrowed cognates in languages of the Archipelago:

Karo-Batak *jaba*, Madurese *jhʌbʌ(h)* 'foxtail millet';
Tb. *jaba-ure*, Balinese *jawʌ*, Ngaju *jawe* 'sorghum';
(Joustra 1907:115; de Clercq 1909: #3113, #3193).

The Toba form is noteworthy because, besides unexpectedly featuring medial *-b-* just like the Karo cognate,[4] it is a compound that has no cognates (of the entire composite) in other Austronesian languages. The compound may have been borrowed as a whole from India, where at least impressionistically cognate-looking compounds, likewise meaning 'sorghum' are reported (Yule and Burnell 1903:465 sub *jowaur*, *jowarree*).

In inherited forms, Toba and Karo lost PMP medial *w, while their medial *b* reflects *b (Adelaar 1981:36). The Toba and Karo forms shown above must have been acquired through a language rendering the medial *v* or *w* of an Indic precursor as *b*.

In the Archipelago, the semantic shift from 'barley' to 'foxtail millet' or 'sorghum' apparently took place in Sanskrit usage as well. The last line of the 7th verse of the 760 CE Dinaya inscription (in Central Java) lists the following items at the local monastery:

yava-yavika-śayyā-cchādanaiḥ '[with] millet, rice, beds, and clothing'
(as interpreted by de Casparis 1941:500–501; Poerbatjaraka 1952:62–63).

Barley at the described monastery would indeed have been a most astonishing archaeological sensation.

With regard to *yava*/*java* as a geopolitical name, the oldest occurrence in Old Malay, in the last line of the 686 CE Kota Kapur inscription (island of Bangka), features the shifted initial of the Prakrit reflex: *bhūmi Jāva*[5] 'the land of *Yava*' (Cœdès 1930:48; Poerbatjaraka 1952:40). The rendering with inital *y-* was preserved in Sanskrit texts, as in line 7 of the 732 CE Canggal inscription (Central Java):

dvīpavaraṁ Yavākhyam,[6] 'an excellent island named *Yava*'
(Poerbatjaraka 1952:52; Sarkar 1959:185).

---

3 I once cited the form as Pali (Mahdi 1994:431), having been misled by the adjectival derivation given as *jāvaka* in the Pali annals of Sri Lanka, the *Cūlavaṁsa* (Kern 1896:240–241), but Turner gives Pali *yava-*, Prakrit *java-*.

4 Madurese regularly has *-b-* < *-w- (Stevens 1966:151; Nothofer 1975:84).

5 The long *ā* in the first syllable indicated place of stress in Old Malay (the Prakrit precursor form *java* has a short *a* in that position). The medial *v* in the Old Malay reflects the spelling, not the actual pronunciation: the script did not allow distinguishing between *b* and *w*, both being written with the Sanskrit character *v* (Vikør 1988:74; Mahdi 2005:186). I will gloss the name as '*Yava*' rather than as 'Java', because it apparently referred not to present-day Java, but to historical *Yava*[*dvīpa*] (Mahdi 2008:119–120).

6 The long *ā* results from fusion at composition junction: *Yava* + *akhyam* '*Yava*-named'.

Three precursor forms corresponding to Chinese and Arabic references to *Yava*

In the first millenium CE, *Yava*[*dvīpa*] was typically referred to with one of the following three precursor forms:

(1) a 'mixed' rendering of Sk. *Yavadvīpa* and its Prakrit reflex *Javadīv*[*a*], i.e. something like **Yavadīv*;[7]

(2) the already mentioned Pk. *Java* < Sk. *Yava*; and

(3) an adjectival derivation of the latter, Pk. *Jāvaka*. < Sk. *yavaka*.

These three alternative forms feature the sequence *-ava-* which is regularly rendered *-awa-* in Malay. Remarkably, early Chinese and Arabic historiographic citations often suggest a *b* instead of the *w* (or *v*) in the immediate precursor form.[8]

Renderings of alternative precursor (1):

Earliest mentions in Chinese sources typically suggest a **Yabadiu* as immediate precursor, i.e. with *b* for medial Sk./Pk. *v* of precursor (1) **Yavadīv*. The Later Han Annals reports the arrival in 132 CE of an embassy from *Yèdiào*.[9] The contemporaneous pronunciation of the latter can be approximated with Pulleyblank's (1991) reconstructed Early Middle Chinese (EMC) *jiap-dɛwʰ which reads as *yap-dɛuʰ in a perhaps more 'customary' transliteration for Austronesianists.[10]

Two centuries later, the itinerary of the Chinese pilgrim Faxian's 413–414 CE voyage from Sri Lanka to Guangzhou writes it as *Yēpótí*,[11] EMC *jia-ba-dɛj (read *ya-ba-dɛi), testifying more overtly to a medial **b* in the immediate precursor form.

Renderings of alternative precursor (2)

The most frequent renderings ultimately derive from precursor (2), Pk. *Java* < Sk. *Yava*. Again, the earliest Chinese renderings point to an intermediate form with **b*. Two variant renderings occur: *Shépó* (EMC *dʑia-ba, read *ja-ba), and, rarely, *Shèpó* (EMC *dʑia'-ba, read *jaʔ-ba). The former is attested in 430 and 435 CE notations in the Liu-Song annals (Ferrand 1916:526; Wolters 1967:36, 151–152), while both appear as doublet variants in the New Tang Annals (Groeneveldt 1877:13 fn.1; Pelliot 1904:279–280; Wolters 1967:216, Mahdi 2008:116 and 127). It is still written *Shépó* in the annals of the 960–1279 CE Song dynasty (Groeneveldt 1877:15 fn.1), and in the *Zhūfānzhì* of Zhao Rugua (Hirth and Rockhill 1911:75), and is only replaced by *Zhǎowā*, a rendering of Malay and/or Old Javanese *Jawa*, in the annals of the 1280–1367 CE Yuan dynasty (Groeneveldt 1877:20 fn.1).

---

7 On 'mixed' Sanskrit-Prakrit renderings of toponyms, resulting from imperfect mastery of Sanskrit by scribes and priests, see Kern (1869:638).

8 Arabic and Chinese have no *v*, but both have a *w*. For Old Greek too, the Ptolemaic renderings *Iabadíou*, *Sabadeĩba*, and *Zábai* (Cœdès 1910:61, 41) imply immediate precursors with *b*, but theoretically one could also ascribe this to an earlier loss of *w* (formerly spelled with *digamma*) in Greek itself.

9 Pelliot (1904:266), Ferrand (1916:521), Stein (1947:136), Mahdi (1995:165–166; 2008:114).

10 Here and further, the 'read as' transliteration is not a correction of Pulleyblank's EMC, but an alternative transcription of the same. Pulleyblank follows IPA convention which, for example, has *j* for palatal glide, for which Austronesianists customarily write *y*.

11 Legge (1886:42–43 of Chinese-script supplement), transliterated in Mahdi (1995:168–169).

In Arabic too, the earliest renderings of alternative precursor (2) indicate an intermediate **Jaba*, rendered as *Ǧāba* in the c. 870 CE *Kitāb al-masālik wa'l-mamālik* ('Book of Roads and Kingdoms') of Ibn-Khordadbeh (Ferrand 1913–14:22 fn.9, 27 fn.7); subsequently also in an 1154 CE work of al-Idrisi; in a work of al-Qazwini who lived 1203–1283 CE; and in a 1339 CE work of Hamdullah Mustawfi (ibid.:185 fn.1; 307 fn.2; and 421 fn.4 respectively).

As in Chinese, renderings with medial *w* in Arabic, *Ǧāwa*, only appear some centuries later than the *b*-cognate. The earliest example noted by Ferrand (1913–14:204 fn.7, read *al-Ǧāwa*) is from the 1224 CE *Mu'ǧam al-buldān* ('Dictionary of Countries') of Yaqut al Hamawi. It persists to this day, and its adjectival derivation *ǧawī* has been borrowed back into Malay as *jawi*, glossed 'Malayan; appertaining to the Malayan peoples and countries' by Wilkinson (1901–03:218).

Renderings of alternative precursor (3)

For alternative precursor (3), Pk. *Jāvaka*, there are several Chinese renderings, all suggesting intermediates with **b*. One rendering of presumably **Jabaka* is found in the 3rd century CE *Nánzhōu yìwù zhì* ('Recount of Curiosities of the Southern Islands') of Wan Zhen as *Shèbó* (Pelliot 1904:277), EMC * dʑia-bak (read *ja-bak). In the report of a 245–250 CE Chinese mission to Cambodia (*Fúnán*) it is rendered *Zhūbó* (Pelliot 1904:270), EMC *tɕia-bak (read *ca-bak). Besides that, there seems to have been a variant version **Jəbaka* which is rendered by Guo Po (who lived in 276–324 CE) as *Qíbó* (Wylie 1897:149), EMC *gji-bak (read *gyi-bak).

The putative intermediate **Jabaka* was apparently borrowed into Arabic as *Zābaǧ*, as indicated by Kern (1885:553 fn.1). But although *ǧīm* in the Arabic-script is read as velar stop *g* in the Egyptian dialect, it is rendered as a palatal affricate *j* in most other Arabic dialects. Hence, Arabic loanwords in Malay typically have *j* for precursor *ǧ*, as noted by Snouck Hurgronje (1906) who therefore questioned Kern's *Zābaǧ – Jāvaka* alignment. Nevertheless, there is an early stratum of apparent mutual borrowing that suggests (possibly indirect) contacts with an Arabic dialect featuring velar reflexion of *ǧ* as *g*. Thus, archaic Malay *jaŋgi ~ jəŋgi* 'black-skinned' is related to the Arabic adjectival derivation *zanǧī* 'id.' of *az-Zanǧ* 'the Horn of Africa, Azania', possibly via Persian *zangī*. Regardless of whether Malay borrowed the form via Persian or directly from Arabic, or whether the Arabic and Malay forms are in fact borrowed from a Persian original[12], this is evidence of an alternative relationship that allows for a Malay velar correspondence to Arabic *ǧ*.

Arabic *Zābaǧ* occurred frequently (Ferrand 1913-1914:v and 701 sub *Djāwaga*; Tibbetts 1979:284 sub *Zābaj*). The earliest citation by Ferrand (1913–14:23 fn.6) is from the already quoted c. 870 CE 'Book of Roads and Kingdoms' of Ibn-Khordadbeh. Considering the doubts expressed by Snouck Hurgronje with regard to the *Zābaǧ – Jāvaka* alignment, one noteworthy passage in the *Kitāb aǧā'ib al-Hind* ('Book of the Marvels of India'), credited to the 10th century sea captain Buzurg ibn Shahriyar, gains particular significance:

> it was the custom among the kings of *Zābaǧ* and [the Land] of Gold that no one was allowed to sit before them, neither indigenous, foreigner nor Muslim, without having his legs crossed, this position being known as *al-barsīlā*.
>
> (Kern 1885:554; Ferrand 1913–14:585; Tibbetts 1979:46–47).

---

[12] Steingass (1892:626, 627) has Persian *Zang* 'Ethiopia', *zangī* 'Egyptian, Ethiop[ian], Moor, etc.'

The word spelled *barsīlā* in Arabic script (Ferrand, ibid. fn.3) is Malay *bersila* (Jawi script *barsīla*, cf. Wilkinson 1901–03:431 sub *sīla*). Although the Sankritism *sila* 'sit with crossed legs' also occurs in other languages, the prefix *ber-* (*bar-* following old spellings) is unique to Malay. Hence, the text unambiguously identifies the language spoken in *Zābağ* as Malay (Mahdi 1995:171).

The alternative name '[(is)land] of Gold' is a calque of Sanskrit *Suvarṇadvīpa* ('gold island')[13]/*Suvarṇabhūmi* ('gold land'), a reference to Sumatra and (perhaps also) the Malayan Peninsula.

For the 7th century, the Persian historians Beladhori and Tabari mention skilled seafarers, the *Sayābiğa*, who formed settlements in the Persian Gulf (de Goeje 1894; 1903:18, 20, 86–91); these are possibly settlers from the Malayan Archipelago. The name was identified by de Goeje as plural form of *Sābağ*, ~ *Sēbağ*, assumed to be cognates of Arabic *Zābağ*.[14] However, it is not a Persian plural form but basically an Arabic one, so the word must have been borrowed in its plural form, and perhaps as rendered by the *Sayābiğa* themselves, because Persian normally retains *z* in direct loans from Arabic.

There does not seem to have been renderings of alternative precursor (3) with *-w-* in either Chinese or Arabic. One reason for this may be, that cognates do not seem to have occurred in (phonologically 'regular') Malay or Javanese either.

4 Identifying characteristics of the mediating language with *-aba-*

It appears that no matter whether the name that was used ultimately originated from Sk. *Yavadvīpa*, *Yava*, or *Yavaka*, Chinese and Arabic renderings originally had *-aba-* for Sk. *-ava-*. It is only in later times, if at all, that doublet forms with *-awa-* appear. The earliest citations of the *b*-modes in Chinese texts predate China's own shipping activities in the Archipelago by several centuries. Similarly, 7th century activity of the *Sayābiğa* cited by Persian historians predates Arabic or Persian sailing to the Archipelago. The precursor-forms with *b* were thus in all likelihood introduced in the Far- and Near-Eastern languages by speakers of the unknown precursor language who must have had a sufficiently extensive area of activity to bridge such a geographical gap.

We can evidently eliminate Indic languages from the list of suspects. The cognate forms in Sanskrit and Prakrit all have *v*. For Sinhalese, Clough (1892:194) gives *jāwā* 'native of Jáva; Malay'. In Tamil, the principal and earliest Dravidian language on the maritime routes, the cognates of the three precursor modes are respectively *Yāvattīvu*, *Yāvam*, and *Cāvakam* (Kern 1869:643, with spelling adaptations).

Politically established speech communities in direct contact with the Malayo-Javanic world seem not to have been influenced by the unknown mediator language, compare Old Khmer *Javā* (Pou 1992:186) and (Old) Cham *Javā* (Aymonier and Cabaton 1906:149). Hence, one cannot even appeal to the fact that the Later Pallava script used in Old Malay did not allow distinguishing between *b* and *w*. If the source of the *b*-modes had been Old Malay, this should also have been reflected in the Khmer rendering. But even in the Sdŏk Kŏk Thom inscription—which refers to a holy ceremony in 802 CE to free Cambodia from

---

[13] That 'islands of *az-Zābağ* are named in India *Suwarn Dīb* meaning islands of gold' is also indicated in al-Biruni's 11th century *Al-Hind* 'India' (Ferrand 1923:1; Tibbetts 1979:50; Mahdi 1995:171).

[14] De Goeje's reconstruction was subsequently supported by Ferrand (1934).

its allegiance to *Javā* (Cœdès and Dupont 1943:108–109) after having been under the latter's immediate overlordship—its name was rendered with *v*.[15]

Meanwhile, there are definite hints in early Indian and Chinese sources, providing a suggestive picture of the seafarers Indians and Chinese had come across in their earliest dealings with the Archipelago. Indian sources refer to *Nāga* communities in a variety of contexts. These sources were written at several points in time in various North and South Indian languages, and they belong to Hinduist as well as Buddhist traditions. They refer in particular to peoples inhabiting the coast or islands in the sea, practicing 'piracy', but also trade, etc. (Mahdi 1999b:169–170, 177–180). Although *Nāga* was a cover term for a variety of ethnic groups, Sea-People (*Orang Laut*) communities operating in the Strait of Malacca and its western approaches must have made up a significant part of them.

Indeed, Solheim (1980:334) suggested on the basis of archaeological data, that seafarers assumed to be Malay-speakers (cf. Mahdi 1994:188–191, 1995:162–165) were already involved in maritime trade with India and the Near East between the 2nd century BCE and the 2nd century CE. More recent archaeological studies indicate that trade with India, the Near East, and the Mediterranean must have existed since the 4th century BCE (Bellina and Glover 2004:73–80), and that large trading ships of a unique Southeast-Asian construction type apparently played a major role (Manguin 2004:283).

Chinese sources possibly allow a concrete identification. In a passage in the 817 CE *Yīqiè jǐngyīn yì* ('Comprehensive pronunciation dictionary') of Hui Lin, which quotes various 3rd century sources, crews of *kūnlún bó* large sea ships from abroad (cf. Christie 1957; Manguin 1980) are said to consist of *Gǔlún* people (Pelliot 1925:257; Needham 1971:459; Mahdi 1999a:163). The pronunciation of the latter designation at different times was reconstructed as EMC *kwət-lwən (read *kuət-luən), LMC[16] *kut-lun; cf. Hakka-Chinese *kwut-lun*, (Giles 1912: #6234 and #7464). A further passage indicates that these *Gǔlún* were 'South-Sea barbarians' and 'entirely black and naked-bodied' (Pelliot 1925:261; Mahdi 1999a:164).

Not only does the description suggest negrito Sea People, but the reference to them as *Gǔlún* (occasionally also *Kǔlún*)[17] is apparently a rendering of a reflex of Proto West-Hesperonesian (PWH)[18] *qulun 'person' (Aceh *ulōn*, Lampung *ulun*, Bintulu *ulun*, Maanyan *ulun*, Malagasy *ólona* 'person'). In some languages, the reflex acquired the meaning 'servant', e.g. Malay, Cham. *hulun*.[19] The word is attested in Old Malay in the expression *hulun-haji* 'the king's vassal subjects' (*haji* 'king'; de Casparis 1956:20, 32;[20] Mahdi 1994:204, 2005:194).

---

15 In the 3rd–4th centuries, the Khmer empire of Funan claimed suzerainty over the Malayan Peninsula, having apparently wrested the latter from former allegiance to *Yava*. This doubtlessly led to 'diplomatic complications' with *Yava*, and that would already have been sufficient opportunity to acquire the *b*-mode, if that had indeed been the form in Old Malay.

16 Late Middle Chinese, likewise after Pulleyblank (1991).

17 Subsequently replaced by the non-cognate sound-alike *Kūnlún* (see Mahdi 1999a:164–165).

18 Cited here as last common protolanguage of Malay, Javanese, Lampung, Cham, Bintulu, Maanyan, and Malagasy (probably conform with Proto West-Indonesian of Nothofer (1975:29), that however does not explicitly include Lampung and Bintulu).

19 Note a similar semantic shift between 'name of Aslian ethnic group' and 'person in subordinated position' in the use of *Sakai ~ sakai* (cf. Wilkinson 1901–03:363, Alwi and Sugono 2001:980) and, less widespread, also of *Semang ~ semang* (cf. Alwi and Sugono 2001:1025).

20 De Casparis glosses the expression 'royal slaves', but there is no explicit context to certify an interpretation as 'slaves'. The latter meaning is only attested much later, in literary Malay of the classical period, but here

It seems likely under these circumstances, that the word for 'person' in the language or dialect(s) spoken by the Sea-People sailors on Malayan ships reaching Chinese ports had been a reflex of *qulun. In languages in which the cognate can mean 'servant' or even 'slave', the word often additionally had the meaning 'I, me (your humble servant)', e.g. Aceh *ulōn ~ lōn*, Cham *hulun*. If the sailors similarly used the word for 'I, me', it would have naturally been adopted in Chinese as reference to them.

Meanwhile, the Chinese rendering *Gǔlún* (< * kuət-luən) suggests that the initial *q of the PWH form was reflected as **k* in the language of those sailors,[21] and this may be significant for identifying their speech community.

5 Pre-Moken as a possible donor language of borrowings with *-aba-*

Assuming that the donor language was spoken by those sailors, we should look for a language having medial *-aba-* corresponding to Malay *-awa-*, and reflecting initial PWH (and PAn) *q- as *k-*. Such languages are indeed known, i.e. Moken (Mo.) and Moklen, spoken by communities of Sea People in the Mergui Islands (on the border between Thailand and Burma). The following examples show *q > *k*;

Mo. *kolon* 'servant, person, man' (Lewis 1960:67, 68; cf. Ml. *hulun* 'servant, etc.'; PWH *qulun) ; *kujan* 'rain' (Ml. *hujan* 'rain'; PWMP *quZan).

The following three examples show *-aba- > *-aba-*:

Mo. *kabang* 'boat, ship' (Lewis 1960:62; cf. Old Ml. *pu-hawang* 'shipmaster', with [*em*]*pu* 'master'; PWMP *qabaŋ, Blust 1973: #307);

baba 'low, short' (Lewis 1960:48; Larish 1999:950 has Mo. *babak*; cf. Ml. *bawah* 'under'; PWMP *ba-baq, Dempwolff 1938:18);

aban 'company, flock' (Lewis 1960:46; cf. Ml. *kawan* 'friend, companion, flock'; *kaban,[22] Dempwolff 1938:71).

A reflex of *-awa- is seen in Moklen [*olan*] *chaba* 'python' (Ml. [*ular*] *sawa* 'id.'), see Larish (1999:538), which reflects PWMP *sawa (Dempwolff 1938:149), cf. Tg. *sawá*, Tb. *sa*, the latter with regular reflection of medial *w as zero (see Adelaar 1981:36).

Some exceptions with *-awa-* are likely to be relatively recent borrowings. For example:

Mo. *bawa* 'bring' (Ml. *bawa* 'carry, bring');
nyawa 'spirit' (Ml. *nyawa* 'soul')
(Lewis 1960:49, 88).

Indeed, Moken and Moklen must have experienced continuous contact with influential Malayic isolects having *-awa-* in historical times.

The Moken and closely related Moklen may seem unlikely candidates for sailors roaming the seas between China and Arabia. But in the early first millennium CE, Sea-

---

too, *hulun* apparently meant more often 'servant', only rarely 'slave' or 'serf'. By contrast, *hamba* seems to have more often meant 'slave, serf' rather than 'servant'.

21 The 'superfluous' final *t of the first syllable must not render some particular feature of the precursor form. A final consonant was frequently added to an originally open syllable in loanwords into Chinese, when the subsequent syllable began with a consonant (cf. Ferrand 1919:265ff.).

22 This reconstruction with *b (Tb. *haban*) assumes that Tg. *kawan* is irregular (perhaps a Malay loan).

People communities speaking an early form of Moken could have played an active role, not through political influence, but as source of manpower for Malay shipping. Regardless of the relationship of Moken and Moklen to either Malayic or Malayo-Chamic,[23] I will provisionally refer to that early first-millennium precursor language as Pre-Moken.

Apart from the Chinese and Arabic words for *Yava* reflecting **-aba-*, there is much additional indirect evidence that there once existed a language showing the same phonological developments as Moken, but spoken in a considerably wider area (Mahdi 1994:203–204 #23):

A reflex of *qulun with initial *k-* (instead of expected *h-*) for *q- was apparently borrowed into Old Javanese (O.Jv.) as part of the compound *pwakulun* ~ *pwangkulun* 'I/me, sir, milord' (< *puaŋ 'master'[24] + *qulun). The form also has a doublet with regular realisation of (*-ŋ+ *q- >) *-ng* + *h-* → *-ngh-* at morpheme junctions in compounds and derivations, i.e. O.Jv. *pwanghulun* 'id.' (Juynboll 1923:393; Mardiwarsito 1978:259). Compare the following examples, chosen for close similarity of the phonological environment:

O.Jv. *karanghulu* 'pillow' ← *karang* 'place' + *hulu* 'head';
panghulu 'leader, chief' ← *paN-* '[ACTOR prefix]' + *hulu* 'head';
sanghulun 'I, me' ← *sang* '[PERSON article]' + *hulun* 'servant, slave'
(Zoetmulder 1982:804, 648, 1667).

Noteworthy is that, although Old Javanese had *hulun* 'servant, slave' (see above example), it does not seem to have had *pwang* 'master',[25] other than as part of this compound. The compound with *pwang* could therefore hardly have been originally coined in Old Javanese. Evidently, the compound was first borrowed as a whole as *pwangkulun*, and only subsequently 'corrected' to the semi-calque *pwanghulun* involving an extant cognate O.Jv. *hulun* 'servant, slave' as second component.

The donor language for O.Jv. *pwangkulun* may also have provided the Toba and Karo renderings with *-aba-* of Pk. *java* 'barley' mentioned above. Similarly, it could also have been the source of Malay forms with *-aba-* for expected *-awa-*, such as *cabang* 'branch' and *jabat* 'hold, grasp'.

With regard to *cabang* 'branch', the initial *c in PWMP *[c]abaŋ is actually uncertain. Based on Tg. *sabang*, Tb. *sabang* 'id.' alone, one would reconstruct *sabaŋ. The initial *c- was based on Jv. *cawang* 'id.' and the Malay reflex that has irregular *-aba-*. For Moken, however, this *-aba-* is regular, and so is *ch* for *s. Compare:

Mo. *cha* 'one' (Ml. *esa* ~ *se-* 'id.');
chochoi 'milk [breast]' (Ml. *susu* 'id.');
pèchang 'banana' (Ml. *pisang* 'id.')
(Lewis 1960:52, 56, 91).

---

[23] See Hogan (1988:3) and particularly Larish (1999:362–415). Although Moken and Moklen 'sound' considerably different to the point of being mutually unintelligible, they share most phonological features (cf. Larish 2005:515–519), so that I will simply treat Moken data as representative of the group as well as of their common ancestral language, citing a Moklen form only when no Moken cognate is recorded.

[24] Blust (1986: #236).

[25] There is only the O.Jv. homonym *pwang* 'and, though', which is not a cognate.

Thus, Ml. *cabang* ‘branch’, having an irregular reflex *-b-*, would become perfectly regular if it were a borrowing from (Pre-)Moken (cf. Ko’-Surin Moken *cabaŋ* cited in Larish 1999:626), even if the protoform had initial *s. That leaves us with Jv. *cawang* that could however be an early loan (before the *b* > *w* shift in Javanese) either directly from Pre-Moken, or via Malay.

In the early 1st millennium CE, influence of Pre-Moken on the Martaban coast must have been substantial enough for (Pre-)Mo. *kabang* ‘boat, ship’ (see above) to be borrowed into Old Mon as *kbaŋ* ‘ship’ (Shorto 1971:67).

There are Old Mon reports about raids from the sea by *rakṣasa* ‘cannibal demons’[26] on the Martaban coast (Luce 1965:145–146; Forbes 1878:234). These *rakṣasa*-s, termed ‘Malayan Vikings’ by Luce who regarded them as ancestors of the Moken, were known in Myanmar tradition as *Bilù*—glossed ‘a kind of monster which eats human flesh and possesses super-human eyes’ by Stevenson and Eveleth (1921:727)—see Mahdi (1994:204). It is difficult to ascertain whether these represented the Pre-Moken speech community, but considering that Bataks of the Sumatran highlands used to be characterised in a similar fashion (as alleged ‘cannibals’), the Old Mon and Myanmar descriptions were arguably also skewed.

6 Another early Sea-People Malayic isolect featuring *-aba-*

Apart from Pre-Moken, there evidently were other isolects spoken on the sailing routes in the past, and at least one of these must also have reflected *-aba- as *-aba-*. This is apparent from borrowed cognates of the aforementioned protoform *qabaŋ ‘boat, ship’, i.e. Chinese *bó* ‘sea-going ship’ (EMC *baɬjk / bɛɬjk) along with Favorlang *abak* ‘boat’.[27] Though both indicate that the donor language must have had *-aba-*, they do not feature the initial *k* < *q expected in a loan from (Pre-)Moken, compare Mo. *kabang* ‘boat, ship’.

Another contrast with (Pre-)Moken is the rendering of the final velar nasal as stop. The language that transported the word to Guangzhou and Taiwan had evidently undergone nasal preplosion (Court 1967; Adelaar 1992b:386–387), often leading to the appearance of a voiceless stop before final nasals (*-m* > *-pm*; *-n* > *-tn*; *-ŋ* > *-kŋ*). Sometimes, the secondary prenasal stop is voiced, as in Bonggi (Banggi island, Sabah) cf. Blust (1997:155–156) who provides an overview of the nasal preplosion phenomenon, and in Orang-Darat (Od.), e.g. Od. *ayabm* ‘chicken’ (Ml. *ayam*), which in some environments inserts an additional anaptyctic schwa, e.g. Od. *dosudən* ‘village’ (Ml. *dusun*) (Kähler 1960:33–38).

Important for the present study is a further development, in which the nasal-preplosion cluster loses the nasal, leaving only the unvoiced stop, for example *-kŋ* > *-k* as in Mentawai *abak* ‘boat’ (< *qabaŋ, Mahdi 1994:476 #145). It is this further development of nasal preplosion that apparently led to Chinese *bó* ‘sea-going ship’ and Favorlang *abak* ‘boat’. There must have been a Sea-People isolect featuring this sound shift and at the same time having *-aba-* in place of standard Malay *-awa-*, but without reflecting initial *q as *k*.

The final nasal-to-stop shift is also recorded for Urak Lawoi’ (Ul.), formerly often confused with Moken (Hogan 1988:2). Compare:

---

[26] A borrowing from Sanskrit *rakṣasa* ‘id.’

[27] Also borrowed, because in Favorlang inherited forms, the expected reflex of final *-ŋ is *-n* (Marsh 1977: #4.3.1.2.9).

Ul. *urak* ‘man, person’ (Ml. *orang* < PMP *uRaŋ);
kawat ‘friend, group of’ (Ml. *kawan* < PWMP *kaban)
(Hogan 1988:159, 125; Dempwolff 1938:160, 71).

But as the latter example demonstrates, Urak Lawoi’ reflects *-aba- as *-awa-* and thus cannot be the unknown donor language. Compare also:

Ul. *lawat* ‘endure, oppose, fight against’ (Ml. *lawan* < PWMP *laban);
tawal ‘tasteless, fresh [not salty]’ (Ml. *tawar* < PWMP *tabaR)
(Hogan 1988:131, 154; Dempwolff 1938:88, 124).

The combination of the final nasal-to-stop shift and retention of *-aba- as *-aba-* is reported for Belangin; that is however spoken far from the sea, inland from Kendayan with which it is closely related (Adelaar 2006:68, 70). Compare:

Bl. *tulak* ‘bone’ (Ml. *tulang* < PWMP *tuqelaŋ);
pañjak ‘long [shape]’ (Ml. *panjang* < PWH *pañjaŋ),

but

babàh ‘under’ (Ml. *bawah* < PWMP *ba-baq).

It seems possible, however, that a meanwhile extinct Sea-People dialect of Malay may have shared this combination of sound correspondences, so as to serve as donor for Chinese *bó* and Favorlang *abak*. Alternatively, it is equally possible that that donor language still featured the original nasal-preplosion cluster, and that the reduction to final stop took place upon borrowing into Chinese and Favorlang (i.e. donor *-kŋ* > Early Chinese *-k*).

7 Distribution of Sea-People isolects reflecting *-aba- as *-aba-* and as *-awa-*

On the sea routes there must have also been other seafarers who spoke more ‘regular’ Malay dialects reflecting *-aba- as *-awa-*. Specifically, this reflex seems to have prevailed on the routes to the Moluccas and Nusa Tenggara, transporting cloves, nutmeg, and sandalwood. This follows in the first place from the borrowing of Malay [*bunga*]*lawang*, which nowadays means ‘mace [of nutmeg]’ but originally meant ‘clove’ (< [*buŋa ‘flower’] + *labaŋ ‘nail’), and was borrowed into Sanskrit as *lavaṅga* ‘clove’ (Mahdi 1994:188, 215 #92). It is already attested in Valmiki’s *Rāmāyaṇa* (Gonda 1932:326–329), suggesting a date before the 1st century BCE, or, allowing for early interpolations that are difficult to distinguish from the authentic text, not later than the 2nd century CE.[28]

East Central Maluku cognates likewise feature an *-awa-* but render the intervocalic velar nasal in the first component (Ml. *bunga* ‘flower’) as *-k-*, which is irregular and indicates that these cognates must be borrowed (Mahdi 1994:189), for example:

> Asilulu *pukalawa-e*, Haruku *pokolawan*, Nusalaut *pekalawan*, Piru *pokelawan*, etc. ‘clove’.

[28] In China, cloves are mentioned in the annals of the Earlier (Western) Han Dynasty (206 BCE–24 CE) (Burkill 1935:961, Mahdi 1994:189–190), confirming the early date of clove transportation from North Maluku.

Final *-n* for *-ŋ is regular. The irregular rendering in medial position, sugesting borrowing, is rather rare, but it is attested in an East Central Maluku set of cognates for 'cat', being borrowings of Malay *singa* 'tiger' (from Sanskrit). This sound correspondence applying to borrowings was apparently operative for a relatively short period at the time of the earliest acquisition of Sanskritisms (Mahdi 1994:189).

Funan's hegemony over the Gulf of Thailand (3rd–4th centuries CE) may have caused a further separation of the groups. Sea-People communities around the Peninsula and the Kra Isthmus, which were under immediate Funan suzerainty, could continue sailing through the South China Sea, while other communities that did not submit to Funan had to take a roundabout route to China through the Java Sea, the Strait of Macassar, and the Philippines (cf. Mahdi 1994:185–188). This seems to follow from the borrowed cognates of Pk. *java*, apparently via Ml. *jawa*, in languages of the Philippines, featuring *-awa-* rather than *-aba-*:

> Maranao *daoa?* 'barley (actually 'millet'?[29])', Tausug, Cebuano Bisaya, Tagalog *dawa* '[foxtail] millet', Aklanon *dawah*;
>
> also:
>
> Isneg *dáwa* 'ear, head, spike (of grain)', Ilokano *dawa* 'ear of grain (esp. of rice)', Itbayaten *um-rawah* 'to appear out of the ears when nearly ripe (of grain)', Pangasinan *dawá* 'rice grain'. (Mahdi 1994:433)

8 Conclusions

In the light of the above, languages or dialects spoken at the beginning of the current era by the Sea-People crews of high-sea merchant ships may have been the immediate source of geopolitical terms in Chinese and Arabic, in which a medial *v* of the Sanskrit-Prakrit ultimate source language was rendered as *b*. It likewise became evident, that more than one isolect must have been involved, including one reflecting initial *q as *k*, and another one which did not do so, but which apparently shifted a final *-ŋ to *-k*.

The identification of Pre-Moken, the ancestral language of present day Moken and Moklen, as the former of the two source isolects, seems particularly fruitful in that it provides possible explanations for other seemingly irregular sound correspondences in the region, particularly the *k* in O.Jv. *pwangkulun*, the *-aba-* for Indic *-ava-* in Toba and Karo Batak words for sorghum, and perhaps also the *c* in Ml. *cabang*. Jv. *cawang*.

In the context of the apparent retention of *-awa-* in the transmission of words for 'clove' and 'grain, millet' within insular South East Asia as far as the Moluccas and the Philippines, the main areas of activity of speakers of dialects featuring *-aba-* must have been more specifically restricted to sailing routes closer to the Asian mainland.

More study is needed to substantiate these further considerations. It seems likely, however, that Sea-People indeed played an important role in early Southeast-Asian maritime navigation and trade with China, India and the Near East, and the languages or dialects they spoke were probably responsible for the *-aba-* modes of the geopolitical names discussed above.

---

[29] McKaughan and Macaraya (1967) gloss it 'barley', but as the cereal does not grow in the Archipelago, the authors apparently meant 'millet'.

References

Adelaar, K. Alexander. 1981. Reconstruction of Proto-Batak phonology. In Robert A. Blust, ed. *Historical linguistics in Indonesia*, Part I, *NUSA*: Linguistic Studies in Indonesian and Languages in Indonesia 10, 1–20. Jakarta: Badan Penyelenggara Seri NUSA.

—— 1992a. *Proto-Malay, the reconstruction of its phonology and parts of its lexicon and morphology*. Canberra: Pacific Linguistics.

—— 1992b. The relevance of Salako for Proto-Malayic and for Old Malay epigraphy, *Bijdragen tot de Taal-, Land- en Volkenkunde* 148. 381–408.

—— 2006. Where does Belangin belong? In Fritz Schulze and Holger Warnk, eds *Insular Southeast Asia*: *linguistic and cultural studies in honour of Bernd Nothofer*, 65–84. Wiesbaden: Harrassowitz.

Alwi, Hasan and Dendy Sugono, eds, 2001. *Kamus Besar Bahasa Indonesia*, edisi ke-3, Balai Pustaka no. 3658. Jakarta: Balai Pustaka.

Aymonier, Étienne and Antoine Cabaton. 1906. *Dictionaire čam – français*, Publications de l'École Française d'Extrême-Orient 7. Paris: Ernest Leroux.

Bellina, Bérénice and Ian Glover. 2004. The archaeology of early contact with India and the Mediterranean World, from the fourth century BC to the fourth century AD. In Ian Glover and Peter Bellwood, eds *Southeast Asia*: *from prehistory to history*, 68–88. London: Routledge Curzon.

Blust, Robert A. 1973. Additions to 'Proto-Austronesian addenda' and 'Proto-Oceanic addenda with cognates in non-Oceanic Austronesian languages'—II, *University of Hawaii Working Papers in Linguistics* 5/3:33–61.

—— 1986. Austronesian etymologies—III, *Oceanic Linguistics* 25:1–123.

—— 1997. Nasals and nasalization in Borneo, *Oceanic Linguistics* 36:149–179.

Burkill, I.H. 1935. *A dictionary of the economic products of the Malay Peninsula*. London: Crown Agents for the Colonies.

Casparis, J.G. de. 1941. Nogmaals de Sanskrit-inscriptie op den steen van Dinojo, *Tijdschrift voor Indische Taal-, Land- en Volkenkunde* 81, 499–513.

—— 1997. *Sanskrit loan-words in Indonesian*: *an annotated check-list of words from Sanskrit in Indonesian and Traditional Malay*, NUSA Linguistic Studies of Indonesian and Other Languages in Indonesia 41. Jakarta: Badan Penyelenggara Seri NUSA.

Christie, Anthony. 1957. An obscure passage from the Periplus: Κολανδιοφωντα τα μεγιστα, *Bulletin of the School of Oriental and African Studies* 19:345–353.

Clough, B. 1892. *A Sinhalese–English dictionary*, new and enlarged edition. Colombo: Wesleyen Mission Press.

Çœdès, George. 1910. *Textes d'auteurs grecs et latins relatifs à l'Extrême-Orient depuis le IVe siècle av. J.-C. jusqu'au XIVe siècle*. Paris: Ernest Leroux.

—— 1930. Les inscriptions malaises de Çrīvijaya. *Bulletin de l'École Française d'Extrême-Orient* 30:29–80.

Çœdès, George and Pierre Dupont. 1943. Les stèles de Sdŏk Kăk Thom, Phnom Sandak et Práḥ Vihār, *Bulletin de l'École Française d'Extrême-Orient* 43:56–154.

Court, Christopher. 1967. Some areal features of Mentu Land Dayak, *Oceanic Linguistics* 6:46–50.

Dempwolff, Otto. 1938. *Vergleichende Lautlehre des austronesischen Wortschatzes*, Bd. 3, *Austronesisches Wörterverzeichnis*, Beihefte zur Zeitschrift für Eingeborenen-Sprachen 19. Berlin: Dietrich Reimer.

Ferrand, Gabriel. 1913–14. *Relations de voyages et textes géographiques arabes, persans et turks relatifs a l'Extrême-Orient du VIII^e^ au XVII^e^ siècles*. Paris: Ernest Leroux.

—— 1916. Ye-tiao, Sseu-tiao et Java, *Journal asiatique* (11e série) 8:521–532.

—— 1923. Notes de géographie orientale, *Journal asiatique* 202:1–35.

—— 1934. Sayābidja. In M.Th. Houtsma et al., eds *The encyclopaedia of Islām*, vol. 4, 200–202. Leiden: E.J. Brill.

Forbes, C.J.F.S. 1878. On the connexions of the Mons of Pegu with the Koles of Central India, *Journal of the Royal Asiatic Society of Great Britain and Ireland* new series 10:234–243.

Giles, Herbert A. 1912. *A Chinese–English dictionary*, second ed. Shanghai: Kelly and Walsh.

Goeje, M.J. de. 1894. De Sajâbidja. In *Feestbundel van taal-, letter-, geschied-, en aardrijkskundige bijdragen* [...] *aan Dr. P.J. Veth* [...] *aangeboden*, 10–12. Leiden: E.J. Brill.

—— 1903. *Mémoire sur les migrations des Tsiganes à traverse l'Asie*. Leiden: E.J. Brill.

Gonda, J. 1932. Etymologica, *Acta Orientalia* 10:326–335.

Groeneveldt, W.P. 1877. *Notes on the Malay Archipelago and Malacca compiled from Chinese sources*, Verhandelingen van het Bataviaasch Genootschap van Kunsten en Wetenschappen 39/1/1. Batavia: W. Bruining.

Hirth, Friedrich, and W.W. Rockhill. 1911. *Chau Ju-Kua: his work on the Chinese and Arab trade in the twelfth and thirteenth centuries, entitled Chu-fan-chi*. St. Petersburg: Imperial Academy of Sciences.

Hogan, David W. in collab. with Stephen W. Pattemore. 1988. *Urak Lawoi'*: *Basic structures and a dictionary*. Canberra: Pacific Linguistics.

Joustra, M. 1907. *Karo-Bataksch woordenboek*. Leiden: E.J. Brill.

Juynboll, H.H. 1923. *Oudjavaansch-Nederlandsch woordenlijst*. Leiden: E.J. Brill.

Kähler, Hans. 1960. *Ethnographische und linguistische Studien über die Orang darat, Orang akit, Orang laut und Orang utan im Riau -Archipel und auf den Inseln an der Ostküste von Sumatra*, Veröffentlichungen des Seminars für Indonesische und Südseesprachen der Universität Hamburg 2. Berlin: Reimer Verlag.

Kern, Hendrik. 1869. Java en het goudeiland volgens de oudste berichten. *Bijdragen tot de Taal-, Land- en Volkenkunde van Nederlandsch-Indië.* 3e volgreeks 4:638–648.

—— 1871. De naamsoorsprong van Java. *Bijdragen tot de Taal-, Land- en Volkenkunde van Nederlandsch-Indië.* 3e volgreeks 6:116–120.

—— 1885. Geschied- en oudheidkundige nasporingen in Britsch Burma. *Bijdragen tot de Taal-, Land- en Volkenkunde van Nederlandsch-Indië.* 4e volgreeks 10:532–556.

—— 1896. Twee krijgstochten uit den Indischen Archipel tegen Ceilon. *Bijdragen tot de Taal-, Land- en Volkenkunde van Nederlandsch-Indië.* 46:240–245.

Larish, Michael David. 1999. The position of Moken and Moklen within the Austronesian language family. University of Hawaii. PhD dissertation. Ann Arbor: University Microfilms International #99–25, 288.

—— 2005. Moken and Moklen. In Alexander Adelaar and Nikolaus P. Himmelmann, eds *The Austronesian languages of Asia and Madagascar*, 513–533. London: Routledge.

Lassen, Christian. 1852. *Indische Alterthumskunde*. Bd. 2. *Geschichte von Buddha bis auf die Ballabhi- und jüngere Gupta-Dynastie*. Bonn: Koenig.

Legge, James. 1886. *A record of Buddhist Kingdoms, being an account by the Chinese Monk Fa-Hien of travels in India and Ceylon (A.D. 399–414) in Search of the Buddhist Books of Discipline.* Oxford: Clarendon Press.

Lewis, M. Blanche. 1960. *Moken texts and word-list.* A Provisional Interpretation. London: Luzac and Co.

Luce, Gordon H. 1965. Rice and religion, a study of Old Mon-Khmer evolution and culture. *Journal of the Siam Society* 53:139–152.

McKaughan, Howard P. and Batua A. Macaraya. 1967. *A Maranao dictionary*. Honolulu: University of Hawai‘i Press.

Mahdi, W. 1994. Some Austronesian maverick protoforms with culture-historical implications. *Oceanic Linguistics* 33:167–229, 431–490.

—— 1995. Wie hießen die Malaien, bevor sie “Malaien” hießen? In A. Bormann, A. Graf, M. Meyer, M. Voss, eds *Südostasien und wir*. Austronesiana: Studien zum austronesischen Südostasien und Ozeanien 1, 162–176. Hamburg: Lit.

—— 1999a. The dispersal of Austronesian boat forms in the Indian Ocean. In Roger Blench and Matthew Spriggs, eds *Archaeology and language III*: Artefacts, *Languages and texts*, 144–179. London: Routledge.

—— 1999b. Linguistic and philological data towards a chronology of Austronesian activity in India and Sri Lanka. In Roger Blench and Matthew Spriggs, eds *Archaeology and language IV*: *language change and cultural transformation*, 160–242. London: Routledge.

—— 2005. Old Malay. In Alexander Adelaar and Nikolaus P. Himmelmann, eds *The Austronesian languages of Asia and Madagascar*, 182–201. London: Routledge.

—— 2008. Yavadvipa and the Merapi Volcano in West Sumatra. *Archipel* 75:111–143

Manguin, Pierre-Yves. 1980. The Southeast Asian ship: an historical approach. *Journal of Southeast Asian Studies* 11:266–276.

—— 2004. The archaeology of early maritime polities of Southeast Asia. In Ian Glover and Peter Bellwood, eds *Southeast Asia*: *from prehistory to history*, 282–313. London: Routledge Curzon.

Mardiwarsito, L. 1978. *Kamus Jawa Kuna (Kawi) – Indonesia*. Ende: Nusa Indah.

Marsh, Mikell Alan. 1977. Favorlang-Pazeh-Saisiat: a putative Formosan subgroup. Washington State University. Dissertation. Ann Arbor: University Microfilms International #78–314.

Needham, Joseph. 1971. *Science and civilisation in China*, with the collaboration of Wang Ling and Lu Gwei-djen, vol. 4, *Physics and physical technology*, Part III, *Civil engineering and nautics*. Cambridge: Cambridge University Press.

Nothofer, Bernd. 1975. *The reconstruction of Proto-Malayo-Javanic*. Verhandelingen van het Koninklijk Instituut voor Taal-, Land- en Volkenkunde 73. 's-Gravenhage: Martinus Nijjhoff.

Pelliot, Paul. 1904. Deux itinéraires de Chine en Inde à la fin du VIII^{e} siècle. *Bulletin de l'École Française d'Extrême-Orient* 4:131–413.

—— 1925. Quelques textes chinois concernant l'Indochine Hindouisée. In *Études asiatiques, publiées à l'occasion du vingt-cinquième anniversaire de l'École Française d'Extrême-Orient*, tôme 2, Publications de l'École Française d'Extrême-Orient 20, 243–263. Hanoi: École Française d'Extrême-Orient.

Poerbatjaraka. 1952. *Riwajat Indonesia*, djilid 1. Djakarta: Jajasan Pembangunan.

Pou, Saveros. 1992. *Dictionaire vieux khmer–français–anglais* / *An Old Khmer–French–English dictionary*. Paris: Centre de Documentation et de Recherche sur la Civilisation Khmère.

Pulleyblank, Edwin G. 1991. *Lexicon of reconstructed pronunciation in Early Middle Chinese, Late Middle Chinese, and Early Mandarin*. Vancouver: University of British Columbia Press.

Sarkar, Himansu Bhusan. 1959. The Stone Inscription of King Sañjaya (Canggal Inscription), 654 Śaka. *Journal of the Asiatic Society* (Calcutta) 1:183–188.

Shorto, H.L. 1971. *A dictionary of the Mon inscriptions from the Sixth to the Sixteenth Centuries*. London Oriental Series 24. London: Oxford University Press.

Snouck Hurgronje, C. 1906. Een miskende Arabische palataal. *Tijdschrift voor Indische Taal-, Land- en Volkenkunde* 48:85-88.

Solheim, Wilhelm G. II. 1980. Neue Befunde zur späten Prähistorie Südostasiens und ihre Interpretation. *Saeculum* 31:275–317, 319–344.

Stein, R.A. 1947. *Le Lin-Yi*: *Sa localisation, sa contribution à la formation du Champa et ses liens avec la Chine*. Han-Hiue, Bulletin du Centre d'Études Sinologiques de Pékin vol. 2, fasc. 1–3. Pékin: Université de Paris, Centre d'Études Sinologiques de Pékin.

Steingass, Francis Joseph. 1892. *A comprehensive Persian-English dictionary*. London: Routledge and Keegan Paul.

Stevens, Alan M. 1966. The Madurese reflexes of Proto-Malayopolynesian. *Journal of the American Oriental Society* 86:147–156.

Stevenson, Robert C. and F.H. Eveleth, eds. 1921. *The Judson Burmese–English dictionary*. Rangoon: American Baptist Mission Press.

Tibbetts, G.R. 1979. *A study of the Arabic texts containing material on South-East Asia*. Leiden and London: E.J. Brill.

Turner, R.L. 1966. *A comparative dictionary of the Indo-Aryan languages*. London: Oxford University Press.

Vikør, Lars S. 1988. *Perfecting spelling. Spelling discussions in Indonesia and Malaysia 1900–1972*. Verhandelingen van het Koninklijk Instituut voor Taal-, Land- en Volkenkunde 133. Dordrecht: Foris Publications.

Wilkinson, R.J. 1901–1903. *A Malay–English dictionary*. Singapore: Kelly and Walsh.

Wolff, John U. 1982. Proto-Austronesian *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, 1–30. Canberra: Pacific Linguistics.

Wylie, Alexander. 1897. Asbestos in China. In A. Wylie, *Chinese Researches*: III. *Scientific*, 141–154. Shanghai: [s.n.].

Yule, Henry, and A.C. Burnell. 1903. *Hobson-Jobson*, new ed. edited by William Crooke. London: Murray.

Zoetmulder, P.J. in collaboration with S.O. Robson. 1982. *Old Javanese–English dictionary*. ’s-Gravenhage: Martinus Nijhoff.

8 *The sounds of Southeast Babar*

HEIN STEINHAUER

1 The Babar archipelago and its languages

The Babar archipelago in east Indonesia consists of Babar Island and the smaller islands of Wetan to its west, Dai, Dawera, Dawelor[1] to its north, and Marsela to its southeast. It is administratively a *kecamatan* (district) in the *Maluku Tenggara Barat* regency (Western Southeast Moluccas), which is part of Maluku province.

Recent anthropological research (Dijk 2000) confirmed that the islands from Kisar and Leti at the eastern tip of Timor up to the Babar archipelago form a regional 'field of anthropological study', a concept defined by J.P.B. de Josselin de Jong in the 1930s as a limited cultural region with a population whose culture appears to be sufficiently homogeneous and distinct to form in its totality a separate object of ethnological research, while at the same time it is sufficiently diverse as to be a promising topic for comparative research (cf. Josselin de Jong 1935). Common cultural characteristics are the status of Luang island as a cultural centre and the existence of moieties (boat-owners and land-owners, Engelenhoven and Hajek 2000:113–114), a traditional belief system manifested in yearly rituals from Leti all the way to Marsela, and the existence of a common 'sung language' (Engelenhoven 1996).

However, the region shows considerable linguistic variety, being one of the areas where Austronesian and non-Austronesian languages meet. Until recently the Babar archipelago and the islands thereabout were linguistically more or less terra incognita. Attention to the languages there was restricted to line-drawing: lines separating languages on a map and lines connecting languages in a tree diagram. While the lines tended to move somewhat over time, data on the languages themselves—if presented at all—were never more than ancillary in this approach.

[1] Also spelled Daweloor and Davelor in the literature.

Alexander Adelaar and Andrew Pawley, eds
Austronesian historical linguistics and culture history: a festschrift for Robert Blust, 91–105.
Pacific Linguistics, 2009.
© This edition vested with Pacific Linguistics.

The most detailed language map of the area is Map 40 in Wurm and Hattori (1982). In contrast to earlier maps it is based on first hand information provided by Nico de Jonge and Toos van Dijk, who at the time did anthropological fieldwork in Marsela and Dawelor. Discussing this map and earlier ones, Taber (1993:390–391) questions the validity of the boundaries they indicate. However, the major problem with maps like those of Wurm and Hattori and their predecessors is that they give a skewed picture of the linguistic reality on the ground, suggesting precision, stability and homogeneity, where in fact multilingualism, generational differences, sociolectal stratification and language shift, are the rule.

The first one to systematically collect data on the Austronesian languages of the Moluccas, including the Southwestern islands, was the Russian anthropologist M.A. Členov (1976; see also Steinhauer 1980). The large scope of his survey caused him to use a wide-mesh grid, with the result that Babar was represented by only one list, creating the impression that the island was totally Wetan speaking, with the single exception of Letwurung, a village on the east coast. Basing himself on a probably mistaken observation by Riedel (cf. Riedel 1886:334), he asserts that the language there was the same as the one spoken on Dawelor.

The most detailed information to date on these languages are Taber's (1993) 'raw phonetic' wordlists from 24 southwestern Moluccan languages distinguished on the basis of lexicostatistical calculations. These calculations were made on wordlists of 193 items from forty-one villages, twelve of which are located in the Babar archipelago. Ten Babar lists are published in an appendix (Taber 1993:411–435). They represent the ten languages or dialect groups, which, as far as we know, are not found outside the Babar archipelago: Dai (on the island of Dai), Dawera-Dawelor (on the islands Dawera and Dawelor), Serili, East Masela, and Central Masela (on the island of Marsela), and North Babar, Southeast Babar, Emplawas, Tela-Masbuar, and Imroing (all on Babar Island proper).

The location and classification of the languages of the Babar archipelago by Taber (Taber 1993 and 1996) differ in no insignificant detail from the ones presented in Wurm and Hattori (1983, Map 40). Table 1 illustrates the classificatory differences between these sources (names of languages are in italics, language clusters in bold, dialects in regular type). As the table indicates, Taber identifies a separate North Babar language, 'spoken by nearly 1500 people in six villages on the north side of Babar Island' (Taber 1993:407), which in Taber (1996) has become: 'There are over 1400 speakers living in five villages.'[2] Wurm and Hattori lack such a language, but picture the north coast as Wetan speaking with a Dai and a Dawera-Daweloor enclave. According to Taber's lexicostatistical calculations, on the south coast there are three languages, spoken respectively in Emplawas, Tela [tə^{l}la] (with its daughter village Masbuar), and Imroing. Wurm and Hattori do not mention Imroing and consider Emplawas and 'Tela-Masboar' as dialects of their Marsela-South Babar dialect group. Both classifications are impressive in their endeavour to be precise, but the insights they present are limited at best.

[2] 'terdapat 1.400 lebih penutur yang menempati lima desa' (Taber 1996:120).

Table 1: Languages and language groups according to Wurm and Hattori (1982) and Taber (1993 and 1996).

Wurm and Hattori (1982)	**Taber (1993**:396, 407 and 409; **1996**:111, 117–121**)**
Timor and Islands Subgroup	**Southwest Maluku Group**
Letri-Lgona: Wetan-West (North) Babar (the whole of Babar except the Dai and Dawera-Daweloor pockets and the southeast part of the island)	*Luang:* Wetan (on the island of Wetan and around Tepa)[3]
–	**Babar group: North Babar subgroup**
	North Babar
Dai (on Dai and in an enclave on Babar's north coast)	*Dai* (on Dai only)
Dawera-Daweloor (on Dawera and Dawelor and in a northeast Babar pocket)	*Dawera-Daweloor* (on Dawera and Dawelor only)
Masela-Southeast Babar cluster	**Babar group: South Babar subgroup**
Masela-South Babar dialect group	**Southwest Babar cluster**
–	*Imroing*
Emplawas	*Emplawas*
Tela-Masboar	*Tela-Masbuar*
	Masela-Southeast Babar cluster[4]
West Masela	*West Masela*
East Masela	*East Masela*
Central Masela	*Central Masela*
Serili	*Serili*
Southeast Babar	*Southeast Babar*

As Table 1 shows, there is agreement on the existence of a Southeast Babar language. According to my informants, this language is spoken with minor variations in lexicon and intonation along the southeast Babar coast in the villages of Koroing, Letwurung,[5] Kokwari, Wakpapapi, Asnari, Analutur, and Manuweri. North of Koroing and along the north coast, another language (Taber's North Babar) is spoken in the old villages of

[3] Luang is not listed in the tables in Taber (1993).

[4] The break-up of this cluster presented here is from Taber (1996). In Taber (1993) the cluster is called Masela-South Babar with the languages: Southeast Babar, Serili, East Masela, and Central Masela (p.396 and 407); on p.409 the latter language is simply called Masela without further geographic qualification. West Masela is mentioned nowhere.

[5] Wurm and Hattori mark Letwurung (one of the four Babar villages indicated on the inset for the Babar archipelago on their map) as Masela-South Babar, and as belonging to an unnamed dialect different from Southeast Babar and the other dialects listed in Table 1 above.

Yatoke, Nakramto (Taber's Nakarahamto), and Ilyara. West of Ilyara, along the northwestern tip of the island, varieties of Wetan are spoken in the villages of Yautubung, Manuwui, Watrupun, Letsyara, Wetota, and Tepa. My informants identified Tela-Masbuar in the southwest as a different language again, but they were not sure about the status of Imroing (between Tepa and Tela), nor of Emplawas (at the southern tip of the island), vis-à-vis Tela-Masbuar. But they both agreed that the inhabitants of Tutuwawang between Emplawas and Manuweri 'did not have a language' at all, meaning that they speak a variety of Malay. Whether they are original inhabitants who shifted to Malay or immigrants, or both, remains unclear. On the oldest map of the island (Hoëvell 1890:198–199), Tutuwawang[6] is marked with an anchor, suggesting that in the late 19th century this village had most contact with the outside world. Today Letwurung is the main anchoring place on the east coast.

The Babar archipelago has long been isolated. In Van Hoëvell's time all villages with the exception of Tepa were still walled and relatively inaccessible. In the last two decades of colonial rule this situation must have changed. According to the *Encyclopaedia van Nederlandsch-Indië* (Encyclopaedia 1917) the inhabitants of Babar still 'worshipped their ancestors and the sun'. But the supplement to this *Encyclopaedia* (1939 vol.VIII) reports that the whole population had been converted to Christianity. In modern Indonesia the outside world has become part of everyday life. Linguistically this means that multilingualism has become general (involving indigenous languages, regional Malay, and increasingly also standard Indonesian). In many villages children are shifting to regional Malay, threatening the viability of the indigenous languages. In the not too distant future, Taber's wordlists may become all that will ever be known about several of these languages. The following notes are a modest addition to what can be deduced from Taber's data[7] regarding one of the Babar languages, Southeast Babar.[8]

2 Southeast Babar: phonemic and morphological observations

2.1 Data

The observations below are based on the dialect of Letwurung (or [lɛχˈwyɔrl]), the mother tongue of Mrs Fien Laun. She was born in 1958 and raised in the village of Letwurung in a Southeast Babar speaking family. In 1983 she moved to the Netherlands, for good. In the Netherlands she rarely had the opportunity to speak her mother tongue, and her home language became Moluccan Malay, on occasions mixed with Dutch. In the last few years she has become the only speaker of her language in the Netherlands, with no one left to speak to. The raw data of her language were collected in weekly elicitation sessions as part of a field training course at the University of Nijmegen in 2005, and in a number of sessions at Fien Laun's home in the city of Groningen. On two of these latter occasions she had a Malay speaking friend of Chinese descent with her who was also born

---

6 Spelled as *Toetoewawang* at the time. Most of the villages mentioned above already existed in Van Hoëvell's time and are indicated on his map.

7 As is customary with such survey wordlists, the items are written in raw phonetic notation, with no reference to possible morphological complexity. Nor are they free from printing errors: for Southeast Babar, I noted cases such as *urn* for *um* '1SG', and *mo'pan* instead of *moˈɲan* 'bird'.

8 I am grateful to Sander Adelaar for his valuable remarks on an earlier version of this paper.

in Letwurung but claimed only to have a passive understanding of Fien Laun's mother tongue.[9]

Section 2.2 is a phonology of Fien Laun's language; §2.3 is a brief discussion of the phenomenon of synchronic metathesis; §2.4 introduces Southeast Babar verbal morphology, in particular the many forms of the subject agreement prefixes.

2.2 Phonology

The Southeast Babar consonants are:

stops:	p	t	k
	b	d	(g)
nasals:	m	n	(ŋ)
fricatives:	(f)	(s)	x
trill, lateral:		r, l	
glides:	w	y	

The fricative /x/ has a rather backward and tense realisation in the position before the end of the word and before another consonant. In other positions its realisation tends to be more lax and plainly velar. The bilabial glide /w/ has a voiced bilabial fricative allophone in coda position.

/g/, /s/ and /ŋ/ only occur in loanwords such as /nasgoreŋ/ 'baked rice' (Indonesian: *nasi goreng*), *sag'eru* 'palmwine' (Moluccan Malay *saguer*), /teŋkur/ 'teacher' (from Malay/Indonesian *tuan guru* 'Mister teacher'?), *pasr* 'market' (Indonesian *pasar*), *-sux* 'like to' (Indonesian *suka*), *kakus* 'toilet' (Indonesian *kakus* < Dutch *kakhuis* 'shit house').

/f/ was attested in the verbal root *-tuf* 'close' (possibly also a loanword; cf. Indonesian *tutup* 'close, cover') and in *foto* 'picture'.

To what extent /b/ and /d/ can be called inherited is questionable. Whereas /t/ (which is always interdental) is omnipresent the voiced alveolar stop /d/ was only attested in one root as the second component of a root-final consonant cluster: *und* [und] 'wind' as opposed to *ukunt* [ukˈunt] 'high'. In loanwords *d* in the source language is assimilated to /t/: *tapur* 'kitchen' (Indonesian *dapur*).

The phoneme /b/ was found in a few words only, namely in *berk* 'heavy' and its 'excessive' derivation *be-berk-lol* [bɛˈbɛrklɔl] (RED-heavy-EMPH)[10] 'very heavy' (cf. Indonesian *berat*), in *basta* 'k.o. cloth (traditionally used as part of the bride price)' (origin unknown), *l-bel* '3SG-call (by telephone)' (from Dutch *bellen*). In (older?) loanwords *b* in the source language is reflected as *p*: the Southeast Babar equivalent of Dutch *bottel* 'bottle', Malay/ Indonesian *botol* is *potn*.

---

[9] Fien Laun deserves true admiration for her efforts to recall the language, which she had not used for such a long time. I am grateful for her help, hospitality, enthusiasm, and cheerfulness. Needless to say, any mistakes in the data presented below are a result of my own analysis and untrained ears.

[10] Abbreviations used are: C – consonant, CLASS – classifier, CMP – Central Malayo-Polynesian, EMPH – emphasis, INC – inchoative, PCMP – Proto Central Malayo-Polynesian, PL – plural, POSS – possessive, PRF – perfective, RED – reduplicated morph, SG – singular, V – vowel, 1,2,3 – first, second, third person, 1+2 – first person plural inclusive, 1+3 – first person plural exclusive.

In word-final position /y/ is often syllabified before a pause, i.e. it is followed by a non-phonemic optional schwa, even if /y/ is preceded by a vowel: e.g. *uty* [uty, 'utyə] (1) 'banana', (2) 'dog', *loy* [lɔy, 'lɔyə] 'proa', *waxy* [wɑχy, 'wɑχyə, 'wɑχəy] 'stone'.[11]

The language has a five-vowel system: /i, u, e, o, a/.

The default realisation of the mid vowels /e, o/ is relatively low: [ɛ, ɔ]. Word-finally there is a phonetic opposition between lower and higher mid realisations: [lɛ] 'day' vs [we] 'water'; [nɔ] 'in, at' vs. [o(w)] 'you (SG)'. These higher mid realisations are analyzed as realisations of diphthongs: /ey/ [e, ey], /ow/ [o, ow]. Especially for the back vowel the realisation with the offglide is regular.

In word-initial position, vowels may be preceded by a non-phonemic glottal stop.

Vowel length seems to be marginally distinctive. Fien Laun commented on her tendency to lengthen certain vowels, indicating that the generation of her mother did not yet do that. This lengthening seems to be frequent in (stressed?) monosyllabic words closed by /l/ or /m/. Vowel length appears to be phonemic, however, in some lexical roots, and at the boundary of subject (agreement) prefixes and some verbal roots, where they are possibly the historic result of vowel fusion. Compare the following oppositions:

(1)	*non* (1) 'name', (2) '3SG.with'	*no:n* '3SG.eat'
	ol '1SG.with'	*o:l* '1SG.eat'
	kal 'ground'	*ka:l* '1+2PL.eat'
	tel 'PRF'	*e:l* 'fish'

Stress does not seem to be phonemic at the word level. Often the same word occurs with different stress in comparable environments. Below I refrain with some exceptions from indicating stress.

As a result of vowel loss in prefixes and final syllables, Southeast Babar is rich in consonant clusters, both word-initially and word-finally. Some examples are:

(2)	*x-weapk*	'1+2PL-speak'
	l-tol	'3SG-hurt'
	t-lurk	'3PL-write'
	apl	'belly'
	irl	'nose'
	uty	'banana; dog'

Word-initially and word-medially geminate consonants are found, in most cases bridging root and affix or reduplicated roots. Some examples of geminate vs. single consonant oppositions are given in (3). Especially geminate stops tend to be shortened in allegro speech.

(3)	*lim-mo:l*	'2SG-hand'	~	*lim-o:l*	'1SG-hand'
	t-tol	'3PL-hurt'	~	*toyt*	'money' (< Indonesian/Dutch *duit*)
	l-loy	'3SG-dance'	~	*loy*	'prao'
	nnom	'sweet potato'	~	*non*	'name'
	otti	'a'	~	*foto*	'picture'
	m-mox	'2SG-vomit'	~	*mox*	'eye'

[11] Especially the sequence /-xy/ is variable as the latter example shows. This explains why Taber has *waxai* 'stone', *oxoi* 'louse', and *moxoi* 'die' where I have *waxy*, *oxy*, and *-moxy* respectively.

lal-l<y>aly ‘yellow’[12] ~ *i-l<y>a-l<y>a* ‘I am going’
(1SG-go-RED)

2.3 Metathesis

The form for ‘yellow’ in the final example is an instance of progressive metathesis, a phenomenon best known from Leti and most extensively described in publications by Van Engelenhoven (see for instance Engelenhoven 2004). In Leti such progressive metathesis is a regular syntactic and morphological process, occurring at the boundary of lexical roots, when a root ending in a high vowel is followed within the same syntagm or word by a root beginning with CV (provided that C not be a glide, and V not be a high vowel):

(4) …Ci + C_1V_1… > …CC_1yV_1…
…Cu + C_1V_1… > …CC_1wV_1…

In Southeast Babar word-final high vowels have merged and were reduced to *-y*. This *-y* metathesises with the initial consonant of the following root or word:

(5) …Cy + C_1V_1… > …CC_1yV_1…

Constraints as to the nature of C_1V_1 appear to be complicated: in any case V_1 may not be a high or a long vowel, and C_1 may not be a glide. If metathesis is blocked, *-y* is dropped unless it is preceded by a vowel. In (6) some examples of metathesis are presented.

(6)

laly + *laly*	>	*lal-l<y>aly* ‘yellow’
l-kary ‘3SG-work’	>	*l-kar-k<y>ary* ‘s/he is working’
l-moxy ‘3SG-die’ (< *n-matay) + *tel* ‘PRF’	>	*lmox t<y>el* ‘s/he has died’
l-moxy ‘3SG-die’ + *kay* ‘INC’ + *tel* ‘PRF’	>	*lmok*[13] *k<y>ay t<y>el* ‘s/he had already died’[14]

This metathesis may have originated in an earlier echo vowel, which became an infixed glide, after which the word-final high vowel (or *-y*) in that syntactic[15] or morphological position was dropped. The sequence *k<y>ay t<y>el* in the last example of (6), in which the final *-y* of the enclitic *kay* is not dropped, is a corroboration of this scenario.

2.4 Verbal inflection

As is generally the case in Blust’s Central Malayo-Polynesian languages, Southeast Babar verbs are inflected for subject. The prefixes involved are not always similar to the free personal pronominal forms and their Central Malayo-Polynesian reconstructions. They are presented in Table 2.

---

[12] Most equivalents of English adjectives are the result of some form of reduplication. The root in those cases seems to be verbal in some cases, but my data are insufficient to determine to what extent this is a regular pattern. Moreover, reduplicated forms may also be used as verbal stems.

[13] This assimilation of *-x* to *-k* before *k-* reflects a constraint on the historic change of *t (> $^{+}$k) to *x*, which did not occur before *-i if the following word began with *k-*. Compare *l-mox* ‘3SG-vomit’ (< *n-mutaq) + *kay* > *l-mox kay* ‘s/he starts vomiting’.

[14] Southeast Babar has a complicated system of aspectual markers. The glosses and translations given here are of a preliminary nature.

[15] The syntactic constraints need further research.

Table 2: Independent personal pronouns in Southeast Babar and the reconstructed cognates of Central Malayo-Polynesian (see Blust 1993:284).

		SEB	CMP
SG	1	*um*	**i-aku*
	2	*(y)ow*	**i-kau*
	3	*iy*	**s-ia*
PL	1+2	*exy*	**k-ita*
	1+3	*am*	**k-ami*
	2	*miy*	**miu*
	3	*ity*	**sida*

Sound changes in the verbal roots and possibly other factors have given rise to a rather bewildering variation in paradigms. In Table 3 some Southeast Babar verbal paradigms are presented as an illustration of this variety. Basically there are two patterns, those with syllabic and those with consonantal prefixes.

Table 3: Variety of verbal paradigms

		-xlil 'weep'	*-el* 'call'	*-la* 'go'	*-tol* 'see'	*-moxy* 'die'	*-wuty* 'make'	*-iml* 'drink'	*-(V)l* 'eat'
SG	1	*o-xlil*	*o-el*	*i-l<y>a*	*i-t<y>ol*	*i-m<y>axy*	*i-wuty*	*iml*	*o:l*
	2	*mo-xlil*	*mo-el*	*m-l<y>a*	*m-t<y>ol*	*m-m<y>axy*	*m-puty*	*m-iml*	*mo:l*
	3	*le-xlil*	*le-el*	*l-la*	*n-tol*	*l-moxy*	*l-wuty*	*l-eml*	*no:n*
PL	1+2	*ke-xlil*	*ke-el*	*x-la*	*x-tol*	*x-moxy*	*x-wuty*	*k-eml*	*ka:l*
	1+3	*me-xlil*	*me-el*	*m-la*	*m-tol*	*m-moxy*	*m-puty*	*m-eml*	*ma:l*
	2	*mi-xlil*	*mi-el*	*m-l<y>a*	*m-t<y>ol*	*m-m<y>axy*	*m-puty*	*m-iml*	*mi:l*
	3	*te-xlil*	*te-el*	*t-la*	*t-tol*	*t-moxy*	*t-wuty*	*t-eml*	*ta:l*

The first person singular forms in this table with a prefix *i-* and an echo glide *-y-* after the first stem consonant (as in the paradigms of 'go', 'see', and 'die') may drop the prefix after the first verb in a sequence of verbs, and in allegro speech in all syntactic positions.

At another occasion I hope to return to the Southeast Babar morphology and other aspects of grammar. In the following sections I shall focus on the formulation of some hypotheses about the history of the Southeast Babar sounds.

3 Southeast Babar: sound changes

3.1 Introduction

The most striking characteristics of Southeast Babar in terms of sound change are the reduction of final syllables, and a number of rather uncommon reflections of Proto Central Malayo-Polynesian consonants and vowels. One of these (*t > x) was discussed by Blust in his paper dealing with the change from *t to *k* (Blust 2004:393, 404–405).[16]

[16] Blust based his observations on Taber's Southeast Babar word list, which is not in all respects reliable, as indicated above. Some of the differences between Taber's data and mine are due to the fact that his list is based on the dialect of Kokwari. That said, however, a rather misleading aspect of the list is its lack of morphological analysis. As a consequence a variety of prefixed verb forms are presented as lexical roots.

The following is a more detailed survey of the sound changes that seem to have occurred in the shaping of Southeast Babar. It should be noted that the lexical corpus I elicited is limited, and that Southeast Babar lacks a reflex for quite a number of Proto Central Malayo-Polynesian (PCMP) reconstructions. In discussing the sound changes I base myself as much as possible on Blust's PCMP etymological reconstructions. In some cases I had to go further back or take recourse to a more regional cognate (notably from Leti or Wetan).[17] Section 3.2 discusses the loss of the consonants *z, *q, *h, *k, *p, *R, *j, and *-y, and §3.3 the preservation of *b, *w, *m, *l, and *r, and the change or *d, *t, *s, *n, and *ŋ. Section 3.4 deals with vowel developments. Section 3.5 contains some closing remarks.

3.2 Loss of consonants

Several Proto Central Malayo-Polynesian phonemes appear to have been lost in Southeast Babar. There are no traces, for instance, of Proto CMP *q, *h, and *z. Compare:

(7) *q > 0 :
*qapuR 'chalk' > *uir*, *qatəluR > *kely*,
*ma-qitem 'black' > *mexm*, *ma-qudip > *-mory*, *maRuqanay 'male' > *myal*,
*bunuq 'kill' > *-wuly*, *buaq 'fruit' > *wu* , *tanaq 'earth' > *kal*.

(8) *h > 0 :
*hikan 'fish' > *e:l*, *hapuy 'fire' > *uy*.

(9) *z > 0 :
*zalan 'road' > *al*

*k was also lost, for example:

(10) *k > 0 :
*kaRat 'bite'> *-ax*, *kayu > *ay*, *kita '1+2PL' > *ixy*, *kutu 'louse'> *oxy*,
*i-kau > *ow* '2SG', *yow* '2SG (+ polite)', *hikan 'fish' > *e:l*,
*burak 'white' > *wo-wor*, *tasik 'sea' > *kat*, *utak 'brain' > *ox* 'head'.

An exception to the loss of *k, noted by Blust (2004:393), is Taber's reflex of PMP *kaen 'to eat', namely *kal*. However, since Taber's verbs generally are inflected forms, I suspect that his *kal* is my *ka:l* '1+2PL.eat' (see Table 3 above). Yet 'eating' is still an exception, since it has a rather irregular parallel paradigm with syllabic prefixes and the root *-kkVl* [-k:Vl]. These forms were always translated into Malay as *sedang makan* 'be eating'.

Table 4: 'be eating'

SG	1	*o-kkol*
	2	*mo-kkol*
	3	*no-kkol*
PL	1+2	*ka-kkal*
	1+3	*mo-kkol*
	2	*mi-kkil*
	3	*ta-kkal*

[17] I am grateful to my colleague Aone van Engelenhoven for his input and suggestions in this respect.

Since partial reduplication of verbal roots marks 'progressive aspect', these forms are the result of a combination of such reduplication and partial analogy (as far as the variable root vowel is concerned) with the paradigm of the non-reduplicated base. Consequently the retention of *k was the result of reduplication: *-ka-kan > *-kkVl*.

The reflections of *p, *R, and *j are also zero, but not without exceptions.

(11) *p > 0 :
*(ə)pat 'four' > *(wo-)ax*[18], *pitu 'seven' > *(wo-)exy*, *punti 'banana' > *uty*,
*hapuy 'fire' > *uy*, *nipən 'tooth'> *lil*.
*malip 'laugh' > *-moly*, *maqudip 'live' > *-mory*

*p is reflected as *p* in *pant* 'hot, warm' (< *panas), in *pipy* 'goat (< *pipi), in *Kep*, the Southeast Babar equivalent of the Wetan speaking town of Tepa, and in the position *-mp- (which sequence later lost its nasal):[19] *kampung 'belly'> *apl*.[20]

(12) *R > 0 :
*Rumaq 'house' > *em*,[21]
*kaRat 'bite' > *-ax*, *maRi 'come' > *-moy*, *daRaq 'blood' > *ra*,
*baqəRu 'new' > *wa-way*,
*qateluR 'egg' > *kely*.

*R is reflected as *r* in *uir* 'lime' < *qapuR, and in *berk* 'heavy' < *ma-bəRat.

*j was lost in *qaləjaw 'day > *le*, and *ŋajan 'name'> *non*, but was preserved as *r* in *irl* < *ijuŋ 'nose', possibly because of the preceding *i.

Final *y was lost after *u. It was preserved after *a, which latter became lost, for example:

(13) *-y > 0 :
*hapuy 'fire' > $^{+}$apu > $^{+}$upi > *uy*, *babuy 'pig' > $^{+}$wawu > $^{+}$wawi > *wawy*;
*-ay > *-y* : *matay 'die' > $^{+}$mati > *-moxy*.

*y was maintained in the one PCMP instance where it occurred intervocalically:

*kayu 'wood' > *ay*

3.3 Preservation and change of consonants

The reflections of *b, *w, *m, *l, *d and *r are not much different from what they were in PCMP. As some of the above examples show, *b generally became realised as *w*:

(14) *balik 'turn' > *-waly*, *batu 'stone' > *waxy*, *burak 'white' > *wo-wor*,
*babuy 'pig' > *wawy*.

It has been preserved, however, in the root *berk* 'heavy' < *ma-bəRat, which is also exceptional because of its reflection of *R.

---

[18] The cardinal numerals from 2 to 9 have a prefix *wo-* (*wu-* before syllables with a high vowel), which probably derives from a classifier **buaq* 'fruit'.

[19] This may be a more general phenomenon: *t in the position *-nt- did not change either (see the discussion on *punti 'banana' below'). Our data do not contain an example of *-ŋk-.

[20] As a phoneme, /p/ occurs frequently, but not in words reflecting a protolanguage of any time depth.

[21] I cannot explain the vowel *e* in this word, nor in the reflection of *pitu in (11) above.

The following three reflexes of *w show that it was preserved word-initially, but not word-finally:[22]

(15) *waiR ‘water’ > *wey*, *lakaw ‘go, walk’ > *-la*, *qaləjaw ‘day’ > *le*

*l remained unchanged, and so did *m, except in *-mp-* clusters (as in *kampung ‘belly’ > *apl*):

(16) *laŋit ‘heaven’ > *lalk*, *maRi ‘come’> *-moy*, *lima ‘hand’ > *lim*, *daləm ‘inside’ > *ralm*;

*d merged with *r in *r*:

(17) *daləm ‘inside’ > *ralm*, *dua ‘two’ > *ru*, *dəŋəR ‘hear’ > *-rel*, *ma-qudip ‘alive’ > *-mory*.

Wo-wor ‘white’ < *burak is an example of preserved *r.

The remaining PCMP consonants *t, *s, *n, *ŋ underwent more radical changes. After the loss of non-geminate *k in all positions, the articulation of *t shifted backwards, resulting in a secondary velar consonant $^{+}$k. This $^{+}$k remained *k* in in root- or word-initial position before a vowel, and in word-final position after loss of the preceding vowel. In other positions it developed further into a fricative *x*:

(18) *t > *k*: *tanaq ‘soil’ > *kal*, *təlu ‘three’ > *wo-kely*, *ma-takut ‘afraid’ > *mkak*, *laŋit ‘heaven’ > *lalk*, *ma-bəRat ‘heavy’ > *berk*

Compare also:

(19)

Southeast Babar	*mork* ‘ hair’	Leti	*murut/murtu*
	Kep (town in West Babar)	Wetan	*Tepa*

Examples of *t in other positions are:

(20) *t (> $^{+}$k) > *x*: *teliŋa ‘ear’ > *xlil*, *batu ‘stone’ > *waxy*, *ma-qitəm ‘black’ > *mexm*, *matay ‘die’ > *-moxy*, *mata ‘eye’ > *mox*, *(ə)pat ‘four’ > *(wo-)ax*.

Compare also:

(21)

Wetan	*-tati*	Southeast Babar	*-kaxy* ‘sit’
Leti	*-tutun/-tutnu* ‘set fire to’		*ay kuxl* ‘fire wood’

Note that part of the above development (that is, the change from intermediate $^{+}$k to *x*) is also demonstrated in the following loanword:

Dutch *rekenen* [ˈrɛkənən] > Moluccan Malay *rekeŋ* [ˈrɛkɛŋ] > *-rexl* ‘count’

The only position where *t was preserved as *t* was directly after a homorganic nasal, as in *punti ‘banana’ > $^{+}$unti > $^{+}$uti > *uty*.[23] In other positions *n was preserved, either as *l* or as *n* (see below).

The change from *t to k/x paved the way for the following change:

(22) *s > *t*: *sulu ‘torch’ > *tuly*, *susu ‘milk’ > *-tuty*, *asu ‘dog’ > *uty*, *panas ‘warm’ > *pant*.

---

[22] My data lack examples of *-w-, but *-w-* was preserved in the loanword *kawl* ‘marry’ (< Indonesian *kawin*).

[23] It is possible that the reduction of *-i to *-y* occurred prior to the loss of *-n- before *-tV.

This change has not taken place in the loanword *pasr* 'market' (< Indonesian/Malay *pasar*).

The form *pant* suggests that *n was preserved before *t* < *s, and that the apparent merger of *ŋ (which presumably became $^{+}$n first) and *n with *l > *l* post-dated the change of *s. Before primary *t, however, *n is reflected as *l*, so the change must have occurred *after* *t had become a velar. The most likely scenario for the apparent preservation of *n as *n* before *t* < *s is that it is not a retention but a secondary development, since the loanword for 'bottle' (probably through Malay/Indonesian *botol*) is *potn*. This suggests that the merger of *ŋ and *n with *l was without exception, and that under the influence of neighbouring *t* $^{+}$l later changed (back) into *n*. This would also explain the 3SG subject (agreement) prefix *n-* before secondary *t-* in the non-syllabic inflection pattern (see Table 3 above), whereas *l-* is found nearly everywhere else.[24] Nearly, because there are also some other cases where the change of $^{+}$n (<*ŋ,*n) to *l* did not take place, namely in syllables which had a nasal in both onset and coda, such as *ŋajan 'name' > *non*. Compare also *t-ol* '3PL-be.with', *k-ol* '1+2PL-be.with', and *n-on* '3SG-be.with'; *ta-kan '1+2PL-eat' > $^{+}$ka-an > *ka:l*, and *na-kan '3SG-eat' > $^{+}$na-an > *no:n*.[25] Cases like *non* and *no:n* indicate that *j and *k had already disappeared when $^{+}$n changed to *l*, while the fact that *nipen 'tooth' is reflected as *lil* suggests that some reflection of *-p- was still present when that change occurred. Some examples of the merger of *n, *ŋ, and *l, and of the preservation of *l are:

(23) *nipən 'tooth' (> $^{+}$lipəl?) > *lil*, *bulan 'moon' (> bulal?) > *wol*, *bunuq 'kill' > *-wuly*, *inum 'drink' (> $^{+}$imun) > *iml*, *(ə)nəm 'six' > *(wo-)lem*, *ijuŋ 'nose' > *irl*, *təliŋa 'ear' > *xlil*, *laŋit 'heaven' > *lalk*, *lima 'hand, arm' > *lim*, *təlu 'three' > *wo-kely*.

In (24) examples are given of the synchronic alternation of *n* and *l* in reduplication and before the enclitic *tel* 'PRF' and *to:l* '1+2PL.POSS.SG'. As indicated above, the appearance of *n* in these examples must be a secondary development.

(24) *pant* 'warm, hot', *pal-pant* 'very hot'
-tol 'be ill', *-ton-tol* 'be very ill'
l- '3SG' + *-tol* 'see' + *tel* 'PRF' > *nton tel* 's/he has seen'
lewal 'language' + *to:l* '1+2PL.POSS' > *lewan to:l* 'our (incl.) language'
mo-mmyaly 'you (SG) get better' + *tel* 'PRF' > *mo-mmyan tyel* 'you (SG) have recovered'

3.4 Vowel change

After the loss of *-q, *-k, *-p, *-R, and *-y, final syllables were reduced; *a and *ə disappeared, while *-u and *-i merged, first in $^{+}$-i, to be reduced to *-y* in nearly all positions later on.[26] Some examples are:

---

[24] I cannot explain the initial nasal in *no-k-kol* '3SG is eating' (see Table 4).

[25] In loanwords a syllable containing both *n* and *l* appears to be avoided. Malay/Indonesian *kenal* 'know (someone)' has become *-kanan* (NB here *l* is reflected as *n*).

[26] It seems that this change did not occur when *u was stressed. In one set of possessive enclitics, which always clearly attracted phrasal stress in Fien Laun's pronunciation, *u is preserved as *u*, e.g. in *ox='u*, *ox=m'u* 'my head, your (SG) head'.

(25) *-ə > 0 : *dəŋəR ‘hear’ > *rel*;
*-a > 0 : *lima ‘hand’ > *lim*, *mata ‘eye’ > *mox*, *təliŋa ‘ear’ > *xlil*, *dua ‘two’ > *wu-ru*, *buaq ‘fruit’> *wu*, *tanaq ‘land’ > *kal*, *burak ‘white’ > *wo-wor*;
*-u > *y* : *batu ‘stone’ > *waxy*, *sulu ‘torch’ > *tuly*, *bunuq ‘kill’ > *-wuly*, *qatəluR ‘egg’ > *kely*, *babuy ‘pig’ > *wawy*, *baqəRu ‘new’ > *wa-way*;
*-i > *y* : *malip ‘laugh’ > *-moly*, *balik ‘turn’ > *-waly*.

*-ə- was lost in historical trisyllabic words but became *-e-* in roots which were or became monosyllabic:

(26) *təliŋa ‘ear’ (> +tliŋa)> *xlil*, *baqəRu ‘new’ (> +baRu) > *wa-way*.[27]
*nəm ‘six’ > *wo-lem*, *dəŋəR ‘hear’ > *rel*, *təlu ‘three’ > *wo-kely*,
*qatəluR ‘egg’ (> +təluR) > *kely*,

In originally penultimate syllables, *u was generally maintained, but it was lowered to *o* before open syllables with a non-high vowel, and before *x* < *t:

(27) a. *u > *u* : *buaq dua/lima ‘CLASS two/five’> *wu-ru/-lim*, *bunuq ‘kill’ > *wuly*, *susu ‘milk’ > *tuty*.

b. *u > *o* : *buaq təlu/(ə)pat/(ə)nəm ‘CLASS three/four/six’ > *wo-kely/-ax/-lem*, *mutaq ‘vomit’ > *-mox*, *utaq ‘brain’ > *ox* ‘head’, *bulan ‘moon’ > *wol*, *burak ‘white’> *wo-wor*, *kutu ‘louse’ > *oxy*.

This latter change presumably occurred after the classifier *buaq had developed into the prefix +bu- or +wu-, but before the loss of *a in the final syllable of independent words.

The greatest variation is found in the reflections of *a in originally non-final syllables.

(28) *-au > *-ow* : *i-kau ‘2SG’ > *ow*, *yow*.

(29) *-ay > +-i > *-y* : *matay ‘die’ > +mati > *-moxy*
*a > *e* if it was directly followed by *i (which became word-final and was reduced to *-y*)

(30) *waiR ‘water’ > *wey*.

In word-initial position (whether or not after loss of a historical initial consonant) *a- became *u-*:

(31) *asu ‘dog’ > *uty*, *hapuy ‘fire’ > *uy*, *kapuR ‘chalk’ > *uir*, *qabu ‘ash’ > *uwy*, *ama ‘father’ > *um*.

After a nasal *a became *o*:

(32) *mata ‘eye’ > *mox*, *malip ‘laugh’ > *-moly*, *mati ‘die’ > *-moxy*, *maRi ‘come’ > *-moy*, *na-kan ‘3SG-eat’ (> +naan) > *no:n*, *ŋajan ‘name’ (> +naan) > *non*.

In all other originally penultimate positions *a is retained as *a*:

[27] The change of the tri-syllabic words in (25) into bi-syllabic ones must have been part of a more general tendency, which also gave rise to the verbal inflection patterns with consonantal prefixes.

(33) *daləm ‘inside’ > *ralm*, *panas ‘warm, hot’ > *pant*, *k-ami ‘1+3PL’ > *am*, *zalan ‘road’ > *al*, *babuy ‘pig’ > *wawy*, *laŋit ‘heaven’ > *lalk*, *batu ‘stone’ > *waxy*, *tanaq ‘soil’ > *kal*, *daRaq ‘blood’ > *ra*, *balik ‘turn’ > *-waly*.

*a is also preserved in the first syllable of verb roots after an initial *m- in forms of the consonantal verbal paradigm in which the infix *-y-* appears, i.e. when the reconstructed pronominal prefix ended in a high vowel (see the paradigm of *-moxy* ‘die’ in Table 3)[28]. Compare also the 1SG and 3SG form of the non-reduplicated and the reduplicated stem of the verb *malip ‘laugh’:

(34) *ku-malip ‘1SG-laugh’ > *i-m<y>aly*,
*na-malip ‘3SG-laugh’ > *l-moly*,
*ku-malip-malip ‘1SG-RED-laugh’ > *i-m<y>al-m<y>aly* ‘I keep laughing’
*na-malip-malip ‘3SG-RED-laugh’ > *l-mol-m<y>aly* ‘s/he keeps laughing’.

Where a final consonant was preserved, the preceding vowel was often dropped, resulting in a word-final consonant cluster.

(35) *-CVC > *-CC*:
*matay ‘die’ > *-moxy*, *panas ‘warm, hot’ > *pant*,
*ma-qitəm ‘black’ > *mexm*, *daləm ‘inside’ > *ralm*,
*ijuŋ ‘nose’ > *irl*, *inum (> +imun) > *-iml*, *kampung ‘belly’> *apl*,[29]
*laŋit ‘heaven’ > *lalk*.

The loss of intervocalic consonants gave rise to vowel clusters, which subsequently underwent contraction, as shown below:

(36)

*aCi > +ai > *e*:	*ma-qitəm ‘black’ (>+maitm) > *mexm*	
*-aCu- > +-au- > *o*-:	*ma-qudip ‘alive’ (> +maudip) > *-mory*	
*aCa > +aa >	(1) *a*: *ŋajan ‘name’ (>+naan) > *non*	
	(2) *a:*: *nakan ‘3SG eat’ (>+naan) > *no:n* [30]	
*iCa > +ia > *e:*:	*hikan ‘fish’ (> +ial) > *e:l*	
*iCə > +iə > *i*:	*nipən ‘tooth’ (>+liəl) > *lil*	

3.5 Concluding remarks

The intermediate stages in the development of PCMP phonemes to their modern Southeast Babar reflexes (indicated above by a raised plus sign) are merely illustrations of the effect of certain early sound changes, setting the stage for later ones. For a more comprehensive chronology of sound changes, more data on Babar languages in general and on Southeast Babar proper are a prerequisite. The present survey is just the preliminary outcome of research ‘in progress’. But this progress will necessarily be slow because of the relative inaccessibility of the speaking community, and especially because the descriptive

---

[28] The development of the subject agreement prefixes of the consonantal inflection type can be traced as follows: *ku- ‘1SG’> +u- > *i-<y>-* , *mu- ‘2SG’ > +mi- > *m-<y>-*, *na- ‘3SG’ > +n- > *l-* (> *n-* before t-), or *na- > +la- > *l-* (> *n-* before *t-*), *ta- ‘1+2PL’ > +t- > +k- > *x-*, *ma- ‘1+3PL’ > *m-*, *mi- ‘2PL’ > *m-<y>-*, *sida- ‘3PL’ > +sa- > +s- > *t-*. The syllabic prefixes of the other inflection type are less easy to explain.

[29] This *p* was probably still preceded by a nasal when *p in +lip(ə)l (< *nipən ‘tooth’) was lost.

[30] I assume that the contraction to a short vowel is older, and that the loss of *k was a (much) later change than the loss of *j.

study of minority languages falls outside current academic priorities. Let us hope that at least some of the Babar languages will have been properly documented before the poles are melted.

References

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

—— 2004. *t to k: an Austronesian sound change revisited. *Oceanic Linguistics* 43.2:365–410.

Členov, M.A. 1976. *Naselenie Molukkskix Ostrovov*. Moskva: Nauka.

Dijk, Toos van. 2000. *Gouden eiland in de Bandazee. Socio-kosmische ideeën op Marsela, Maluku Tenggara, Indonesië*. Leiden: CNWS Publications.

Encyclopaedia, 1917–39. *Encyclopaedia van Nederlandsch-Indië* (IX volumes). ’s-Gravenhage: Nijhoff.

Engelenhoven, Aone van. 1996. Lirasniara: the sung language of southwest Maluku (East Indonesia). Paper presented at the XIth KITLV Workshop on Southeast Asian Studies: Endangered Languages and Literatures of Southeast Asia.

—— 2004. *Leti: a language of southwest Maluku*. Leiden: KITLV Press.

Engelenhoven, Aone van, and John Hajek. 2000. East Timor and the Southwest Moluccas: language, time and connections. *Studies in languages and cultures of East Timor*. Academy of East Timor Studies, 107–124. Sydney: University of Western Sydney.

Hoëvell, G.W.W.C. Baron van. 1890. De Afdeeling Babar. *Tijdschrift voor Indische Taal-, Land-, en Volkenkunde* XXXIII:187–199.

Josselin de Jong, J.P.B. de. 1935. *De Maleische archipel als ethnologisch studieveld*. Leiden: Ginsberg.

Riedel, J.G.F. 1886. *De sluik- en kroesharige rassen tusschen Celebes en Papua*.’s-Gravenhage: Martinus Nijhoff.

Steinhauer, Hein 1980. Review of M.A.Chlenov, *Naselenie Molukkskix Ostrovov*, Moscow 1976. *Bijdragen tot de Taal-, Land- en Volkenkunde (BKI)* 136:182–185.

Taber, Mark. 1993. Toward a better understanding of the indigenous languages of Southwestern Maluku. *Oceanic Linguistics* 32:389–441.

Taber, Mark, ed. 1996. *Atlas Bahasa Tanah Maluku.* Ambon. Pusat Pengkajian dan Pengembangan Maluku. Ambon City: Universitas Pattimura, Summer Institute of Linguistics.

Wurm, S.A. and Shirô Hattori, eds. 1983. *Language Atlas of the Pacific Area, part 2: Japan area, Taiwan (Formosa), Phillipines, Mainland and Insular South-East Asia.* Canberra: The Australian Academy of the Humanities in collaboration with the Japan Academy.

9 *Motherese and historical implications*

SHIGERU TSUCHIDA

1 Introduction[1]

Motherese is a non-standard speech register used by parents and other adults when talking to infants and/or pets. It used to be called baby talk before the 1970s.

Motherese words are usually ignored in comparative studies because it is believed that the relation between their shapes and meanings is not arbitrary, but often natural, as in onomatopoeia. It is, however, not necessarily so. Babies acquire not only their standard language but also motherese only through their parents or other adults. In this sense even motherese is a part of *langue* as against *parole* in Saussurian terms.

In this short paper I show some examples of motherese words in two Austronesian languages, Saaroa and Bunun, spoken in Central Taiwan, and I try to show that some words in Saaroa and Bunun most likely originated from the respective motherese registers in these languages.

2 Background

When I was investigating Saaroa in 1968 in Taiwan, I noticed that one of my Saaroa consultants, an old lady (born in 1908), was talking to her 2-year old granddaughter in quite a different tone, often using quite different words from those usually spoken. Just out of curiosity I asked her what she was saying, and I was able to collect a number of motherese words in Saaroa. I then tried to collect such words, whenever possible, in Kanakanavu, Tsou (Duhtu dialect), Bunun (southern dialect), and in Rukai (Maga and Mantauran). Readers who are further interested in motherese in Formosan Austronesian languages are referred to Tsuchida (1973). Although this article is written in Japanese, the main body is a comparative vocabulary list consisting of standard words in English followed by the corresponding motherese terms from the above six languages.

[1] I thank Professor Li Jen-kuei, Sander Adelaar and an anonymous referee for corrections and valuable comments on an earlier version of this paper.

Alexander Adelaar and Andrew Pawley, eds
Austronesian historical linguistics and culture history: a festschrift for Robert Blust, 107–114.
Pacific Linguistics, 2009.
© This edition vested with Pacific Linguistics.

It should be noted, however, that, due to the strict time limit imposed on foreigners staying in the mountain area in those days, I had to collect all the motherese words by asking my consultants directly how they would say such and such words when they talked to babies, and I was not in a position to make natural observations. According to what I was told, motherese words are not only used by parents or grandparents, but in fact by everybody talking to babies; brothers and sisters, adults and youngsters, men and women.

In June 2008 I had an opportunity to investigate Kanakanavu again, and found that it is not actively spoken anymore. Chinese and Bunun are predominant languages in the Kanakanavu area. Elderly people could recall the motherese words that I had collected about forty years ago, but they no longer used it with their grandchildren. Bunun people still talk to their babies in Bunun motherese. I believe that the Saaroa and Mantauran dialects of Rukai are in a similar situation, i.e. on the verge of extinction, and therefore their motherese registers are not used anymore.

2.1 Languages treated

In this paper I shall discuss only two languages, Saaroa and Bunun, and I will point out that some words in these languages most likely originated from their respective motherese registers.

Saaroa is spoken in Kaochung village, Taoyuan County, Kaohsiung Prefecture in southern Taiwan. The exact number of the ethnic population and of the speakers is not clear because, when counted, they are included together with the Kanakanavu in the Tsou ethnic group. Approximate figures would be about 400 people, but those who can speak their own language properly will be perhaps less than 20 or so by now.

Bunun is spoken by about 47,000 people in a large area in central to southern Taiwan, assimilating Saaroa and Kanakanavu. Its main dialects are northern, central, and southern Bunun. I investigated the southern dialect in Kaochung (the same village where I investigated Saaroa) and in Duhtu a (Tsou) village in Nantou Prefecture. These villages are geographically far apart. I did observe some dialectal differences, which are shown in Tsuchida (1973). However, the differences are not treated here because they are not relevant to the present discussion.

2.2 List of phonemes in Saaroa and Bunun

Phonemes in the two languages are as follows:[2]

Sar	p	m	t	n	k	ng	ʔ (1)	c (2)	v (3)	s	r	l (4)	lh (5)
Bun	p b (6)	m	t d (7)	n	k	ng	ʔ (1)		v	s z (8)	h	l (9)	

Sar	a	i	u	e (10)
Bun	a	i	u	

[2] Abbreviations are as follows: (A) – address term; Kan – Kanakanavu; (R) – reference term; RukMg – Maga dialect of Rukai; RukMn – Mantauran dialect of Rukai; Sar – Saaroa; Tso – Tsou.

Explanation of the above phoneme symbols:

(1) Glottal stop [ʔ].
(2) Voiceless affricate [ts].
(3) Either [v] or voiced bilabial fricative [β].
(4) Flap [ɺ].
(5) Voiceless lateral fricative [ɬ].
(6) Preglottalised [ʔb].
(7) Preglottalised [ʔd].
(8) Voiced interdental fricative [ð].
(9) In southern Bunun, a voiceless lateral fricative [ɬ].
(10) Schwa [ə].

3 Peculiarities in Motherese

The languages in central Taiwan have several peculiar motherese phenomena in common. Motherese words show (1) phonological peculiarities; (2) formal differences, especially in syllable structure; and (3) suppletion. Rules given below are not very strict and perhaps better qualified as general tendencies.

3.1 Phonological peculiarities

3.1.1 Replacement of phonemes

Some phonemes were replaced or lost. In the examples below, the numbers in parentheses indicate the serial numbers in the original list of comparative words in Tsuchida (1973). A hyphen indicates morpheme boundaries. Only those examples that have more than two parallel cases are mentioned.

In Saaroa the following replacements are observed:

- *l* → *ʔ* in initial position and zero in medial position:

Sar	*laeve*	→	*ʔaeve*	‘friend’	(53)
Sar	*liusu*	→	*ʔusu*	‘hips, buttocks’	(11)
Sar	*vulailhi*	→	*vaaʔi*	‘eye’	(2)
Sar	*calinga*	→	*cuunaa*	‘ear’	(4)

- *r* → zero in medial position except for in the last syllable where it is replaced by a glottal stop:

Sar	*um-a-ariva*	→	*ʔaiva*	‘hold in arms’	(21)
Sar	*tupuru-a* (Imp.)	→	*puupu-a*	‘Sit!’	(28)
Sar	*tikuru*	→	*kuukuʔu*	‘clothes’	(74)
Sar	*uuru*	→	*uaʔa*	‘cooked rice’	(44)

- *s* → *c*:

Sar	*saʔau*	→	*caaʔau*	‘tasty’	(52)
Sar	*saa-sare-ana*	→	*caacaeana*	‘earth, soil’	(77)

- *ng* → *n*:

Sar	*t-um-a-tangi*	→	*taataanii*	‘weep’	(22)
Sar	*calinga*	→	*cuunaa*	‘ear’	(4)
Sar	*macarengece*	→	*caaneece*	‘itchy’	(85)

- *v* → *ʔ*:

Sar	*eteve*	→	*teeʔe*	‘sugarcane’	(49)
Sar	*ma-vacange*	→	*ʔaacange*	‘good’	(79)

In (21) Sar *um-a-ariva* (→ *ʔaiva*) ‘hold in arms’, the replacement of *v* by *ʔ* does not occur, which may be the result of metathesis: *aʔiva* → *ʔaiva*.

In Bunun, the following replacements are observed:

- initial consonant → *ʔ*:

Bun	*mata*	→	*ʔata*	‘eye’	(2)
Bun	*tangis*	→	*ʔangis*	‘weep, cry’	(22)

There are 39 examples of such replacement, but there are also many exceptions, such as:

Bun	*taki*	→	*taci*	‘excreta’	(13)
Bun	*ngulus*	→	*ngulus*	‘nose’	(5)

- *-z-* → *-l-*:

Bun	*hazam*	→	*halam*	‘bird’	(63)
Bun	*makazav*	→	*kalav*	‘cold (weather)’	(83)

3.1.2 Simplification of consonant clusters

In Bunun, the first consonant of a consonant cluster is deleted except when the second consonant is *l*, in which case the *l* is deleted:

Bun	*bunbun*	→	*bubun*	‘banana’	(50)
Bun	*hasbing*	→	*ʔabing*	‘sneeze’	(30)
Bun	*cibuklav*	→	*bukav*	‘stomach, belly’	(9)

3.1.3 The appearance of non-standard sounds

Nasalised vowels, which do not appear in the adult language in Saaroa, are observed:

Sar	*civuka*	→	[ʔũka:]	‘belly’	(9)
Sar	*kalavungu*	→	[ʔõ:ĩ:]	‘water buffalo’	(68)

These nasalised vowels seem to have developed in onomatopoeia and in certain interjections and exclamations:

(i) in onomatopoeia:

Sar	[ʔõ:ĩ]	‘water buffalo’	(68), or:
Sar	*tarukuuka* [ʔũʔũã:]	‘chicken’;	(64)

(ii) in certain interjections and exclamations:

Sar	*mu-a-tii*	→	[ʔə̃:ʔã:]	‘Defecate!’	(14)
Sar	*maa-ta-tusuru*	→	[ʔãĩã] *mau*	‘Sleep!’	(18)
Sar	*ma-ruaru kia mana naani*	→	[ʔũã] *kia mana naani*	‘Stay here!’	(26)

Examples (66) and (16) require further explanation:

Sar *talaku* → [ʔĩ:a:] 'pig' (66)

This could be an onomatopoeic word designating the squeaking of a pig. However, judging from the shouting [ʔĩ:ã: ʔĩ:ã: ʔĩ:ã:] in the Mantauran dialect of Rukai calling for pigs nearby (as against the shouting [o:a o:a o:a] calling for pigs far away (invisible)) and the fact that the Mantauran motherese for pig is *ooaoa*, this Saaroa motherese [ʔĩ:a:] originally probably refers to the shouting to calling pigs.

Sar *taruaila* → [ʔõʔõko] 'carry on one's back' (16)

It is noticed that the other two Tsouic languages (Kanakavu and Tsou) have a very similar motherese term; compare:

Tso *s-m-ovri* → *ʔoʔoa, ʔooʔo* 'carry on one's back' (16)
Kan *um-a-ava* → *ʔooʔo* 'carry on one's back'

The Tsou, Kanakanavu, and Saaroa people usually carry their babies on their back to pacify them while rocking them and saying [ʔo:ʔo:ʔo:]. It is most likely that the motherese word for 'carry on the back' in these three languages comes from this dandling and pacifying call.[3]

In the southern dialect of Bunun the phoneme /l/ is always realised as a voiceless lateral fricative [lh], but in motherese it very often appears as a voiced lateral fricative [ɮ]:

Bun *ngulus* [ngulhus] → *ngulus* [nguɮus] 'mouth' (16)
Bun *matudul* [matudulh] → *tudul* [tuduɮ] 'sleepy' (17)

4 Syllable structure and formation of Motherese words

Syllables before the penultimate syllable are very often deleted, including prefixes. The remaining syllables may be called 'motherese stems.'

4.1 Deletion of syllables

In Bunun, the replacement of the initial consonant of motherese stems with the glottal stop often appears together with the deletion of pre-penultimate syllables:

Bun *ma-pa-haungun* → *ʔaungun* 'quarrel' (34)
Bun *ma-sabah* → *ʔabah* 'sleep' (18)

4.2 Syllable lengthening

In Saaroa, the two vowels of motherese stems are lengthened:

Sar *ramucu* → *muucuu* 'hand' (7)
Sar *sapale* → *paalee* 'foot/leg' (8)
Sar *civuka* → [ʔũ:ka:] 'belly' (9)

Either the ultimate or penultimate consonant is taken as the base, and the following canonical motherese word is formed: CV:CVʔV, as shown in the following examples:

---

[3] Incidentally, the Bunun people say [ʔive: ʔive:] in such a case.

Sar	*salumu*	→	*muumuʔu*	‘water’	(43)
Sar	*ngatili*	→	*tiitiʔi*	‘penis’	(12)
Sar	*amaʔa* (R), *kaamu* (A)	→	*maamaʔa*	‘father’	(53)
Sar	*tikura*	→	*kuukuʔu*	‘clothing’	(74)
Sar	*apulu*	→	*puupuʔu*	‘fire’	(76)

5 Suppletion

Many motherese words are not derived from the standard language but are the result of suppletion. Nasalised vowels mentioned above (see §2.1.3) appear all in the words which came from suppletion by onomatopoeia or interjections. Apart from these, there are also the following ones:

- In Saaroa:

Sar	*ʔapase*	→	*kuukiʔi*	‘crab’	(58)
Sar	*taamia*	→	*ʔaita kia*	‘Don’t!’	(24)
Sar	*mu-a-siri*	→	*ʔekaa*	‘stand up’	(29)
Sar	*mairange*	→	*taataʔa*	‘sweet potato’	(47)

- In Bunun:

Bun	*susu*	→	*vuvu*	‘breasts, milk’	(42)
Bun	*ma-daing*	→	*nanah*	‘big’	(87)
Bun	*sapuz*	→	*baba*	‘fire’	(76)
Bun	*ma-ʔansuh*	→	*ʔabuh*	‘tasty’	(52)

6 Historical Implications

Based on the analysis in the previous sections, we can infer some interesting historical changes that must have occurred in Saaroa and Bunun.

6.1 The Saaroa word for ‘meat’

It is well known that in motherese the word for food is *papa* or *baba* in many languages in the world (Ferguson 1964), such as Latin *papa*, Spanish *papa* ‘food,’ Moroccan Arabic *bappa*, *babba*, *pappa* ‘bread,’ etc. In four Formosan Austronesian languages that I could investigate, the same result was observed:

	Standard	Motherese	Meaning
Tso	*boni*	*paapa*	‘eat’
Kan	*ʔalame*	*paapa*	‘meat’
RukMg	*broo*	*papa*	‘cooked rice’
RukMn	*vaʔoro*	*papaa*	‘cooked rice’

In Bunun *papa* is the motherese for *ʔamaong* ‘carry on the back.’

The word for ‘meat’ is *papaʔa* in standard Saaroa, and there is no corresponding motherese word for it. On the other hand, the Kanakanavu word for ‘meat’ is *páapa*. Thus:

	Standard	Motherese	Meaning
Kan	*ʔaláme*	*páapa*	‘meat’
Sar	*papaʔa*	—	‘meat’

What is more, in Saaroa motherese, the syllable structure $C_1V_1V_1C_1V_1ʔV_1$ often appears (see §4.2 and §4.3). Thus *papaʔa*, the Saaroa word for ‘meat’ is most likely a back-formation from a motherese form **paapaʔa*.

6.2 The Saaroa word for ‘grandparent’

The Proto Austronesian word for ‘grandparent’ is *e(m)pu or *apu, whereas the Saaroa motherese word for the same meaning is *paapuʔu* as against its corresponding standard word *tamu* (A) and *tamuʔu* (R). It is very likely that this Saaroa motherese *paapuʔu* came from PAn *epu/*apu. It appears that the word *epu/*apu was replaced by *tamu* in Proto Kanakanavu-Saaroa (cf. Kan *tamu* (both reference and address)). Thus this is an interesting case of an original PAn form which survived in motherese.

6.3 The word for breasts

The Proto Austronesian word for ‘breasts’ is *susu. However, we can also reconstruct a Proto Hesperonesian word *nunu[H] based on the following cognate words: Tsou *nunʔu*, Pazeh *nunuh*, and Malagasy *nono* ‘breasts’. It is interesting to note that the Saaroa word for breasts is *ʔu-susu*, but its corresponding motherese is *nuunuʔa*, whose form conforms to the rules of Saaroa motherese word formation. We may well infer then that the protoword *nunu[H] is derived from the motherese word for breasts in the Hesperonesian protolanguage.

6.4 The Bunun word for ‘eat’

The Bunun word for ‘eat’ is *maun* (in actor focus), or *kaun-un* (in patient focus), and the word for ‘food’ is *kaun-an*. These words are apparently derived from Proto Austronesian *k-um-aʔen, *kaʔen-en, and *kaʔen-an respectively, but in that case, why do we find *maun*, and not **kumaun*? In §3.1 we explained that in Bunun motherese, only the last two syllables of the corresponding standard adult word are maintained. For example:

Bun	*ma-lanuhu*	→	*nuhu*	‘sit’	(28)
Bun	*ma-lodah-un*	→	*dah-un*	‘hit, strike (PF)’	(31)
Bun	*la-batu*	→	*batu*	‘throw’	(33)
Bun	*ʔis-ha-hailang-un*	→	*lang-un*	‘toys’	(78)

Thus it may be reasonable to consider that Bun *maun* ‘eat (actor focus)’ came from the corresponding motherese of *[ku]maun, which on turn derived from PAn *k-um-aʔen. But since then, a new motherese word *ʔaun* was derived from *maun* through back-formation (see §2.11).

The difficulty here is that there is no *-um-* infix observed in present-day Bunun, where PAn *-um- seems to have become a prefix *mu-* perhaps by metathesis, as in *mu-taki*

'defecate' or *mu-halhal* 'drop, fall.' If *maun* 'eat' is from *kumaun, we must assume that this change had occurred before *-um- became *mu-*.[4]

Although Bunun *mu-* seems to occur mostly in intransitive verbs (as in *mu-taki* and *mu-halhal*), it does occur in transitive verbs such as *mu-slud* 'remove something to (somewhere).'

Paul Li, on the other hand, suggests (pers. comm.) that it may be simply the result of alternation of *k*-initial verb with *m*-initial. Here is, however, again another difficulty of its own that *k-* and *m-* alternation appears usually in stative verbs, not in action verbs such as eating.

7 Concluding remarks

In this paper I discussed the possibility that

1. some words in the standard adult language may have originated from motherese;
2. some words in the standard adult language are innovations whereas more archaic forms may be found in motherese;
3. some words in the standard adult language may be reformations based on motherese.

Although motherese words are usually not treated at all by comparative linguists, it may not be a wise thing to ignore this register altogether in comparative linguistic studies.

References

Ferguson, Charles A. 1964. Baby talk in six languages. *American Anthropologist* 66.6.2:103–114.

Starosta, Stanley, Andrew K. Pawley and Lawrence A. Reid. 1982. The evolution of focus in Austronesian. In S.A. Wurm and L. Carrington, eds *Papers from the Third International Conference on Austronesian Linguistics,* vol. 2, 145–170. Canberra: Pacific Linguistics.

Tsuchida, Shigeru. 1973. Nursery words in central Formosan native languages. *Asian and African linguistics 2*. Tokyo: Institute of Asian and African Languages and Cultures, University of Tokyo Foreign Affairs, 1–51.

[4] Paul Li suggests (pers. comm.) that *mu- is older than *-um- as already pointed out in Starosta, Pawley and Reid (1982), and *mu-* is also attested in Pazih, and that it is probably original, whereas *-um-* is a later development and an innovation. However, it is very unlikely, I believe, that the prefix *mu- changed to the infix *-um-* in so many different languages all over in the western Austronesian languages including Taiwan independently, exactly because the infix *-um-* is rather unnatural and highly marked.

10 *The Proto Austronesian laryngeal*

JOHN U. WOLFF

1 Introduction

This paper is dedicated to Professor Robert Blust. Not only has he gathered a tremendous treasure trove of cognates that expand vastly the number of forms reconstructed for Proto Austronesian (PAn),[1] but he has also written on almost every aspect of historical phonology of the Austronesian (An) languages and brought resolution to innumerable thorny problems. This only begins to skim the surface, for there is hardly an area in An linguistics where Bob's contributions have not been significant and informative of what we know. My own work rests heavily on Bob's research, which he has always generously shared with me. It is in recognition of these achievements that I dedicate this paper to him.

In a seminal article David Zorc (1982) argued for the reconstruction of two laryngeals which he symbolised 'H' and 'ʔ' in addition to the reflexes of PAn *s[2] and *q, which had been termed 'laryngeals' by Dyen (1953). *s and *q need not concern us here, for they

[1] Abbreviations: An – Austronesian; PAn – Proto Austronesian; MP – Malayo-Polynesian; PMP – Proto Malayo Polynesian; Akl – Aklanon; Am – Amis; At – Atayalic; AtMx – Mayrinax Atayal; AtSq – Squliq Atayal; Bun – Bunun; Cb – Cebuano; Ilk – Iloko; Kav – Kavalan; Knn – Kanakanavu; Mar – Maranao; Ml – Malay; OJv – Old Javanese; Paz – Pazih; SedTko – Taroko Sediq; St – Saisiat; Tg – Tagalog.

[2] Most of the historical Austronesian literature symbolises this phoneme as *S. I assume a PAn phonology as follows, giving my symbols and those used by Blust (1995):

Wolff	p	c	t	none	k	q
Blust	p	s	t, C	c	k	q
Wolff	b	j	d	g, none	g	ɣ
Blust	b	Z	d	g	j	R
Wolff	s					h
Blust	S					H
Wolff	m	ɬ	n		ŋ	
Blust	m	ñ, N	n		ŋ	
Wolff	w	y	none	l		
Blust	w	y	r	l		

Alexander Adelaar and Andrew Pawley, eds
Austronesian historical linguistics and culture history: a festschrift for Robert Blust, 115–125.
Pacific Linguistics, 2009.
© This edition vested with Pacific Linguistics.

were not laryngeals in PAn. This paper is about *H and *ʔ and reexamines the data to determine whether in fact there were two PAn laryngeal phonemes, or whether there was only one. The thesis of this paper is that there was only one laryngeal phoneme and that there is no evidence for the reconstruction of a second.

2 The reconstruction of *h

First, let us look at the evidence for '*H', or as I shall call it, *h.[3] The first recognition of this phoneme published in a western language was Dyen (1965). In this influential paper Dyen argued that the Formosan evidence, e.g. the Am reflex of the word for 'woman' *fafahi* 'wife', indicates the need to reconstruct a new phoneme '*H' (Dyen 1965:302). At the time in my own thought (but unpublished) I brushed it off as a transitional [h] that had developed in the process of disyllabizing a form that I reconstructed as *bay.[4] However, attestations of /h/ occur not only in Amis (Am) but also in Atayalic (At), Pazih (Paz) and Saisiat (St), and a correspondence among the various forms that manifest /h/ in the same PAn etyma in these languages make it clear that this /h/ cannot have arisen as a product of transitional processes. This phoneme reflects PAn *h.

Dyen's proposal for *H was a very small part of his 1965 paper, and it was exemplified there only by three forms. Since then this phoneme has been discussed in a number of places. It might be worth noting here that among the earliest studies dealing with *h was a remarkable and lucid paper Bob Blust wrote as a graduate student in the seventies that discussed the word for 'woman' and the evidence it offers for the reconstruction of *h. Somewhat later Tsuchida's remarkable doctoral thesis appeared (Tsuchida 1976), which provided exemplification of almost all the Formosan data that leads to the reconstruction of *h. Tsuchida in fact reconstructed two *h's and believed them to be present in a host of forms that never had *h (cf. §2.5, below). Zorc (1982) reviews Tsuchida's data, rejecting some of Tsuchida's reconstructions, but again reconstructs forms with *h that never had an *h. One of the aims of this paper is to establish the criteria for reconstructing *h and exemplifying the forms that meet these criteria.

2.1 Root-final *-h

*h may be reconstructed in medial and final positions in the root, but the clearest cases are those in final position. *-h remained /-h/ in At, (when not obscured by the affixation of the 'male forms'[5]) Paz, St, and Am.

---

3 The tradition in An historical linguistics is to use 'H' to represent the reconstruction of this phoneme in PAn, and 'h' to represent the reconstruction of a phoneme in the protolanguage of a later sub-group of the family, usually Proto Malayo-Polynesian (PMP). I break with that tradition here.

4 There is a similar transitional phenomenon in Tg that parallels the putative development of transitional [h] in Am: *balay > Tg *báhay* 'house'. One assumes *báhay* developed from an earlier *baay (cf. Cb *baay*) that had come into being when the intervocalic *l was lost.

5 As is well known, the Atayalic languages developed 'male' forms, most of them by infixing a 'male' morpheme in the final syllable of the root, or affixing it to the end of the root. Because of syncopation and consonant cluster simplification, the male form of the root was substantially changed in its make-up from the original, such that the original form is often unknowable without reference to the inherited female form. As the male-female distinction had largely died out by the time our first citations of Atayalic forms were recorded and it was always the male form that was retained, most of the attestations are male forms and may not clearly reflect the PAn etymon. The Mayrinax dialect was the most conservative in this regard and some of the female forms are remembered by older speakers and can be cited. (Cf. Li (1980) for a discussion and exemplification.)

Table 1: *-h

*baɣah 'embers' [6]	At (Sq) *bagah* Paz *bahah* St *bælæh* 'charcoal' Am *valah* 'embers'
*baqeɣuh 'new'	At (Sq) *bah* Am *faɬloh* 'new'
*buŋuh 'head'	Am *fongoh* 'head'
*capuh 'sweep'	At (Mx) *sapuh* Paz *sapuh* St *sapuh* 'sweep'
*eqatih 'water recede'	St *maʔsih* 'ebb'
*galih 'dig'	St *kalih* 'dig' Paz *saa-kari* 'digging stick' (lack of Paz /-h/ unexplained)
*iqetah 'rice hull'	St *kæʔsæʔ* [7] Am *ɬtah* 'rice hull'
*jaɬih 'near'	At (SedTko) *dalex* Paz *alih* St *aɬʔaɬih-an* 'near'
*kukuh 'nail, claw'	At (SedTko) *kokox*
*nunuh 'breast'	At (SedTko) *nnux-an* 'has been sucked on' Paz *nunuh* 'breast'
*pukuh 'joint, node'	Am *pokoh* 'joint'
*paqah 'thigh'	St *ʔæpæh* 'thigh'
*qaciɣah 'salt'	Am *cilah* 'salt'
*quluh 'head'	St *ta-ʔœlœh* 'head'
*qumah 'field'	At (Mx) *qumqumah* Paz *umamah* St *ʔœmʔomæh* Am *omah* 'cultivated field'
*tuɬuh 'roast'	St *sœɬœh* Am *toɬoh* 'roast'
*tutuh 'strike'	St *totœh* 'hit, beat'
*uɬah 'go first'	Paz *ulah* 'go first' St *minʃaʔ-laʔ* [7] 'first'

2.2 Medial *-h-

It is clear that in medial position *h must be reconstructed in those cases in which Am or occasionally one of the other *h-preserving languages evinces /h/. However, in most cases medial *h has been lost in languages other than Am through syncopation of the penult and CC simplification. In those cases *h can be reconstructed, if at all, on indirect evidence, to be discussed presently. First, we take up the examples where medial *h is directly reflected.

---

[6] In this paper only the portion of each of the forms that reflect the laryngeals can be discussed. The reconstruction of many of the forms here cited involve complexities of development aside from the reconstruction of a laryngeal that need to be addressed. They can only be taken care of in a complete discussion of PAn phonology and with citation of many more reflexes than is presented here. My forthcoming *Proto Austronesian phonology and glossary* (in preparation) provides the complete explanation of the form of the reconstructions I cite here.

[7] In the St reflexes of *iqetah and *uɬah *-h assimilated to a /ʔ/ to the left—that is, became /ʔ/.

Table 2: Medial *h directly reflected

*bahaɬiɣ ‘board’	St *baɬihæh* ‘board’ (< *baɬihaɣ — metathesis and assimilation of *-ɣ to /h/ on left) Am *fahɬil* ‘cypress’
*buhet ‘squirrel’	At Mx *buhut* Paz *buhut* St *ka-bhœt* Am *fohet* ‘squirrel’
*buhuwaŋ ‘hole’, pit’	Am *fohang* Pu *buwang* Gorontalo *huwango* ‘hole, pit’
*cahebay ‘hang’	Am *ɬahpay* ‘hang’ [8]
*caheɬaɣ ‘shine’	Am *cahɬal* ‘come out between rain showers (sun)’
*lahuj ‘seawards’	Paz *rahut* ‘downstream’ St *læhœr* ‘downhill’
*paheɣaw ‘hoarse’	Paz *puhaw* ‘hoarse’ (< *pehaw < *pahehaw)
*paheku ‘fern’	Am *pahko* ‘an edible fern’

Now to look at the cases where the reconstruction of medial *-h- rests on indirect evidence. In some of these cases the reconstruction of *-h- is definitive. These will be considered first. The evidence for these comes from clusters in Philippine languages consisting of /ʔC/ or /Cʔ/ (a consonant plus a glottal stop or glottal stop plus consonant) where /ʔ/ is not a reflex of PAn *q,[9] e.g. Cb *bugʔat* OJv *bwat* ‘heavy’. OJv reflects PAn *q with /h/ and Cb reflects *q with /ʔ/, but in this case, OJv does not evince /h/. Therefore, the PAn form did not have a *q. My hypothesis is that this correspondence reflects PAn medial *h. Table 3 gives examples. Our hypothesis is supported by a sequence of events that jibe closely with other facts of phonological development from PAn to the current languages: namely, first, that PAn did not allow consonant clusters. Second, when *h was lost in pre-Proto Philippine (or Proto Philippine—it is hard to know how early *h was lost), hiatus between vowels developed.[10] Third, in most of the Philippine languages [ʔ] developed automatically between two vowels that came to abut. (The complete rule is that [ʔ] developed automatically before syllable-onset vowel.) Finally, in some cases, the penultimate vowel of the root was elided, giving rise to a consonant cluster consisting of a consonant and a glottal stop. In other cases an intervocalic glottal stop remains. The following table gives forms in which a Philippine language provides evidence for a medial *h. In most of these cases At, Paz, St, and Am, which overtly reflect *-h-, have syncopated the medial syllable or do not have a reflex of the root in question.

Table 3: Medial *h indirectly reflected by *ʔ in a Philippine language

*baɣehat ‘heavy’	Cb *bugʔat* OJv *bwat* ‘heavy’ (Lack of OJv /h/ indicates that there was no PAn *q.)
*betihec ‘calf’	Cb *batiʔis* OJv *wetis* ‘calf’
*kanuhec ‘squid’	Mar *kanoʔos* Ml *nus* ‘squid’ (Lack of Ml /h/ indicates that there was no PAn *q.)
*lahuwen ‘long time’	Cb *laʔun* ‘aged’ Muna *lao* ‘long (of the dry season)’. (Lack of Muna /ɣ/ proves that there was no PAn *q.)
*luhab ‘boil over’	Ilk *luʔab* ‘bubble up’ Muna *lua* ’bubble up’

[8] Am *ɬahpay* shows nasalisation of initial *c (*c nasalised is *ɬ̣) and the devoicing of *b (> /p/) by assimilation to the preceding /h/.

[9] Most Philippine languages reflect *q with /ʔ/. Here we are talking about cases where Philippine /ʔ/ does not reflect *q—that is, where no language has a reflex of *q but Philippine languages evince /ʔ/.

[10] In the extra-Formosan languages hiatus only developed when *h was the onset of the final syllable. When *h- was the onset of the penult or earlier syllable in tri-syllabic roots, no hiatus developed: the abutting vowels contracted.

Now we come to the least definitive cases of reconstruction of *-h-. In these cases no glottal stop is attested in Philippine languages, to my knowledge. I hypothesise a reconstruction of a tri-syllabic root and *h to account for these attestations. However, it cannot be ruled out that instead of the processes I hypothesise, some unknown analogy or borrowing has given rise to the attested forms:

Table 4: Not totally definitive reconstructions of *-h-

*buhelaɣ ‘impaired vision’	At Mx *mabuluq* ‘blind’ Kav *burar* ‘blurred vision’ Bun *buʔal* ‘cataract’ Tg *bulág* ‘blind’ Cb *bulúg* ‘cateract’
*bacuheq ‘wash’	Paz *bazu* Am *facaħ* Pu *vaseħ* Ml *basuh* ‘wash’
*paheliq ‘spleen’	Knn *páali* ‘bile, gall bladder’ Ml *pelih* ‘liver’
*sehapuy ‘fire’	At Mx *hapuy* Paz *hapuy* St *hapoy* Knn *apúlu* Bun *sapuð* Pai *sapuy*Tg *apuy* Ml *api*

I hypothesise that these forms descend from a tri-syllabic root with *h to account for irregular correspondences in the attestations. In the case of *buhelaɣ, Kav, Bun, and Tg reflect elision of the penult; Cb and At reflect metathesis of the penultimate and the final vowels. In the case of *bacuheq the attestations reflect different outcomes of vowel contraction after *h was lost:[11] Am and Pu reflect contraction of *ue > *e (and subsequently, /a/ and /e/ respectively), Paz and Ml reflect contraction to /u/. In the case of *paheliq, *h is reconstructed to account for the weakened penult in Ml. (In Ml the antepenult is weakened in tri-syllabic roots, but the penult of disyllabic roots is not weakened.) In the case of *sehapuy ‘fire’ there is a discrepancy in the correspondence of initial consonants that must be accounted for. Whereas Pa and Bun have /s-/, which can only reflect *s, Knn and the MP languages (Tg, Ml, and OJv) have Ø, which cannot originate in *s. The northern Formosan (Paz, St, and At) cognates have /h-/ and are ambiguous on this point, for /h-/ there could possibly reflect *s-, but may also reflect *h. To account for this discrepancy we assume a tri-syllabic root *sehapuy, with elision of *eh to produce Bun *sapuð* and Pa *sapuy* (a normal, well attested process in those languages) and loss of the antepenult *se- in northern Formosa to produce At *hapuy* and Paz *hapuy* and in the extra-Formosan languages to produce Tg *apúy* and Ml *api* (a normal process in those languages).[12]

2.3 Is *h directly reflected outside of Formosa?

*h is not directly reflected anywhere except in the four Formosan languages, At, St, Paz, and Am. No forms with clear reconstruction of *h (those of Tables 1 and 2, above) nor any of the forms with a less definitive reconstruction of *h have an /h/ in a language outside of Formosa. That does not mean, however, that given an /h/ in At, St, Paz or Am, a

---

[11] In Am *h was lost in tri-syllabic roots when it was the onset of the final syllable. Otherwise it was retained. This development is different from that undergone by the extra-Formosan languages, discussed in fn.10, above.

[12] Although the northern Formosan languages and the extra-Formosan languages have undergone the same development—loss of the antepenult, there is no reason to think that these developments were shared. They surely took place independently. Syncope (motivated by stress) was a widespread process in all branches of the An languages and happened repeatedly in many languages and in many inherited forms. All sorts of combinations of languages making parallel innovations are attested in the data from numerous protoforms that underwent syncope.

cognate in any other language will reflect Ø. In fact there are several cases where /h/ is reflected in one or more of these four languages, and a cognate outside of these languages and even in an MP language evinces /h/ as well. In those cases /h/ derives from *s, for in all but one of these cases a cognate with /s/ is attested in some language. In short, in the cases of /h/ in these four languages that correspond to /h/ in another language, the protoform is to be reconstructed with *s.[13] The following table presents the forms that evince the correspondence At, Paz, St or Am /h/ – MP /h/.[14]

Table 5: Etyma with reflexes evincing the /h/ – /h/ correspondence

*kasiw 'tree, wood'	At *kahoy* Paz *kahuy* St *kœhœy* Pa *kasiw* 'tree' Am *kasoy* 'firewood' Tg *káhoy* 'tree'
*kałusekus 'fingernail'	St *ka-kłœkœh* Kav *qnuqus* Pa *kałuskus-an* 'fingernail' Cb *kalukhu* 'scrape off' (< *kaluhkuh)
*(qałi)sipec[15] cockroach'	Paz *hipet* St *hipih* Kav *sipes* Am *łałipis* Ifugao *hipot* Ml *lipas* 'cockroach'
*qulasipan 'milliped	Paz *haripan* St *alongœhipan* Kav *lusipan* Cb *ulahípan* 'milliped'
*qusuŋ 'mushroom'	At *quhung* Cb *uhung* 'mushroom'[16]
*iseyup 'blow'	Paz *hium* St *hiop* Kav *s-m-iup* Cb *huyúp* 'blow'
*suɣ 'string beads'	At *lu-huɣ* St *sœ-hœl* Kav *tu-sur* Bun *tu-sul* 'string beads' Tg *tú-hog* 'put on spit'
*suni 'soft sound'	Paz *huni* Am *soni* 'sound' Tg *húni* 'pleasant sound'

2.4 Was there a second *h?

Tsuchida (1976:131–136) suggests the reconstruction of a second h-laryngeal which he termed '*H_2'. He found good correspondences only in final position, and these are the ones that we examine here. Although Tsuchida cites from a large number of languages, it is the forms from only two of them that are decisive: Bunun and Aklan from the Philippines. These two languages are not (have not been shown to be) in a subgroup, and if there is a

[13] There is no space here to discuss the reflexes of *s in the Formosan languages. Suffice it to say that *s has a double reflex in At, Paz, St, and Amis as well as in Tsouic and Thao—i.e, *s is reflected by /s/ and /h/ (or Ø < *h) in around twenty PAn forms. The languages that evince this double reflex have /h/ (or Ø) for half or less of these forms, and no two languages evince /h/ (or Ø) for the same PAn etyma. This is clear evidence of a sound change that took place in one dialect and spread on a word-by-word basis, but not to all of the Formosan languages—Rukai, Pa, Pu, and Kav were not affected. These forms are treated extensively by Tsuchida (1976:159ff.), Tuschida suggests that there were six different protophonemes to account for all the combinations of /h/, /s/, or Ø that the attestations evince.

[14] This list does not give all of the forms in this correspondence. There is a handful of forms that show /h/ in place of /s/ in Thao or in Tsouic, but not in these four languages, and have cognates with /h/ or deriving from /h/ in MP languages.

[15] Ml, Am and other languages manifest a reflex with the 'animal prefix' *qałi- (discussed in detail by Blust 2001). (The following sequence gives an account of the Ml development: *qałisipec > *qałihipec > *łihipec > *lihipec > *lipec > *lipas*.)

[16] It so happens that *qusuŋ does not have an attested reflex that evinces /s/. This is pure happenstance, and the /h/ in both At and Cb is sufficient to prove that the PAn etymon had *s.

set of regular correspondences between them, the only conclusion that can be drawn is that they reflect a phoneme of PAn.[17]

The correspondence we consider here, and which was believed to reflect *-H_2 is northern and central Bun Ø[18] and Akl /-h/ as opposed to northern and central Bun /-ʔ/ and Akl Ø.[19]

There are no other pairs of languages not in a sub-group from which attestations are cited in the literature to prove the existence of *H_2. Other pairs of languages that are cited to provide attestations proving a second PAn laryngeal are in fact in a low order sub-group. In short if Bunun and Aklan comparisons do not show regular correspondences in their endings of Ø and Aklan /h/ or Bunun /ʔ/, there is no evidence at all for the reconstruction of *H_2.

Table 6: Final /-ʔ/, /-h/, and -Ø in Bunun and Aklanon

PAn	Bunun	Aklanon
*aku	*ðaku*	*akó* 'I'
*ama	*ama*	*amah* 'father'
*batu	*batu*	*batóh* 'stone'
*buta	*butaʔ* 'mote in the eye'	*butáh* 'blind'
*ciku	*ciku*	*síko* 'elbow'
*daya	*daðaʔ* 'above, high'	*i-layá* 'interior region'
*daɣami	*dumali*	*dagámi* 'straw'
*dusa	*dusaʔ* 'two'	*duha-dúhah* 'be of two minds'
*ina	*t-ina*	*iná* 'mother'
*kita	*ita* [20]	*kitá* 'we (inclusive)'
*lima	*himaʔ*	*limáh* 'five'
*kutu	*kutu*	*kúto* 'head louse'
*mata	*mataʔ*	*matá* 'eye'
*mi	*ða-mi*[20]	*ka-mi* 'we (exclusive)
*paŋa	*panga* 'forked'	*panga* 'hook for getting fruit'
*piga	*pia*	*pila* 'how many?'
*qabu	*qabu*	*abóh* 'ashes'
*qaselu	*qusau*	*háeʔo* 'pestle'
*taqi	*takiʔ*	*táʔi* 'faeces'
*telu	*tau*	*tatloh* 'three'

[17] I go under the hypothesis that the extra-Formosan languages form a sub-group, the MP languages. This entails that a form that is attested in an MP language must also have a cognate attested in a Formosan language for us to know that it derives from PAn. (Cf. further discussion in the last footnote to §3.)

[18] Takituduh dialects devoice the vowel when final—i.e., pronounce /VØ/ as [-Vh]. Tsuchida (1976) cites vowel-final Takituduh forms with *-h.* We write *Ø* here.

[19] The source of the Bun data is Tsuchida (n.d.). Data from the southern Bun dialects (Ishbukun) are not indicative, as those dialects do not distinguish [-ʔ] from its absence—i.e., [ʔ] is appended to all vowel-final words: all words not ending in an obstruent, liquid, or nasal, must end in [ʔ].

[20] For Bun *ðami* and *ita* 'we' my only published source is Nihira (1988), which does not mark final glottal stop. I write vowel final and not /-ʔ/, in these forms on the basis of my field notes, made during my excursion to the central Bun area in 1997. As I mention in the §3.2, my informants were not entirely consistent with each other with respect to the final /ʔ/, even when they were natives of the same village.

There is not a great deal of data available for Bun, and there are few forms to compare. This list shows, however, that there is no correlation whatsoever between final vowels in Akl and /-ʔ/ in Bun, or between final /-h/ in Akl and final vowel in Bun. Which correspondence a given pair evinces is pure chance. The /-ʔ/ has an explanation other than that it is a reflex of a PAn phoneme, as discussed in §3.2 below. /-h/ in Akl developed as a transitional phenomenon—i.e., [h] developed between a root final vowel and a vowel initial suffix. This [h] became contrastive probably when glottal-stop transitions between vowels were borrowed from dialects or other languages, and the root-final /h/ that occurred between the root and affix was analogically spread to word-final position.[21]

3 Reconstruction of *ʔ

A number of scholars have expressed a belief that PAn had a phoneme *ʔ (Dyen 1965; Zorc 1996:passim; Tsuchida 1976:182–183).[22] The reconstruction of this phoneme by Zorc was based on the occurrence of glottal stops, especially in final position in Iban and the Philippines, but also some forms in Formosan and other Hesperonesian languages. (Table 7, below, gives examples.) While it is true that these languages evince glottal stops whose origins in many cases have not heretofore been explained, those unexplained glottal stops are not grounds for reconstructing a phoneme of the protolanguage. What is needed is a set of correspondences in two or more languages that are not in a subgroup. Further, if *ʔ can be reconstructed at all, it can only be reconstructed in word-final position. This is because the glottal stops attested in initial and medial positions originate as transitional phenomena in those languages in which /ʔ/ does not originate in *q or another PAn stop (cf. the discussion of glottal stop intercalation in §2.2).[23] Therefore, we will confine this discussion to the reconstruction of *-ʔ in final position. Bun is the only Formosan language that has a contrast between final /ʔ/ and its absence. To reconstruct PAn *ʔ, if this is possible at all, we need a comparison between forms in Bun and in one other language in which there is a two-way contrast reflecting what has been reconstructed heretofore as -V.

3.1 Do Iban and Bunun furnish data to demonstrate the existence of *ʔ ?

Zorc (1996:§2.2) believes that Iban provides evidence for the reconstruction of *ʔ for PAn. Indeed, in Iban /-ʔ/ is contrastive: some roots with a PAn etymology where cognates have /-V/ have /-Vʔ/ in Iban; others of these end in /-V/ in Iban. Zorc hypothesises that Iban /Vʔ/ < PAn *Vʔ, and Iban VØ < PAn *VØ. In order to establish the validity of this hypothesis we need to find a correlation with Bun data. The northern and central dialects of Bun (but not the southern Ishbukun dialect) do indeed have two reflexes for what is reconstructed as PAn *-V. In northern and central dialects this is [VØ] vs [Vʔ]. In Takituduh it is [Vh] vs [Vʔ] (cf. §2.2, fn.18). If it were to turn out that Bun [V] (Takituduh [Vh]) were to correspond closely to one

---

[21] Zorc 1969 is not definitive on the phonemic status of [-h], but in a personal communication he informs me that Akl distinguishes -Ø from /-h/: a word like *mata* ‘eye’ has a final vowel, but *tingah* ‘food particle in the teeth’ ends in /h/.

[22] Adelaar 1992 and Nothofer 1975 reconstruct *ʔ for the protolanguage of a subgroup. Adelaar assumes the existence of PAn *ʔ, and examines the proposition that Iban /ʔ/ reflects PAn *ʔ (65–67). He concludes that it does, albeit with doubts because of unexplainable counter-cases.

[23] Zorc thought otherwise in the case of *ʔ in medial position. If it should turn out that final *-ʔ did indeed exist, then it might be worth reviewing the data for other positions, but if not, it is totally unlikely that the reconstruction of PAn *ʔ will explain attestations of /ʔ/ in other positions.

of the two possible endings in Iban and Bun [Vʔ] were to correspond to the other possible ending in Iban, then voilà, we have incontrovertible evidence for reconstructing PAn *ʔ in final position, for Iban is in the MP group and Bun is not.[24] If it does not turn out to be the case, there is no evidence for *-ʔ, and to date this is not one of the known phonemes of PAn.[25]

The following chart shows the PAn etyma reconstructed with /-V/ that I have been able identify as having Iban and Bun reflexes.

Table 7: Cognates in Bunun and Iban evincing /-ʔ/

PAn	Bunun	Iban
**asu*	*acuʔ*	*asu* 'dog'
**aku*	*ðaku*	*akuʔ* 'I'
**batu*	*batu*	*batu* 'stone'
**buta*	*butaʔ* 'mote in eye'	*butaʔ* 'blind'
**ciku*	*ciku*	*siku* 'elbow'
**daya*	*daðaʔ* 'above high'	*dayaʔ* 'inland'
**dusa*	*dusaʔ*	*dua* 'two'
**duɣi*	*duliʔ*	*duriʔ* 'thorn'
**ina*	*t-ina*	*ina* 'mother'
**juɣami*	*dumali*	*jeramiʔ* 'straw'
**kita*	*ita* 'we' [20]	*kitaʔ* 'you (plural)'
**kutu*	*kutu*	*kutu* 'head louse'
**lima*	*himaʔ*	*limaʔ* 'five'
**ɬuka*	*nukaʔ*	*luka* 'wound'
**mata*	*mataʔ*	*mataʔ* 'eye'
**mi*	*ða-mi* [20]	*ka-mi* 'we (exclusive)'
**paŋa*	*panga*	*pangaʔ* 'forking'
**qabu*	*qabu*	*abu* 'ashes'
**qaɬitu*	*qanituʔ*	*antu* 'evil spirit'
**qaselu*	*qusau*	*alu* 'pestle'
**taqi*	*takiʔ*	*taiʔ* 'faeces'
**taqu*	*taqu* 'inform'	*tauʔ* 'know'
**waɣi*	*vali* 'sun'	*ari* 'day'

[24] I assume that any MP form that has a Formosan cognate or any MP feature that has a Formosan analog remounts to PAn. This is disregarding the fact that more than a thousand years elapsed between the earliest Austronesian settlement on Taiwan and the earliest emigration of the speakers of PMP—i.e., making PAn include languages that were spoken at a time much later than the the earliest settlement of the Austronesians in Taiwan (but earlier than the departure of the community that spoke PMP). In short, a feature shared by the MP languages and one of the Formosan languages may have come into being much later than the earliest Austronesian settlement of Formosa.

[25] Zorc also brings into the discussion forms from the Philippines of PAn provenience that have a final /-ʔ/ where *V has been reconstructed. However, there are only a handful of these, and /-V/ in the Philippines corresponds to both /V/ and /Vʔ/ in northern and central Bun. The explanation of /-ʔ/ in the Philippine languages is most likely going to come from processes or events that took place after proto Philippine times.

It is clear from this handful of correspondences that there is poor correlation. Four Iban forms with /-V/ have a cognate in Bun with /Vʔ/ and five Iban forms with /-Vʔ/have a cognate in Bun with /-V/. Fourteen of the 23 forms agree in their ending. Even if more correspondences can be found, the additional data are not going to change the conclusion: nine exceptions in out of 23 is tantamount to no correlation. It is highly unlikely that Iban can ever provide evidence for the reconstruction of PAn *ʔ.

3.2 The final glottal stops in Iban and Bunun

Many of the Formosan languages and the Hesperonesian languages append a non-contrastive laryngeal to words ending in a vowel—i.e., all words must end in a C, and if they do not end in one of the consonant phonemes, they have [ʔ] or [h] at the end of the word. This is the case for example on western Java: both Sundanese and west Javanese dialects have no final vowels. Words with etymological final vowels have appended a non-contrastive [ʔ]. After Malay was introduced to the area (Jakarta) a certain portion of the new Malay speakers were native speakers of dialects that automatically appended a non-contrastive [ʔ]. Their non-contrastive [ʔ] was introduced into their version of Malay. It was imitated by other members of the community and became a variable for them. Some of the items with [ʔ] came to be taken over by the dialect of Malay spoken in Jakarta for all members of the speech community, some were not, and in this way the final glottal stop came to be contrastive.[26] Some forms remain variably with [ʔ] and without [-ʔ] to this day. Thus we have *juga?* 'also', *minta?* 'ask' and many other forms invariably with a /-ʔ/; *taña?* and *taña* 'ask (a question)', where the glottal stop is variable; and *bisa* 'can' (never with /-ʔ/, to my knowledge).[27] The situation of Bun and Iban is quite similar to that of Jakarta Malay. Like Jakarta Malay, Bun is spoken in an area surrounded by languages and dialects that have non-contrastive glottal stop at the end of words that etymologically ended in a vowel. Further, my very limited first-hand experience with some of the northern and central dialects of Bun made it clear that the final glottal stop, at least in some lexical items, has become (or maybe always was) a variable in the speech communities I visited. Informants who were born and lived their lives in the same village in some cases did not put glottal stops in the same words. Another indication: the word for 'hand' *ima* does not occur in the northern and central dialects with /-ʔ/, whereas the word for 'five' *hima?*, which originates in the same PAn root *qalima, invariably has a glottal stop.

It is quite possible, and I believe likely, that a Jakarta-like situation prevailed in Bun—i.e., that Bun borrowed forms from dialects having non-contrastive glottal stop after word-final vowels, but it was happenstance which forms with final glottal stop were borrowed. This would account for the fact that the appearance of /-ʔ/ does not correlate with /-ʔ/ in another language. The Iban situation is different from Bun only in the fact that Iban is largely in contact with languages and dialects that have contrastive glottal stops. However, the distribution of the glottal stop in Iban is a clear indication that the addition of glottal stops in Iban is due to language or dialect contact (perhaps not so recently as is the case of Bun)—that is, the failure of the glottal stops in Iban to jibe with glottal stops in other languages and dialects (as shown for example in Adelaar 1992:§3.4.2.4) is most certainly due to late development of these glottal stops in Iban. They surely developed in a situation

[26] Another development that made final glottal stop contrastive was the change of final /-k/ to /-ʔ/.

[27] Adelaar (1992:67–69) discusses the final /ʔ/ of Jakarta Ml. His conclusions are tantamount to those presented here.

very similar to that of Bun and Jakarta Malay, for automatic glottal stop in word-final position is characteristic many of the languages of Kalimantan, including some of south and southwest Kalimantan. In short, neither Iban nor Bun present evidence for the reconstruction of a PAn glottal stop.

References

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

Blust, Robert A. 1995. *Austronesian comparative dictionary*. Online manuscript.

—— 2001. Historical morphology and the spirit world: the *qali/kali-prefixes in Austronesian languages. In J. Bradshaw and K.L. Rehg, eds *Issues in Austronesian morphology: a focusschrift for Byron W. Bender*, 15–73. Canberra: Pacific Linguistics.

Dyen, Isidore. 1953. The Proto-Austronesian laryngeals. *William Dwight Whitney Linguistic Series,* No.9. Baltimore: Linguistic Society of America.

—— 1965. Formosan evidence for some new Proto-Austronesian phonemes. *Lingua* 14:285–305.

Li, Paul Jen-kuei. 1980. Men's and women's speech in Mayrinax. *Papers in honor of Professor Lin Yü-k'eng on her Seventieth birthday*, 9–17. Taipei: Wen-hsing Book Co.

Nothofer, Bernd. 1975. *The reconstruction of Proto-Malayo-Javanic.* 's-Gravenhage: M. Nijhoff.

Nihira, Yoshiro. 1988. *A Bunun vocabulary*, Third Edition. Privately printed.

Richards, Anthony 1981. *An Iban–English dictionary.* Oxford: Oxford University Press.

Tsuchida, Shigeru. 1976. Reconstruction of proto-Tsouic phonology. *Study of languages and cultures of Asia and Africa,* Monograph Series, No.5. Institute for the Study of Languages and Cultures of Asia and Africa: Tokyo Gaikokugo Daigaku

—— n.d. File of approximately 1000 lexical items in eight Bunun dialects with a reconstruction.

Wolff, John U. 1999. The monosyllabic roots of PAn. In Elizabeth Zeitoun and Paul Jen-kuei Li, eds *Selected papers from the Eighth International Conference on Austronesian Linguistics*, No.1, 139–194. Taipei: Academia Sinica: Symposium Series of the Institute of Linguistics (Preparatory Office).

Zorc, R. David. 1969. *A study of the Aklan dialect,* vol. 2, 'Dictionary'. Kalibo.

—— 1982. Where, O where, have the laryngeals gone? Austronesian laryngeals re-examined. In A. Halim, L. 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 B. Nothofer, ed. *Reconstruction, classification, description: festschrift in honor of Isidore Dyen*, 41–72. Hamburg: Aberla Verlag.

Part 3

Grammatical change and typology

11 *The various origins of the passive prefix* di-

ALEXANDER ADELAAR

1 Introduction[1]

The prefix *di-* is a frequent passive[2] marker in various West Indonesian languages. It occurs, among others, in Malay and many other Malayic languages as well as in Javanese, Sundanese, Lampung, Toba Batak and some of the South Sulawesi languages. Many explanations have been suggested for Malay[3] *di-* (Adelaar 2005a). While there are strong indications that it originated through cliticisation of the preposition *di*, this is not necessarily the origin of this prefix in other West Indonesian languages. Some of these languages must have obtained it through borrowing, which seems to be an ongoing process in which Indonesian, the national language, nowadays plays a major role. However, for most languages with *di-*, it remains to be demonstrated that this prefix was borrowed.

Important questions that need to be addressed:

1. are the passive *di-* prefixes in languages other than Malay related to Malay *di-*?
2. if yes, how did *di-* end up in each of these languages as well as in Malay itself? Through a common ancestor, through borrowing, or as the result of independent developments?
3. if it developed independently, was *di-* the result of a development parallel to the one proposed for Malay (that is, *di (a preposition) > *di-* (a passive prefix)? Or was it the result of a different process altogether?
4. if it was passed on through borrowing, what was the source, or, what were the sources?

[1] With gratitude to Bob for a first-rate supervision, and a great friendship ever since.

[2] In modern Malay and Indonesian, clauses with verbs prefixed with *di-* are basically passive in terms of their morphosyntactic properties and pragmatic function (see van den Berg (2004:533–535).

[3] In this paper 'Malay' refers to standardised forms of Malay such as Indonesian, standard Malaysian, and their literary Malay predecessors.

Alexander Adelaar and Andrew Pawley, eds
Austronesian historical linguistics and culture history: a festschrift for Robert Blust, 129–141.
Pacific Linguistics, 2009.
© This edition vested with Pacific Linguistics.

I am not able to answer these questions for each language exhibiting *di-*. However, it appears that at least in the case of modern Javanese, *di-* is the result of a very different process involving neither proclitisation of a preposition, nor borrowing. In the present paper I trace its origins, showing that this passive prefix as well as the preposition *dening* introducing (among others) actor both derive from the Old Javanese noun *de* '1. action, way, manner'.[4] I also explore possible connections between the Javanese and Malay *di-* prefixes.

The paper is organised as follows. Section 2 is a short overview of the various explanations of Malay *di-*; §3 and §4 describe the various meanings and applications of *de* in respectively Old and modern Javanese; §5 discusses the data and possible connections between Malay and Javanese *di-*. Section 6 is a conclusion.

2 Theories about the origin of the Malay passive prefix *di-*

At least five explanations have been proposed for the origin of Malay *di-*, some of which are not necessarily mutually exclusive. They are listed and commented on below. Although I am in clear favour of the first of these explanations, none of them may be convincing enough to bring the discussion to a satisfactory close.

1. *Di-* developed from the preposition *di*, which became cliticised to the root of the following passive verb. This theory was proposed by Aichele (1942–43) and developed further by myself (Adelaar 2002, 2004, 2005a, 2005b).

My own argumentation is primarily based on historical linguistics and comparative evidence from several Malayic varieties. These data clearly show that across these varieties, *di* has a wide array of prepositional meanings beyond that of locative preposition, including those of actor (Minangkabau,[5] Kanayatn;[6] cf. also Perak Malay *də*[7]) and genitive (Kanayatn, Banjar Malay[8]). They also show that in Kanayatn, *di* functions as an actor preposition if the actor (which usually occurs directly before the verbal stem) is expressed; if it is not, *di* is prefixed to the verb (see further §5). As a consequence, it appears that *di* as a preposition originally had a much wider application than it has today in Indonesian and was also an actor preposition; as it was cliticised to the following verb by abstention of an actor, it was finally reinterpreted as a passive marker. The problem with this explanation is that from a wider typological perspective, the development from a preposition to a passive prefix is unprecedented, as pointed out by Wolff (2001) and van den Berg (2004). However, there are several Austronesian languages that independently show a parallel development in which the same marker occurs as an actor preposition and a passive prefix,

---

4 I benefited much from discussions about the origins of *di-* with Alexander Ogloblin in Jakarta (June 2005 and email communication), who had noticed the same connection and drew my attention to his insightful analysis of *de* in Ogloblin (2000). Many thanks also to Waruno Mahdi and Arlo Griffith for their valuable comments on an earlier version, and to Elisabeth Riharti for providing me with modern Javanese sample sentences. Early references to the connection between *de* and *di-* are Kern (1899:306–313), Swellengrebel (1933:18–20) and Haaksma (1933:63), although these sources are not very transparent and do not refer to the nominal origins of *de*.

5 Minangkabau (spoken in West Sumatra) has the variant forms *di* and *dè?* (Moussay 1981:242).

6 Kanayatn (spoken in West Borneo) is also known by its Malay name Kendayan; it includes the dialects Ahe and Salako (Adelaar 2005a).

7 Brown (1921:11 note 3); Perak Malay is spoken in Perak, West Malaysia.

8 Banjar Malay is the main language in South Kalimantan and is also spoken in parts of neighbouring Central and East Kalimantan (Indonesian Borneo).

such as Belangin (a Kanayatn-type Malayic variety), Minangkabau Malay (Adelaar 2005a), Sinama Bajau (Akamine 2005) as well as Madurese and, in historical perspective, Javanese,[9] as I will demonstrate in the following pages.

2. *Di-* developed from an earlier passive prefix *ni- (still attested in 7th century Old Malay inscriptions) through denasalisation of initial *n-, in the same way as the intransitive prefix *bər-* must have developed from an earlier *maR- through denasalisation of initial *m- (Casparis 1956; Teeuw 1959; Hopper 1988; van den Berg 2004).

In Adelaar (2002, 2004) I argue that a development from *ni- to *di-* is phonologically unlikely. On the one hand, denasalisation is not part of the phonological history of Malay, and the change from *maR- to *bər-* must be due to the emergence of an epenthetic *b caused by the reduction of *maR- to *mr- (/*mər-), in a development as follows: *maR- > *mr- > *mbr- > *br- > *b(ə)r-*. On the other hand, the neutralisation of antepenultimate vowels to schwa is very much part of this history. A historical prefix *di- should therefore have become *də-, which evidently has not happened. However, van den Berg argues that affixes are not subject to sound change in the same rigorous way as lexical elements and may show changes or retentions that are not shared by the latter. He also finds additional evidence for a change from *ni- to *di-* in South Sulawesi languages, which he considers a parallel but independent development. In Adelaar (2005a) I disagree with van den Berg's exemption of affixes from the effects of regular sound change. Exceptions can only be made if there is a compelling alternative explanation (and such explanation is lacking in the alleged evolution from *ni- to *di-*). I assume that South Sulawesi *di-* is due to influence from Malay, either directly as standard Indonesian influence on various South Sulawesi minority languages, or indirectly via Buginese. The latter must have borrowed Malay *di-* and passed it on to other South Sulawesi languages before it changed this prefix to its present-day form *ri-*. However, it is also possible that the *di-* prefix in some of the South Sulawesi languages is the result of an independent development.

3. *Di-* is originally a cliticised form of the third person pronoun *dia* (van Ophuysen 1903:145, 165; Shibatani 1985).

Such an origin is unlikely because, as Wolff (2001) points out, *dia* is originally the 'dative' form of a nominative pronoun *ia*. Using the dative form to express actor would be odd, especially considering that both other actor prefixes *ku-* (1st person) and *kau-* (2nd person) are not derived from dative pronouns (in spite of the fact that such dative counterparts do exist, cf. *daku* and *dikau* respectively).[10] The derivation of *di-* from a 3rd person pronoun would also leave unexplained why verbs with *di-* still require *–nya* as a default 3rd person actor suffix.

4. Old Malay had the verbal prefixes *bər-* and *di-*, but in the 7th century court language of the South Sumatran inscriptions, these would have been replaced by respectively *mar-* and *ni-*, which were Batak loan affixes; sub-dialectally and in later inscriptions, however,

---

9 Special thanks to Alexander Ogloblin (email communication) for pointing out a parallel development from preposition to passive prefix involving Madurese *e*/*-e* and Middle Javanese *den*/*den-*.

10 Another of Wolff's (2001) counterarguments is that, 'according to the phonological rules which operate elsewhere in Malay', a cliticised form of *dia* would have lost its first vowel instead of its second one. I assume that Wolff is referring to the fact that the cliticised forms of *aku* and *əngkaw* are based on the second syllable of the free pronouns, whereas the cliticised form of *dia* would be based on the first syllable. While I agree with Wolff's overall conclusion that *di-* does not derive from *dia*, from a phonological point of view I am not convinced that *dia* could not yield a clitic *di-*, as its phonotactic structure is very different from that of *aku* and *əngkaw*.

bər- and *di-* would have remained in use and would eventually outlive *mar-* and *ni-* (Aichele 1942–43; cf. also Mahdi 2005:180).

To assume substantial Batak influence in Srivijaya is somewhat speculative, given our limited knowledge about Srivijaya and the early history of Malay, and bearing in mind that this empire had Malay as a court language, while the Batak languages were spoken in politically dependent uphill areas at its periphery. Moreover, as Teeuw (1959) already pointed out, the nature of the linguistic data does not require us to look for solutions outside the history of the Malayic subgroup itself. In other words: why look for Batak influence per se if we know from comparative evidence that *maR- and *ni/*<in> were already present in Proto Malayo-Polynesian and Proto Austronesian anyway? An additional problem with this explanation is that it does not address the question of the *ultimate* origin of *di-* or *bər-*, (which are traced to Proto Malayic *di- and *bar- respectively [Mahdi 2005:184]), but rather that of *ni-*.

5. The prefix is a conflation of a third person plural enclitic pronoun *da- (< Proto Malayic *sida '3rd person plural') and the passive prefix *ni-* (Wolff 2001).

As van den Berg (2004:548–549) already demonstrated, while this conflation is not inconceivable, in the present case it is based on too many speculative assumptions to be a likely explanation.

3 Old Javanese *de* appears as a noun and a function word

The discussion about the origins of Malay *di-* will have some implications for other West Indonesian languages that have the passive prefix *di-*. This prefix is not inherited from Proto Austronesian or Proto Malayo-Polynesian, and while in some West Indonesian languages it may have a very different history from that of *di-* in Malay, in some other ones it may be borrowed from Malay or have originated through developments parallel to the ones undergone by Malay *di-*.

In Javanese, however, *di-* has a rather different origin. It is derived from *de*, which was still a noun as well as a function word in Old Javanese. In this language, as in Malay, the passive prefix *di-* is a relative newcomer. Old Javanese did not have it, using instead the affixes <*in*> and *ka-*. In addition, the vowel of Javanese *di-* should have become schwa on account of a very regular tendency towards antepenultimate vowel neutralisation (Nothofer 1975:51–71). However, in contrast to Malay *di-*, Javanese *di-* can be traced within Javanese itself and appears to have a lexical source in this language.

Zoetmulder (1982) gives the following meanings to the Old Javanese word *de*: '1. action, way (of acting); condition, cause, reason; 2. by (through the agency of); 3. in relation to, with regard to, toward, to'. Ogloblin (2000:179) describes it somewhat differently: '1. action; cause (= a noun); 2. actor preposition; 3. a nominalising auxiliary word introducing a predicative group (in analogy with modern Javanese *ənggon(-é)* and *olèh(-é)*)'.[11] Below are sample sentences to demonstrate these meanings. They clearly demonstrate the broad semantic and syntactic scope of *de*, to the point of causing translational dilemmas in some cases. (Especially the interpretation of *de* as actor or cause, or as nominaliser or preposition, is ambivalent).

[11] For these nominalising auxiliaries, see further §4.

3.1 *de* as a noun meaning 'action; way, manner'[12]

(1) *a-wəlas i de nṛpa Hāstinendra*
IV-have.pity PREP action;way,manner prince H
'feeling sorry for the actions of Prince Hāstinendra'
(Bhāratayuddha 48.7, Zoetmulder 1982)[13]

(2) *lālitya ri-ng de*
graceful PREP-ART action;way,manner
('graceful in action' [of dancers] =) 'moving gracefully'
(Kuñjarakarṇa Dharmakathana 146; 33.2b; Ogloblin 2000:179)

(3) *ang-ingət-ingət-a de ni-ng l<um>awan-a*
ACT-observe-RDP-IRR action;way,manner GEN-ART <ACT>fight-IRR
'observe the way of fighting' (Bhāratayuddha 36.1, Zoetmulder 1982)

3.2 *de* as a noun meaning 'cause, reason'

(4) *prayatna wruh ri-ng de*
intent.on to.know PREP-ART cause, reason
'intent on knowing the cause/reason' (Rāmāyana 3.82, Zoetmulder 1982)

(5) *de-nira ya ta h<um>urip sira*
cause,reason-3.POSS that.is PTL <ACT>live 3
'his reason (to do so) was to stay alive' (Sarkar (1971) 1:125(3), 255; Ogloblin 2000:180)

3.3 *de* as a preposition

As a preposition, *de* can introduce Actor as well as several other roles. In many of these cases, *de* is part of a compound preposition and is combined with a following genitive preposition *ni* and an article *-ng*, which syntactically belongs to the following noun.

3.3.1 *de* as an actor preposition

This use of *de* is seen in passive sentences. In Old Javanese, passive verbs are derived with the infix *<in>*, putting emphasis on the act itself or on the actor, or they are derived with the prefix *ka-*, emphasising result. Both can occur with *de* as actor preposition.

(6) *Mogha tâku malupa ri kita,*
mogha [ta -aku] ma-lupa ri kita
so it happened -1 STAT-forget PREP 2

---

[12] In the sample sentences that I present in the following pages I make use of the following gloss abbreviations: ACT – active; ADV – adversative; APPL – applicative; ART – article; DIST – distal deictic; GEN – genitive); HON – honorific; IRR – irrealis[12]; IV – intransitive; N – nasal prefix; NOM – nominaliser; PASS – passive; PERF – perfect; POSS – possessive; PREP – preposition; PTL – particle; RDP – reduplication; REL – relative marker; STAT – stative; 1, 2, 3 – first, second, third, person; (note that Javanese personal pronouns do not distinguish number). Note that in some of the sample sentences I do not follow Zoetmulder's morphological parsing (or lack thereof). I also use 'irrealis' instead of Zoetmulder's 'arealis'.

[13] Sources of sample sentences are indicated between brackets behind their translation. For Old Javanese, I follow Zoetmulder (1982, 1983) and Ogloblin (2000) in also indicating the primary source text of these sentences.

tan wruh yan ka-pangan de-ngku
not know that PASS-eat PREP-1POSS
'and so it happened that I forgot about you; I did not know that I had eaten you' (Adiparwa 75; Zoetmulder 1982:239)

(7) *nghulun pwa k<in>on de sang Dewayânî*
1 EMPH <PASS>order PREP HON D.

tamolah [tan N-polah] hana ngke
not ACT-active be here
'My father [= *sang D.*] ordered me to stay here' (Adiparwa 81, Zoetmulder 1983:105)

(8) *Tan dadi ka-ton de ni-ng wwang campur*
not occur PASS-see PREP GEN-ART person mixed
'It is not possible that she can be seen by an impure person' (Adiparwa 13; Zoetmulder 1982:52)

3.3.2 *de* as a causal preposition

Sometimes *de* introduces the cause or reason of an action (corresponding to English 'because of' or 'by').

(9) *ka-panas-an ta sang Uttangka de-nya*
ADV-to.heat,anger-ADV PTL HON Uttangka PREP-3POSS
[Uttangka was very angry because of him =] 'He drove Uttangka mad' (Adiparwa 15, Zoetmulder 1983:52)

(10) *sâkṣât rwan gumantung kakingan de ning lahrû*
sâkṣât rwan g<um>antung ka-king-an de ni-ng lahrû
like leaf <ACT>hang ADV-dry-ADV PREP GEN-ART draught
'like a leaf hanging down, shriveled by the draught' (Adiparwa 25, Zoetmulder 1983:59)

(11) *ndâtan suka sang hyang pitara de nira*
not happy HON HON forefathers PREP 3POSS
[his ancestors were not happy because of him =] 'his ancestors were not happy about him' (Adiparwa 3, Zoetmulder 1983:51)

3.3.3 *de* as a preposition expressing a relationship involving kinship or social hierarchy

(12) *putu de bhagawân Wasiṣtha*
grandson PREP lord W.
'grandson of Wasiṣtha' (Adiparwa 62, Zoetmulder 1983:56)

(13) *mangké tembé yan ibu-ngku mâri hulun-a de-nta*
this; now beginning that mother-1POSS stop be.slave-IRR PREP-2POSS
'from this moment onwards my mother stops being a slave of yours' (Adiparwa 45, Zoetmulder 1983:56)

3.4 *de* as a nominaliser

de can also nominalise verbs, in which case it is followed by a possessive pronoun.

(14) *de-nta* *bhakti* *ri-ng* *guru*
NOM-2POSS dedicated PREP-ART teacher
[your being dedicated to your teacher =] 'Your dedication to your teacher'
(Adiparwa 16, Zoetmulder 1983:55)

(15) *Tan lupa de nira r<um>akṣa ika-ng rât*
not forget NOM 3POSS <ACT>protect DIST-ART world
[he did not forget his protecting of the world =] 'he did not forget
(his obligation/task/intention) to protect the world' (Adiparwa 95,
Zoetmulder 1983:55)

3.5 Derivations based on *de*

There are several derivations that have *de* as a root. Some of these (*de-n-ing*, *day-a*) have become lexicalised and are now separate lexemes in modern Javanese.

de-n-ing has a function that is comparable to that of the preposition *de* 'because of; by'.[14] It means 'because (of the fact that); by', and it is a lexicalisation of < *de* 'cause' + *ni* (genitive preposition) + *-ng* (article to the following noun). It has become a regular actor preposition in modern Javanese (see §4).

(16) *<in>amər* *de* *ni-ng* *watək* *dewata*
<PASS>caress PREP GEN-ART class,category god
'[they were] caressed by the gods' (Arjunawiwāha 73;I.8; Ogloblin 2000:183)

(17) *Suka de ni-ng harəp ni nghulun anugrahâna*
glad PREP GEN-ART wish GEN 1 [in-anugraha-an-a]
PASS-receive-APPL-IRR
'We are glad that we will receive a favour' (Adiparwa 193, Zoetmulder 1983:151)

deya and its variant forms *daya* and *dāya* are basically irrealis forms of *de* rendered as '(future) act, way of acting; plan, what one is going to do' in Zoetmulder (1983).[15]

(18) *kepwan* *ing* *dāya* *yogya*
ka-ipu-an i-ng day-a yogya
ADV-perplexed,desperate-ADV PREP-ART action-IRR suitable
'at a loss as to what to do' (Hariwangśa 14.7; Zoetmulder 1982)

Note, incidentally, that the penultimate *a* in *daya* is most probably the original vowel. The *e* in *de* must be due to monophthongisation of *ay or contraction of *a + *i, as in inherited Old Javanese lexicon, midvowels are the result of the contraction of *a with a preceding or following high vowel. Hence it is likely that *de* had a phonologically more conservative variant form *day or *dai.

Consider also modern Javanese *daya* 'force, energy; influence'. This word must ultimately be a lexicalised form of the Old Javanese derivation **day-a*.[16] It was also

---

[14] Or, at least, this is the way they are translated in reference works and text translations. The exact meaning and use of prepositions and other grammatical words in Old Javanese still need to be sorted out.

[15] Ogloblin (2000:180) gives the additional translations 'duty' and 'influence, impact' or 'cause'.

adopted in Malay as *daya* ‘power, energy, capacity’ and ‘trick, ruse’; compare also *məmpərdaya* ‘to deceive, trick’, *daya-upaya* ‘all possible efforts’ and *apa daya* ‘what is to be done?’ (Echols and Shadily 1989).

Ogloblin also draws attention to the use of *de* as a verbal stem ‘to cause; to urge’ (Ogloblin 2000:180), as in the following sentences:

(19) *ḍuh nohan-ku kaka-nta cuṇḍuk i kita dunung*
gosh happy-1POSS brother-1POSS meet PREP 2 to.come
kadi d<in>e
as.if <PASS>urge
‘Ah, how happy I am, brother, to meet you who arrived here as if being urged’
(Arjunawiwāha 97; XVIII.1a; Ogloblin 2000:180)

(20) *yân-de [ya an-de] sangśaya-ni hyang Indra*
3 ACT-cause doubt-GEN HON I.
‘This caused the divine Indra’s doubt’ (Arjunawiwāha 72; I.6; Ogloblin 2000:180)

den- and *depun*(-) are typically Middle Javanese derivations of *de*; they are passive prefixes which often occur in imperative constructions (Ogloblin 2008:42). *depun* is basically a prepositional phrase consisting of *de* and *pun*. The latter is originally a nominal determiner used before names (‘mister’) or in references to people with a certain craft or skill (Indonesian *tukang*), which had developed into a 3rd person pronoun in Middle Javanese (cf. Ogloblin 2008:41; Zoetmulder 1982:11, 1983).

(21) *sapa kang depun [de-pun] tangis-i*
who REL PASS-3 cry-APPL
‘who was (s)he crying over’ (Kidung Harsa-Wijaya 5.141b; Zoetmulder 1982)

4 Reflexes of *de* in Modern Javanese

In modern Javanese, *de* has become the passive prefix *di-*. The structural developments in Javanese are much easier to follow than in Malay. Old Javanese had the default passive affixes *<in>* and *ka-*. These have survived in modern Javanese as archaic (bookish) and unproductive passive markers and are largely superseded by *di-*, which, in turn, does not occur in Old Javanese and is clearly an innovation. As in Malay, this prefix is ill adapted to the overall word structure of Javanese because it has not undergone antepenultimate vowel neutralisation, which is also a very regular sound change in modern Javanese.

As an archaic form, *de* still exists as a separate word meaning ‘vis-à-vis’ and in various derivations.[17] Pigeaud (1938) lists *de-* as a regional variant form of the general passive marker *di-* (although present-day speakers of Javanese do not always recognise this form). *dèn-* also occurs in modern Javanese but is labelled as bookish in Pigeaud (1938).

The vowel heightening in *di-* can be accounted for. Pigeaud (1938:vii) pointed out that Javanese *i*, *é* and *è* are virtually interchangeable in traditional spelling. A variation between *e* and *i* is also seen in Middle Javanese *depun* (discussed above) and the modern high

16 However, compare also Sanskrit *udaya* ‘rise, success’ proposed by Casparis (1997:16) as a possible but doubtful origin.

17 There are various entries in Pigeaud’s dictionary that can be analysed as derivations based on *de* (including *denta*, *dengku*, *déné*), but they are not immediately relevant for the present discussion.

Javanese passive marker *dipun*/*dipun-* (spelled both as a free form and a prefix in written texts).

dening survived as a monomorphemic preposition marking actor and cause. It has become a more formal and literary counterpart of the usual actor markers *karo* 'with; by' and *ing* 'in, at; by'. In some western dialects of Javanese, however, a cognate of *dening* has survived as a multipurpose preposition, which is also the default actor marker, compare Cirebon Javanese *ning* 'by; to; in, at; in relation to, concerning' (Ewing 2005). This *ning* has an even wider application in Banten Javanese, where it is described as 'in, inside; at; into, towards; from; with; if; by (actor marker); concerning; facing, in front of' (Munadi Patmadiwiria 1977).

The Old Javanese nominaliser *de* has not survived in modern (standard) Javanese. The latter has several equivalent nominalisers, but these are formally different. The most frequent one is the low register ('ngoko') form *òlèh-* (originally 'to obtain'), which is usually shortened to *lé* in spoken language. Such nominalisers are followed by a possessive pronoun (yielding *òlèh-ku, òlèh-mu, òlèh-é*, for respectively first, second, or third person actor). Interestingly, *òlèh-* is related to the Malay actor preposition *oleh*: as an actor preposition in Malay, and a nominaliser in modern Javanese, it assumes the same functions as *de*, but only cross-linguistically so. A somewhat less frequent nominaliser is *anggon-* or *ənggon-* (originally meaning 'place, location'). Examples of *òlèh-*:

(22) *lèh-ku* *m-laku* *pirang-pirang* *kilo*
NOM-1POSS ACT-walk countless kilometre
[my walking has been for miles and miles =] 'I've been walking for miles on end' (Elisabeth Riharti pers. comm.)

(23) *lèh-mu* *arəp* *m-ulih* *kapan?*
NOM-2POSS FUT ACT-come.back when
[your intended coming back is when =] 'When will you come back?
(Elisabeth Riharti pers. comm.)

Modern Javanese nominal constructions mainly involve active verbs; however, they can also have adversative verbs[18] in their scope, as in the following sample sentences:

(24) *Òlèh-é* *kodanan* *[kə-udan-an]* *sə-wəngi* *natas*
NOM-3POSS ADV-to.rain-ADV one,entire-night entire
'he was being rained on all night' (Elisabeth Riharti pers. comm.)

(25) *lé* *kə-maling-an* *nganti* *pitungpuluh* *yuta*
NOM ADV-steal- ADV amount.to 70 million
'they took as much as 70 million rupiah from him' (Elisabeth Riharti pers. comm.)

5 Discussion

The modern Javanese passive prefix *di-* can be traced to Old Javanese *de* 'cause, reason' and 'action; way, manner', a word which is also at the source of modern Javanese *dening*, a multipurpose preposition frequently used to introduce actor. The connection between modern Javanese *di-* and Old Javanese *de* is further evidenced by Middle Javanese *dèn-* and regional modern Javanese *de-*, both passive prefixes. The vowel change in *di-* can be accounted for (§4).

[18] Also called 'accidental passive' verbs.

How does this relate to the general typology of passive formation?

The development of Javanese *di-* from a noun meaning 'cause, reason' and 'action; way, manner' does not immediately fit in with any of the grammaticalisation patterns leading to the creation of passive voice formulated by Haspelmath (1990:54). These patterns are (1) passives derived from inactive auxiliaries, (2) passives derived from (nominal or pronominal) reflexives, (3) passives derived from causatives, and (4) passives derived from generalised (nominal or pronominal) subject constructions. It also does not seem to fit into the grammaticalization patterns for passive formation listed in Heine and Kuteva (2003).

However, there is a connection. Along with an inactive auxiliary, the construction involved in Haspelmath's first pattern also involves a 'resultative participle'. The latter is more generally known as a 'passive participle', but Haspelmath (p.40) is quite explicit in his preference for the term 'resultative participle', because, he argues, these participles do not have to be passive in meaning. For instance, they also include participles like 'escaped' in 'the escaped prisoner'.

The verbs in the nominal constructions on the basis of Old Javanese *de* are by no means identical with these participles, but they are somehow reminiscent of them in the sense that they are also foregrounded at the expense of the actor, which is backgrounded. It is possible that *di-* passives in Javanese evolved from these *de* nominalisations. The verbs in the latter are usually active, although they can also be adversatives derived with *k(a,ə)- -an*, which are more passive in nature. The verb-foregrounding and actor-backgrounding involved in them may have paved the way for a passive interpretation.

Such a development would also explain why *de* as a nominaliser has been replaced by other markers in modern Javanese (*òlèh-* etc.): the semantic bleaching of *de*, and its gradual identification with the more general (syntactic) category of passive marker, made it less useful for other applications and created the need for a new, less ambivalent, clause nominaliser.

A remaining question in this assumed pathway is that *de* as a nominaliser is always combined with a pronominal suffix, whereas *di-* is not. An exception to this is *dipun-*/ *depun-*, where *pun* was once a 3rd person pronoun, suggesting that the present syntactic properties of *di-* are a fairly recent development. The matter needs further investigation.

Malay *di-* was first attested on an Old Malay inscription found on Javanese soil (in Gandasuli village, Kedu, Central Java) and dated 832 AD (Casparis 1950:50–57). Moreover, the history of *di-* leaves no clear trail in Malay, whereas it can be traced to a lexical source in the history of Javanese. On the basis of these factors one could argue that Malay *di-* was originally borrowed from Javanese, and that this happened some time in the 9th century AD. This would pose a chronological problem: *di-* does not appear in Old Javanese but belongs to modern Javanese, which is only recorded from the 16th century onwards. It replaced Old Javanese, which had been a written medium in Java from the 9th until the 16th century (Ogloblin 2005:590–591).[19] The problem could of course be discarded by arguing that modern Javanese already came into existence long before the 16th century. Its sudden emergence in the 16th century as a written language, which was very different from Old Javanese, leaves little doubt that it had already begun to develop into a separate spoken dialect at a much earlier stage. The process may conceivably have

[19] This take over should be seen in terms of a written medium: the periods attributed to these languages tell us little about when modern Javanese came into being as a spoken language. Note also that Old Javanese continues to be used as a liturgical language in Bali (Ogloblin 2005:590).

begun seven centuries before it replaced Old Javanese as a written standard (that is, in the 9th century AD), or even longer ago. If it had, one could still argue that Malay *di-* was borrowed from this burgeoning form of modern Javanese. However, seven centuries is a long time, and while borrowing from modern Javanese is not entirely impossible, the arguments are speculative and bring along their own chronological problems.

It is more likely that Malay *di-* developed from *di*, given that the latter already occurred in 7th century Old Malay as a multipurpose preposition (Adelaar 2005a) and that there are parallel developments from a preposition to a passive prefix in Minangkabau, Kanayatn, Belangin, Sinama Bajau, Madurese and Middle Javanese[20] (§2.1). These frequently observed parallel developments in West Indonesian languages clearly offset typological objections against the development from a preposition to a passive prefix.

Both Malay and Javanese developed a passive prefix *di-*, but the ways in which they did were very different. Nevertheless, it is quite possible that the similar outcome of these developments was partly a consequence of the influence that Malay and Javanese have had on one another and which started more than a millennium ago.

One could, of course, go back one step further in time and argue that Old Malay *di* (that is, the *preposition*) was borrowed and reflects Old Javanese *de*, but this is somewhat countered by the fact that the 7th century inscriptions show no clear evidence of Javanese influence (in contrast to later forms of Malay). Furthermore, Dempwolff (1938) reconstructed a Proto Malayo-Polynesian locative preposition *di (> Malay *di*, Old Javanese *ri*, Bug *ri*), and there is as yet no compelling reason to doubt that Malay *di* is its reflex.

6 Conclusion

The lexical and morphosyntactic evidence provided by Zoetmulder and Ogloblin clearly suggest that the modern Javanese passive prefix *di-* ultimately developed from an earlier word *de* (< *day/*dai) basically meaning ‘cause, reason; action, way, manner’. This word was still extant in Old Javanese, where it was also used as a preposition, a nominaliser, and a verb. *di-* most likely developed from Old Javanese *de* in its function as a nominaliser. The shift from nominaliser to passive marker must have created a structural gap, which would explain why modern Javanese replaced *de* as a nominaliser with other words, such as *òlèh-(é)* and *ənggon-(é)*.

Old Javanese *de* is also at the origin of modern Javanese *déning*, a compound preposition regularly (but not exclusively) used to introduce the actor, and other words such as *déné* ‘regarding’ and *daya* ‘force, energy; influence’ (> Malay *daya* ‘power, energy, capacity’ and ‘trick, ruse’).

Given the longstanding mutual influence between Javanese and Malay, it is likely that this influence enforced the development of identical passive *di-* prefixes in both languages. However, these prefixes have very different origins in each of these languages.

---

[20] A language used from the 14th to late 17th century and more or less intermediate between Old and modern Javanese. It was generally used in Bali for the production of Kidung poetry and hence shows very little influence from Islam; however, as Ogloblin points out, these are no absolute criteria for its definition (Ogloblin 2008).

References

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

—— 1992b. The relevance of Salako for Proto-Malayic and for Old Malay epigraphy. *Bijdragen tot de Taal-, Land- en Volkenkunde* 148:382–408.

—— 2002. Salako morphology and the interrelation between voice, mood and aspect. In K.A. Adelaar and R.A. Blust, eds *Between worlds: linguistic papers in memory of David John Prentice*, 1–27. Canberra: Pacific Linguistics.

—— 2005a. Much ado about *di-*. *Bijdragen tot de Taal-, Land en Volkenkunde* 161/1:127–142.

—— 2005b. *Salako or Badameà: sketch grammar, texts and lexicon of a Kanayatn dialect (West Kalimantan).* Wiesbaden: Harrassowitz.

—— 2008. On the classifiability of Malayic. In Yuri A. Lander and Alexander K. Ogloblin, eds *Language and text in the Austronesian world. Studies in honour of Ülo Sirk*. LINCOM Studies in Austronesian Linguistics 06. 11–22. Munich: Lincom Europa.

Aichele, Walther. 1942–43. Die altmalaiische Literatursprache und ihr Einfluss auf das Altjavanische. *Zeitschrift für Eingeborenensprachen* 33:37–66.

Akamine, Jun. 2005. Sama (Bajau). In A. Adelaar and N.P. Himmelmann, eds *The Austronesian languages of Asia and Madagascar*, 377–396. London: Routledge.

Brown, C.C. 1921. Perak Malay. *Papers on Malay subjects. Second Series (V).* Calcutta: The Baptist Mission Press.

Casparis, J.G. de. 1950. *Prasasti Indonesia. Inscripties uit de Çailendra-tijd*. Bandung: A.C. Nix.

—— 1956. *Prasasti Indonesia, II. Selected inscriptions from the 7th to the 9th Century A.D.* Bandung: Masa Baru.

Echols, John M. and Hassan Shadily. 1989. *An Indonesian–English dictionary*. (3rd ed.). Ithaca: Cornell UP.

Ewing, Michael C. 2005. *Grammar and inference in conversation. Identifying clause structure in spoken Javanese*. Amsterdam/Philadelphia: John Benjamins.

Haaksma, Rémy. 1933. *Inleiding tot de studie der vervoegde vormen in de Indonesische talen.* Leiden: Boekhandel en Drukkerij voorheen E.J. Brill.

Haspelmath, Martin. 1990. The grammaticalization of passive morphology. *Studies in Language* 14:25–72.

Heine, Bernd and Tania Kuteva. 2003. *World lexicon of grammaticalization*. Cambridge: CUP.

Hopper, Paul J. 1988. How ergative is Malay? In Richard McGinn, ed. *Studies in Austronesian linguistics*. Monographs in International Studies. Southeast Asia Series. No.76. 441–454. Athens: Center for International Studies. Ohio University.

Kern, Hendrik. 1889. Eenige imperatief-vormen van het Oudjavaansch. *Bijdragen tot de Taal-, Land- en Volkenkunde* 38 (5de volgreeks, deel IV); reprinted in H. Kern.*Verspreide Geschriften Achtste Deel*. Appendix [Appendix to 8th volume], 293–313. 's-Gravenhage: Martinus Nijhoff.

Mahdi, Waruno. 2005. Old Malay. In A. Adelaar and N.P. Himmelmann, eds *The Austronesian languages of Asia and Madagascar*, 182–201. London: Routledge.

Munadi Patmadiwiria. 1977. *Dialek Jawa Banten–Banten.* Jakarta: Pusat Pembinaan dan Pengembangan Bahasa.

Moussay, Gérard. 1981. *La langue minangkabau*. *Cahier d'Archipel* 14. Paris: Association Archipel.

Nothofer, Bernd. 1975. The reconstruction of Proto-Malayo-Javanic. *Verhandelingen van het Koninklijk Instituut voor Taal-, Land- en Volkenkunde* 73. 's-Gravenhage: Martinus Nijhoff.

Ogloblin, Alexander K. 2000. The Old Javanese word de, in Lokesh Chandra, ed. *Society and culture of South East Asia. Continuities and changes*. Sata-pitaka Series vol. 395. 179–190. New Delhi: International Academy of Indian/Aditya Prakashan.

—— 2005. Javanese. In A. Adelaar and N.P. Himmelmann, eds *The Austronesian languages of Asia and Madagascar*, 590–624. London: Routledge.

—— 2008. In search of Middle Javanese. In Yuri A. Lander and Alexander K. Ogloblin, eds *Language and text in the Austronesian world. Studies in Honour of Ülo Sirk*. LINCOM Studies in Austronesian Linguistics 06. 31–45. Munich: Lincom Europa.

Sarkar, H.B. 1971–72. *Corpus of the inscriptions of Java (up to 928 AD).* vols I–II. Calcutta: Firma K.L. Mukhopadhyay.

Shibatani, Masayoshi. 1985. Passives and related constructions. *Language* 61/4:821–848.

Swellengrebel, Jan Lodewijk. 1936. *Korawāçrama. Een Oud-Javaansch proza-geschrift, uitgegeven, vertaald en toegelicht.* Santpoort: N.V. Uitgeverij vh. C.A. Mees.

Teeuw, Andries. 1959. The history of the Malay language. *Bijdragen tot de Taal-, Land- en Volkenkunde* 115:138–156.

van den Berg, René. 2004. Some notes on the origin of Malay di-. *Bijdragen tot de Taal-, Land- en Volkenkunde* 160:532–554.

van Ophuysen, C.A. 1903. Maleische Taalstudiën I. *Tijdschrift voor Indische Taal-, Land- en Volkenkunde* 46.

Wolff, John U. 2001. The origin of the Malay affixation di-nya. Paper presented at the Fifth International Symposium on Malay/Indonesian Linguistics. Leipzig, 16–17 June.

Zoetmulder, P.J. 1982. *Old Javanese — English dictionary* (2 Parts). KITLV publication, The Hague: Nijhoff.

—— 1983. *De taal van het Adiparwa. Een grammaticale studie van het Oudjavaans.* Dordrecht (Neth.): Foris Publications.

12 *Relative-clause bracketing in Oceanic languages around the Huon Gulf of New Guinea*

JOEL BRADSHAW

1 Introduction[1]

The Oceanic languages around the Huon Gulf of New Guinea are among the few known languages to mark their relative clauses at both ends. There are two more pockets of such languages in central Africa, in two different language families (Hagège 1976:198–201), but relative-clause bracketing is otherwise exceedingly rare in any major language family.[2] For some time after the publication of Sankoff and Brown (1976), it appeared that Tok Pisin was going to join this typologically exclusive club of languages that bracket their relative clauses at both ends. However, subsequent research (see Bradshaw 2007) has cast doubt on both the extent of such bracketing in Tok Pisin and the putative functional parallels between the brackets in Tok Pisin and those in the substrate Oceanic languages around the Huon Gulf, where the Tok Pisin bracketing was first noted. (Sankoff and Brown appealed to universal

[1] My first-ever contribution to a Festschrift (Bradshaw 1982b) was at the invitation of Bob Blust when he was at the University of Leiden. My choice of topic in that instance was inspired by an article by the linguist then being honored (Kähler 1974). This contribution in honor of Bob Blust adds to the coverage in both works. It documents more Austronesian languages that mark their relative clauses like Numbami. And it supplies data from more up-to-date sources than were available when I documented a sketchier sample in Chapter 4 of my never-published (1982a) dissertation. I would also like to thank Bob for regular encouragement over the years to stay involved in linguistics and also for hosting me on my way to two places that have most influenced my subsequent research in linguistics and reading of history, Papua New Guinea and Romania, respectively. Finally, I would like to thank Andy Pawley, Peter Lincoln, and an anonymous referee for comments that helped improve earlier drafts of this contribution.

[2] Hagège's examples include Moru and Mangbetu, two Central Sudanic languages within the Nilo-Saharan family, and Mbum and M'baka (= Ngbaka) in the Adamawa-Ubangi branch of the Niger-Congo family. I have recently learned from Raymond Boyd (pers. comm.) at CNRS that bracketed relatives are found throughout the Adamawa-Ubangi languages. In many of them, the same form—usually a demonstrative—occurs at each end; in others, the initial bracket might derive from an indefinite (like 'which'?) and the final from a locative.

Alexander Adelaar and Andrew Pawley, eds
Austronesian historical linguistics and culture history: a festschrift for Robert Blust, 143–161.
Pacific Linguistics, 2009.
© This edition vested with Pacific Linguistics.

discourse factors to explain the clause bracketing in Tok Pisin—and, by implication, in any other language with similar structures.) The data and analysis compiled below will clarify the functions of the relative-clause brackets, not just for the benefit of creolists investigating substrates, but also for syntactic typologists more generally.

It appears that no one has so far reconstructed relative-clause formations for Proto Oceanic. But several early surveys of relative clauses in various Oceanic languages suggest that the most common pattern is for the relative clause to follow its head and for the intervening marker, if any, to be either a demonstrative or a general subordinator, which we can represent schematically as N_{head} + DEM/REL + CLAUSE. See Sohn (1973) for a survey of Nuclear Micronesian languages; Kähler (1974) for some examples from Fijian, Tongan, and Samoan; and Pawley (1975) for a broader synthesis covering Eastern Oceanic languages more generally. None of these surveys found relative-clause markers at both ends, although Pawley (1975) focuses on the role played by pronominal elements as placeholders for the head noun within its subordinate clause. Such pronominal elements would typically occur in clause-final position only when the head noun functions in a locative or oblique role within the subordinate clause.

Many more Oceanic languages have been documented since the 1970s, but bracketed relative clauses still appear to be a unique areal feature of the Oceanic languages around the periphery of the Huon Gulf. They are well attested in members of the four major subgroups (Ross 1988) in the area: North Huon Gulf (Jabêm, Bukawa), South Huon Gulf (Mapos Buang, Patep, Iwal), Numbami (an isolate), and the Markham languages. There is also a bit of evidence that bracketed relative clauses show up in Sio, a language on the northeast tip of the Huon Peninsula, not far from Jabêm territory and inside Morobe Province geographically, but not part of the Huon Gulf family. In most of these languages, the clause-final bracket is usually demonstrative and anaphoric in origin, and may also mark the ends of other types of subordinate clauses (such as those specifying time, purpose, or condition), while the clause-initial bracket is either another demonstrative or some general subordinator.

Unfortunately, the actual forms used in the various languages are so disparate in shape that it seems impossible to reconstruct any morphemes unique to relative-clause bracketing for those languages that share this syntactic innovation. See, for example, Holzknecht's (1989:139) table of the demonstratives used to bracket relative clauses—at one end or the other or both—in the well-established Markham subgroup. The most we can do for the rest of the Oceanic languages in Morobe Province is to identify the most common etymological sources of the various clause markers, but this tells us no more than the independently reconstructable shapes of those demonstratives, question words, or general subordinators. Moreover, purely syntactic innovations like clause bracketing diffuse quite easily by language contact and thus form a poor basis for subgrouping; and new morphological functions that arise by grammaticalisation of shared etyma in innovative structures are similarly suspect (see Bradshaw 1985). The bracketing of relative clauses in the Oceanic languages of Morobe Province thus seems to me to be an areal feature spread by diffusion across language and subgroup boundaries, and not a very reliable genetic marker of shared innovation.

Although relative clauses bracketed at both ends are typologically unusual, the sources of each set of brackets are not at all surprising if one considers the mixed word-order typologies of Oceanic languages on the New Guinea mainland and adjacent offshore islands (Bradshaw 1982a; Capell 1976). Oceanic languages elsewhere are very strongly VO in their word-order typology, regardless of whether their basic word order is SVO, VSO, or VOS. They are prepositional and tend to place genitive, adjectival, and clausal modifiers after the nouns

they modify. The Oceanic languages on the New Guinea mainland, by contrast, are a mixed bag. Many have SOV basic word order, like most of the neighboring Papuan languages. All have preposed genitives but postposed adjectives, again like their Papuan neighbors. And many have innovative postpositions (and ambipositions) as well as prepositions. Even the OV languages, however, keep relative clauses after the nouns they modify, except those of Central Papua, which have developed fully preposed relative clauses (Bradshaw 1982a).

Clause subordinators in clause-initial position are common in VO languages. When head nouns precede their relative clauses, the markers tend to separate the head noun from the clause, and they often derive from either question words or demonstratives (Wh-forms or Th-forms). Among the Oceanic languages on the New Guinea mainland, however, both the more typologically conservative VO languages and the more innovative OV languages tend to mark the ends of subordinate clauses, and even of NPs more generally, often with formants derived from pronouns or demonstratives. Relative clauses bracketed at both ends appear to be a result of the intersection of these two typological tendencies, as the data compiled here will demonstrate.

2 North Huon Gulf languages

2.1 Jabêm

Jabêm is the best-recorded of the North Huon Gulf languages. It has four full sets of demonstratives, each containing forms correlated to 1st, 2nd, and 3rd person (Dempwolff 1939:30–31; Bradshaw and Czobor 2005:34–37). The meaning distinctions associated with the three persons are as follows.

- 1st person: ‘that which is near me, which concerns me, or which I know; what follows, what I am going to say’
- 2nd person: ‘that which is near you, which concerns you, or which you know; what precedes, what you have just said’
- 3rd person: ‘that which is neither by you nor by me, which neither you nor I know, which concerns neither you nor me’

The semantic distinctions among the four morphological sets are somewhat harder to pin down. Dempwolff says the forms beginning with **t** mark more definite or certain *(bestimmtere)* referents, while the others have a somewhat indefinite or vague coloration. In several examples, the latter class receive glosses like ‘far from us, not visible’. If similar formal distinctions in Numbami, another Huon Gulf language, are truly analogous, then it is likely that the Jabêm *t-* forms, like the Numbami *ta-* forms, indicate stronger deixis and contrast, while the Jabêm *n*-based forms, like Numbami *na*, are more strictly anaphoric.[3]

---

[3] In morpheme glosses, pronominal person is indicated by numerals (1, 2, 3), and number by S for singular, DU for dual, P for plural, PI for plural inclusive, and PX for plural exclusive. Clause brackets of demonstrative origin are usually glossed DEM, while those of other origins are usually glossed REL. Other pronominal glosses include GEN – genitive, OBLPRO – oblique pronoun, PSU – phrase summary, RPRO – resumptive pro-form, REF – referential noun marker, RPL – plural relative pronoun, RSG – singular relative pronoun. Tense/aspect/mood glosses for verbs include CONT – continuative, DUR – durative, FIN – completive, FUT – future, IRR – irrealis, POT – potential mode, PTPL – participial, R – realis, and RDP – reduplicative. In a few cases I have omitted lengthy TAM glosses as immaterial to the current discussion. Other glosses include ADV – adverb marker, COND – conditional, ERG – ergative marker, LOC – locative marker, SAY – complementiser derived from ‘to say’, SUBJ – subject marker, and TR – transitiviser.

JABÊM demonstratives (Dempwolff 1939)

(1)

Deictic (*t*-based)		Definite (*n*-based)		
Long	Short	Long	Short	
tonec	*tec*	*onec*	*nec*	1st person
tonaŋ	*taŋ*	*onaŋ/ônaŋ*	*naŋ*	2nd person
tonê	*tê*	*ônê*	*nê*	3rd person

In order to mark relative clauses, Jabêm utilises the short, *t*-based set in clause-initial position and the short, *n*-based set in clause-final position. In some cases, *naŋ* may be used as a clause-initial marker. However, the short *t*-forms never occur as clause-final brackets (Dempwolff 1939:87, Bradshaw and Czobor 2005:109). (In Jabêm orthography, *ê* and *ô* are upper-mid vowels relative to lower-mid *e* and *o*, while *-c* marks a final glottal stop, and the palatal glide is written *j*.)

JABÊM relative-clause brackets (Dempwolff 1939)

(2)

tec	*... nec*	1st person
taŋ/naŋ	*... naŋ*	2nd person
tê	*... nê*	3rd person

These brackets are used independently of the case relationship of the head noun within the matrix sentence or the coreferential NP in the relative clause. Within the embedded clause, the coreferential NP may be represented as a pronoun or deleted. The choice of 1st, 2nd, or 3rd person forms as brackets depends on whether the information in the clause is associated with the speaker, hearer, or other, as outlined in the beginning of this section.The clause-final bracket is often left off, especially if the matrix sentence terminates at the end of the subordinate clause (Dempwolff 1939:87; Bradshaw and Czobor 2005:109).

JABÊM relative clauses (Bradshaw and Czobor 2005)

(3) a. *na-ôpi* *nip* *[tê kêkô nê]* *ôna*
go-2S-climb coconut DEM 3S-stand DEM 2S-go
'go climb that coconut tree standing over there'

b. *lip* *[tec aê gawa nec]* *gêjac mocseŋ teŋ*
trap DEM 1S 1S-set DEM 3S-catch bushfowl one
'the trap I set caught a bushfowl'

c. *lau* *[taŋ sêwiŋ aê atom naŋ]* *seseŋ aê su*
people DEM 3P-join 1S not DEM 3P-oppose 1S FIN
'people who are not with me are against me'

d. *lau* *[naŋ sêlac sêja Kêla naŋ]* *sêmu* *sêmêŋ me masi*
people DEM 3P-sail 3P-go K. DEM 3P-return 3P-come or none
'have the people who sailed to Kela come back?'

e. *bôc tonaŋ* *bulumakao* *[taŋ bômbôm* *sêpip nê su*
pig that cow DEM whites 3P-squeeze GEN3S milk/udder

ma *sênôm naŋ]*
and.then 3P-drink DEM
'that animal is a cow, whose udder/milk the whites squeeze (out) and drink'

f. *aê Jesu* *[tec kôjanda aê]*
1S J. DEM 2S-persecute 1S
'I am Jesus, whom you persecute'

2.2 Bukawa

Eckermann (2007) presents a different analysis that sheds some light on the distinct roles of the initial and final brackets in the Bukawa dialects, which stretch all along the north coast of the Gulf and as far as the Salamaua Peninsula on the south coast. (In Bukawa orthography, as in Jabêm, *ê* and *ô* are upper-mid vowels relative to lower-mid *e* and *o*, and *-c* marks a final glottal stop. Unlike Jabêm, however, the palatal glide is written *y*, and there is even a voiceless palatal glide, written *yh*.)

BUKAWA demonstratives (Eckermann 2007)

(4)

dau	the/that (one) – demonstrative particle glossed as DEM
(din)dec	this one/these ones, here (close, now, recent)
(din)dê	that one/those ones, there (distant)
dinaŋ	that one/those ones (distant, past)
naŋ	glossed as GIV (given information)

The demonstrative *dau* functions as a kind of definite article, perhaps best translated 'the aforementioned' or 'the very one', and tends to mark nouns previously introduced with the indefinite article *daŋ* 'a/one' (2007:82). Each of the demonstratives has a long and short form: *dindec ~ dec, dindê ~ dê, dinaŋ ~ naŋ*. Relative clauses tend to be marked at both ends with the short form of the most distal demonstrative *naŋ*, although the final marker may be left off if the relative clause falls at the end of the matrix sentence (2007:111).

BUKAWA relative clauses (Eckermann 2007)

(5) a. *ŋamalac andô* *[naŋ kôm tasaŋ* *têŋ iŋ* *ma* *kôc iŋ sa*
person old DEM 3S-do lie 3S-go.to 3S and 3S-take 3S up/out

gi *hu* *iŋ siŋ* *naŋ]*
3S -go 3S -leave 3S divestingly DEM
'the old man who lied to her and took her out and left her'

b. *ŋac sauŋ ...* *gêlic gêŋ sambob* *[naŋ mboc kôm naŋ]*
man little 3S-see thing all DEM snake 3S-do DEM
'the small man saw everything that the snake did'

c. *balôm mboloc dinda* *[naŋ* *yac alic* *têŋ ôbwêc naŋ]*
ghost stupid mother DEM 1P 1PX-see 3S-go.to night DEM
'that great stupid ghost that we saw in the night'

d. *o lau* *bu-ŋga ŋayham* *[naŋ* *Pômdau kêyaliŋ mac sa]*
o people water-of good DEM Lord 3S-choose 2P up/out
'o good Christians whom the Lord has chosen'

3 South Huon Gulf languages

Relative clauses marked in both initial and final position are found in Mapos, a dialect of Buang proper; in Patep, a member of the Mumeng dialect chain within the Buang group; and in Iwal, the most conservative member of the South Huon Gulf group. There is little doubt that bracketed relatives are found throughout the other members of the group as well.

3.1 Mapos Buang

In Mapos Buang, the clause-initial marker is *sën* 'this', a member of the set of demonstratives. (In the Mapos Buang examples, *ë* is upper-mid relative to lower-mid *e*, while *ô* is lower-mid relative to upper-mid *o; kh* and *gh* are the uvular equivalents of velar *k* and *g*.) The clause-final marker is one of a set of abbreviated demonstratives or else the 'emphatic' particle *lo*, which seems to serve as a general clause-boundary marker (Hooley 1970:181).

MAPOS BUANG demonstratives (Hooley 1970)

(6)

sën(ë)	'this'
saga(sën)	'that, there'
sagi(sën)	'this'
sagu	'that, over there'

MAPOS BUANG relative-clause markers (Hooley 1970 and pers. comm.)

(7)

sën ... ë	'here'
lo	emphatic particle
agi	'here'
aga	'there'

As a straight demonstrative, *sën* may stand alone in postnominal position, but as a relative-marker, *sën* remains in clause-initial position when the clause is separated from its head noun, as in (8a).

MAPOS BUANG relative clauses (Hooley 1970 and pers. comm.)

(8) a. *hong re [sën ghesis sa bôôkh ë]*
2S who DEM 2S-hit 1S pig DEM
'who are you who've killed my pig?'

b. *beggang [sën depekwë monë lok lo]*
house DEM 3P-change money in.it REL
'a place for changing money'

3.2 Patep

In Patep, the clause-initial marker is invariant *wê*, a form of uncertain origin. The clause-final marker is usually *ge*, a proximal deictic ('here') which also marks the ends of topic phrases and conditional clauses. Occasionally, other deictics appear in place of *ge* in clause-final position.

PATEP relative clauses (Lauck 1976, 1980)

(9) a. *ông ob tyoo yii yuu nuhu [wê ob lam ge]*
2S POT dodge spear two arrow REL POT come DEM
'you will dodge the spears and arrows that will come'

b. *môp [wê he vông ge], he o xovô da ti lêm*
way REL 3P do DEM 3P NEG know Sunday one NEG
'about the customs they followed, they didn't know about Sunday'

c. *nu byac ngô [wê ma nêl ge] mêgêm i mi dô*
child daughter hear REL father say DEM so 3S HAB stay
'the daughter heard what her father said so she stayed'

PATEP topic marking (Lauck 1976, 1980)

(10) a. *ki-luvac ge] he lungên ên luda dô lec*
potato DEM 3P not.know because sand be on
'the sweet potatoes they couldn't find because there was sand on them'

b. *vi-dii-ên, tride ge], dɨ xe lec kal*
be.day-NOM W'day DEM then 1PX board car
'on Wednesday at midday we got on cars'

PATEP conditionals (Lauck 1976, 1980)

(11) a. *xomxo obêc tulec ge] dɨ ob hi*
person POT find COND then POT hit
'when a man would find them, he would kill them'

b. *ông loc vac nita ge], od ông viac ông nivi-ha*
2S go into grass COND then 2S care.for 2S good
'when you go into the bush, you watch out'

3.3 Iwal

Iwal (also known as Kaiwa) relative clauses are marked in final position with one of a set of demonstrative formants: *nik* 'near me', *nok* 'near you', and *ok* 'over there'. The last appears to be much more common than the other members of the set. A different set of forms occurs in clause-initial position: *ete* or *ebe*. The former is a demonstrative stem apparently cognate with *ta* in Numbami and *t-* in Jabêm. It combines with the demonstrative formants given above to produce a set of demonstrative modifiers: *etenik* 'near me', *etenok* 'near you', *etok* 'over there', and *ete ok* 'long way off'. The stem *ebe* combines with the distal demonstrative formant *ok* to produce *ebok* 'over there' and *ebe ok* 'long way off'.[4] The data available to me (compiled in Bradshaw 2001) do not make clear what differences, if any, exist between the functions of *ete* and *ebe* as clause-initial relativisers.

[4] Clause-initial *ebe* in Iwal may seem related to a putative complementiser derived from a verb *-be 'to say'. Compare the Jabêm all-purpose complementiser *gebe*, lit. 'it-says', and the conditional complementiser *èmbe*, lit. 'it will say'. However, the Iwal preposition and desiderative complementiser *ve* 'to, for, with; want to' is perhaps a better—or additional—reflex of *-be 'to say'. The all-purpose complementiser *iŋgo* in Numbami also derives transparently from the verb *-ŋgo* 'to say'. In both Numbami and Jabêm, the verb 'to say' serves as a desiderative when it takes a sentential complement in the future tense or irrealis mode.

Iwal time clauses are marked in a slightly different manner. The clause-initial marker is *galk* (compare *galkik* ‘earlier today’), and the clause-final marker is *ik* if the clause identifies a time close to the present and *ok* if the clause identifies a time further removed from the present.

IWAL relative clauses (Bradshaw 2001)

(12) a. *ei gi-as uvun [ebe gi-sweng tambok ok]*
3S 3S-hit dog DEM 3S-bark night DEM
‘he hit the dog that barked in the night’

b. *amol [ete ayeu ga-li ok] ei gi-ro mank*
man DEM 1S 1S-see DEM 3S 3S-stab bird
‘this man I saw speared a bird’

c. *au ga-li avie [etok amol etok gi-as ane uvun ok]*
1S 1S-see woman DEM man DEM 3S-hit GEN3S dog DEM
‘I saw the woman whose dog the man hit’

d. *amol [ebe ayeu ga-ab kulkul ve ane ok] Yakob*
man DEM 1S 1S-give work for GEN3S DEM Jacob
‘the man I worked for is Jacob’

e. *amolmol [ebe i-pasang ul ok] eisir i-mbweg Kui*
people DEM 3P-make pot DEM 3P 3P-stay K.
‘the people who make pots are from Kui’

4 Numbami

In Numbami, an isolate within the Huon Gulf subgroup, the clause-final marker is *na*, apparently related to the anaphoric demonstrative formant *na*. The shape of the clause-initial marker varies depending on the nature of the information in the clause. Relative-clause-marking does not vary according to the case role of the relativised NP. The Wh-form *manu* ‘which, where’ is by far the most common clause-initial form. It simply indicates that the referent being identified is specific. The Th-forms *ta(te)* ‘this, these, here’ and *ta(to)* ‘that, those, there’ (both usually reduced to *ta*, which I gloss DEM) indicate that the referent is definite—that is, known to the addressee. The complementiser *iŋgo* (from ‘3S-say’, which I gloss SAY in this function) when used as a relativiser signals that the referent is identified by the purpose to which it will be put. (Purposive relatives are marked for future tense.) Clause-initial *manu* and *iŋgo* contrast in time clauses. Clauses identifying specific, realis times are introduced by *manu*, while those identifying nonspecific or future times (or conditions under which something else may come about) are introduced by *iŋgo.*

The various clause-initial relativisers may co-occur. When they do so, they occur in the following order: *ta(te/to)* + *manu* + *iŋgo.* One or the other of the clause-initial markers is always present, but clause-final *na* is sometimes elided if the end of the clause is made clear by other means, such as by an intonation boundary.

NUMBAMI relative clauses with new heads (Bradshaw, fieldnotes)

(13) a. *lawa lauwa na [manu tima na], payama tima*
people fight GEN REL 3P-come DEM for.good 3P-come
‘the soldiers who had come, they came for good’

b. *te-mbi i lawa [manu ma-ma ma-ndo i-muŋga na]*
3P-take 1PX people REL 1PX-come 1PX-stay 3S-precede DEM
'they took those of us who had come first'

c. *ma-ki bani [manu ma-yaki na] su ulaŋa*
1PX-put food REL 1PX-pare REL into pot
'we put the food we've pared into the pot'

d. *sai ni-wese nomba [manu ti-yota waŋga na]*
who FUT3S-show thing REL 3P-tie canoe DEM
'who will show (us) what to tie the canoe with?'

e. *nomba lua [manu mu-seya woya na], eana e-tate*
thing two REL 2P-ask 1S DEM DEM 3S-here
'those two things you asked me about, here they are'

f. *tiki biŋa de lawa [manu aindi waŋga i-tatala na]*
3P-send word to people REL 3P-GENP canoe 3S-sink DEM
'they sent word to the people whose boat sank'

g. *wa ostrelya [manu tindo Salamaua], eana, ai tinzolo*
and Australia REL 3P-stay S. DEM 3P 3P-scatter
'and the Australians (not previously mentioned) who were at Salamaua, those guys, they took off'

NUMBAMI relative clauses with given heads (Bradshaw, fieldnotes)

(14) a. *kana lawa [ta ti-ndo ti-tabiŋa Salamaua], ai ti-pai kulakula na*
so people DEM 3P-stay 3P-be.near S. 3P 3P-do work DEM

i-muŋga
3P-precede
'so those people (previously mentioned) who were staying near Salamaua, they did this (cargo) work first'

b. *ewesika tiyama-ma [ta mami puta na], inami kulakula bamo ano-ma*
women all-ADV DEM 1PX-live earth DEM 1PXGEN work much true-ADV
'all of us women who live on this earth, we have a lot of work'

NUMBAMI purposive relatives (Bradshaw, fieldnotes)

(15) a. *ina-kalati sa [iŋgo ina-lalaŋi kundu na]*
3PFUT-fix place SAY 3PFUT-scorch sago DEM
'they'll fix a place to scorch the sago at'

b. *ma-ki [iŋgo ni-ye gaya wambana-ma na] iye susuna*
1PX-put SAY FUT3S-lie next.day morning-ADV DEM 3S-lie corner

wa [manu iŋgo mananisi na] ma-ki i-ye maina-ma
and REL SAY 1PXFUT-boil REL 1PX-put 3S-lie separate-ADV
'we put that (portion of the food) for the next morning into the corner, and that which we intend to cook we place separately'

NUMBAMI time and conditional clauses (Bradshaw, fieldnotes)

(16) a. *[manu bembena-ma i-ma teteu na], i-loŋoni biŋa Numbami kote*
REL at.first-ADV 3S-come village DEM 3S -hear talk N. not
'when he first came to the village, he didn't understand Numbami'

b. *[temi iŋgo na-leleu na-ma na], [iŋgo ta-zuzu na], aiya nu-kole*
time SAY FUT1S-return FUT1S-come DEM SAY 1PI-shove DEM 2S FUT2S-turn

nu-ŋgo biŋa deŋa woya kote.
FUT2S-say word to 1S not
'when I come back, when we shove off, don't turn around and say a word to me'

c. *[iŋgo ta-tala kundu tomu na], a kole lua mo toli*
SAY 1PI-chop sago broken DEM perhaps man two or three

ina-wasa ina-tala tomu
3PFUT-go 3PFUT-chop broken
'when we chop down a sago palm, perhaps two or three men will go chop it down'

5 Markham languages

'Bracketing of relative clauses with demonstratives is a common practice among the languages of the Markham' (Holzknecht 1989:138). However, some languages mark both ends and others seem to mark just one end. In South Watut, the demonstrative *tiŋga* can be split to mark both ends. Among the Lower Markham languages, Nafi marks the end rather than the beginning of the clause with the demonstrative *ŋgah*, while Labu marks the beginning rather than the end of the clause with the general relativiser *lake* (Siegel 1984:119).

SOUTH WATUT bracketed relative clause (Holzknecht 1989)

(17) *Jek i-ra jiya? ri naip a [ti ra-gin afu ŋga]*
Jack 3S-cut tree with knife DEM 1s-give DAT DEM
'Jack cut the tree with the knife which I gave him'

NAFI right-bracketed relative clause (Holzknecht 1989)

(18) *kafi [siwu-n ŋi-mbak ŋgah] ŋi-kapuŋ wom iŋgiŋg*
woman husband-3 3S-die DEM 3S-stay house only
'a woman whose husband has died only stays in the house'

LABU left-bracketed relative clauses (Siegel 1984)

(19) a. *ai yô-kôna hêna [lake sê-nda dusuku]*
1S 1S-see woman REL 3P-stay Labubutu
'I saw women who live in Labubutu'

b. *amêna ya ainialê [lake ya hono kô kakala]*
man 3S-hit boy REL 3S-hit theft of chicken
'the man hit the boy who stole the chicken'

c. *ini ainialô [lake ai yô-kona]*
he boy REL 1S 1S-see
'he's the boy whom I saw'

A greater variety of examples is available from Amari, a dialect of Adzera proper in the Upper Markham Valley, where the demonstrative *ugu* 'there, far away, near him or it, already seen or referred to' (Holzknecht 1980:68) both precedes and follows relative clauses. The first *ugu* is optional and, according to Holzknecht (1980:70–71), occurs in post-head noun, not clause-initial, position. (It is not clear what tests Holzknecht performed to determine the syntactic affiliation of the first *ugu*.) In these respects, relative-clause marking in Amari differs from that found in the Huon Gulf languages discussed above, wherein the clause-initial relativiser is obligatory and stays with the clause, rather than the head noun, when the clause is separated from its head. Another way in which Amari differs from the other two groups is that verbs in Amari relative clauses are marked as subordinate with the participial suffix *-(d)a.*

AMARI relative clauses (Holzknecht 1980)

(20) a. *ifab (ugu) [mus-a yup-a intap ugu] i-mamp sib*
pig DEM always-PTPL dig-PTPL earth DEM R-die FIN
'that pig that was always digging up the ground is dead'

b. *dzi na-yu ungar ugu [garam fawaʔ-a sib aga ugu]*
1S HORT-take house DEM man break-PTPL FIN DEM DEM
'I will take the house which the man has broken'

c. *dzantsun (ugu) [dzi rab-a gai gin ugu] i-tatiʔsib*
axe (DEM) 1S cut-PTPL tree OBLPRO DEM R-break FIN
'the axe I cut the tree with is broken'

d. *ungar (ugu) [dzi mus-a ginɡʔ-a gin ugu] dzaf i-ga sib*
house (DEM) 1S always-PTPL sleep-PTPL OBLPRO DEM fire R-eat FIN
'fire has burnt down the house I usually sleep in'

e. *garam (ugu) [dzi i-rim pas rut in ugu] i-yu i-fa taun*
man (DEM) 1S R-give letter go.with OBLPRO DEM R-take R-go town
'the man I gave the letter to has taken it to town'

f. *sagat (ugu) [dzi ni-da nan rut in ugu] i-fa sib gamp*
woman (DEM) 1s say-PTPL talk go.with OBLPRO DEM R-go Fin village
'the woman I spoke to went back to the village'

g. *mamaʔ marub ugu [dzi dzigin-da i gan nam-gan ugu]*
child male DEM 1S steal-PTPL OBL GEN3 food-GEN3 DEM

i-fa uta da ru-fa ginɡʔ
R-go nothing and CONT-go sleep
'the boy whose food I stole went without anything and went to sleep'

5 Bracketing outside the Huon Gulf

Bracketed relative clauses also show up in Sio (also spelled Siâ), a language on the northeast coast of the Huon Peninsula facing the Vitiaz Strait, not far from Jabêm territory and still inside Morobe Province geographically, but outside the Huon Gulf family genetically. I have no firm data on relative-clause marking in other Morobe members of the Vitiaz group.

Relative clauses in Sio are most commonly marked in clause-initial and/or clause-final position, generally in both positions, with one of two deictics, *ŋine* ‘this, near’ and *ŋinde* ‘that, far’ (Clark 1994; Dempwolff 1936). However, the clause-initial bracket can also be a general subordinator like *kala* (usu. ‘also’) or *ŋana*.

SIO relative clauses (Clark 1994, Dempwolff 1936)

(21) a. *nia [ŋine kinda ta-mo] ara*
place DEM 1PI 1PI-stay good
‘the place we’re staying is good’

b. *si-lâ si-kai kelekele [ŋine muŋga i-poro pa-nzi ŋinde]*
3P-go 3P-get things DEM before 3S-talk to-3P DEM
‘they went and got the things that he had told them about’

c. *ku-lâ pw-ai ŋgoa natu [na-pile i-keno luma ŋinde]*
2S-go 2S-get pig child 1S-leave 3S-stay house DEM
‘go get the young pig I left in the house’

d. *i-ma i-ka bole [kala ta-o ndue ŋine]*
3S-come 3S-eat banana REL 1PI-put there DEM
‘he will eat the bananas which we set down’

6 Languages with postposed clauses and markers in NP-final position

Austronesian languages with postposed relative clauses and markers only in NP-final position are widely distributed within Papua New Guinea. The position of the relative clause in such languages is not innovative. Even the absence of a clause-initial marker may not be innovative. It is not uncommon for Oceanic languages to lack clause-initial relativisers. However, the kinds of markers that turn up in final position in these New Guinea languages are almost certainly innovative.

6.1 Schouten languages

In Kairiru, an East Sepik Province language of the Schouten subgroup, relative clauses follow all other nominal modifiers in the NP (Wivell 1981:185). Although word order is somewhat variable, especially for oblique NPs, Kairiru is basically an OV language. It has no special markers for relative clauses. However, Kairiru does have what Wivell (1981:37) calls a ‘Phrase Summary (PSU)’. The phrase summary consists of one of the 3rd person focal pronouns in absolute NP-final position. It presumably has no other function than to announce the end of the noun phrase (or referential unit) and the resumption of the sentence of which the NP is a constituent. Focal pronouns functioning as phrase summaries can only occur after NPs that refer to humans (or possibly higher animals) (Wivell 1981:38). (Wivell does not say whether other elements, demonstratives for instance, can play a role similar to that of the phrase summary in NPs referring to nonhumans or non–higher animals.) Wivell also says (1981:38) that phrase summaries cannot occur after NPs in oblique case roles. However, this second restriction—including Wivell’s supporting example—could just as well be translated into a restriction against phrase summaries appearing in sentence-final position, since NPs in other than oblique case roles do not generally occur in that position.

KAIRIRU NP-final phrase summary (Wivell 1981)

(22) a. *ramat pur tuol rra-nguk ei a-ruong*
man pig three 3P-snort 3S(PSU) 3S-hear
'the man heard the three pigs snort'

b. *moin Rrinrrin ei o-ur wun nau a-q-i*
woman Rr. 3S(PSU) 3S-descend beach saltwater 3S-fetch-3S
'the woman Rrinrrin, she went down to the beach to fetch some saltwater'

c. *kyau rri ramat tuol rri w-un-rri*
1S 3P man three 3P(PSU) 1S-hit-3P
'I hit the three men'

d. *foyeq Masos rru miem Samen rru rri-lieq rryan Smolau*
great.grandparent M. 3DU mother(1S) S. 3DU(PSU) 3DU-go river S.
'great-grandfather Masos and my mother Samen went to the river at Smolau'

In Manam, a language of western Madang Province closely related to Kairiru, relative clauses are postposed and there are no special forms used to mark relative clauses alone. However, certain NPs and subordinate clauses tend to be marked in final position with a 'resumptive pro-form' (Lichtenberk 1983:452–461 and elsewhere). The shape of the resumptive pro-form (here glossed RPRO) matches that of the proximal demonstrative *ng(a)e* 'this' or *ng(a)e-di* 'these', except that the pro-form is frequently cliticised to the end of the constituent it marks off. The distal demonstrative *ngara* 'that' or *ngara-di* 'those' is never used as a resumptive pro-form. However, the resumptive pro-form may co-occur with either the proximal or distal demonstrative.

Among the kinds of constituents marked with the resumptive pro-form are topicalised NPs, time phrases and clauses, relative clauses, and conditionals.

MANAM resumptive pro-forms (Lichtenberk 1983)

After topicalised NPs:

(23) a. *zirápu ne-m nge-Ø] fred óno i-enéno*
mattress GEN2S RPRO-3S F. OBLPRO 3S-sleep.RDP
'as for your mattress, Fred sleeps on it'

After time expressions:

b. *u-múle-nge] píta ábe i-alále*
1S-return-RPRO P. already 3S-leave
'when I came back, Pita had already left'

After conditionals:

c. *ʔáti téʔe-Ø i-púra-nge] bogía n-láʔo*
boat one-3S 3S-come-RPRO B. 1S-go
'if the boat comes, I will (definitely) go to Bogia'

After relative clauses:

d. *tamóata [wabúbu-lo i-púra-nge] ísi i-éno*
man night-in 3S-come-RPRO still 3S-sleep
'the man who came during the night is still asleep'

e. *ʔai sápara-Ø* *sasalága* *[péra* *ʔána-Ø atábala-Ø*
tree branch-3S long(PL) house GEN3S overhead-3S

di-éno ngáe-di] *di-pólo-Ø*
3P-lie RPRO-3P 3P-cut-3P
'they cut off the long (tree) branches that were above his house'

6.2 Papuan Tip languages

In Are (aka Mukawa), a language of Milne Bay Province, relative clauses are postposed to head nouns. Resumptive pronouns signal the end of the clause (and NP) and the resumption of the matrix sentence. The resumptive pronoun in the Are examples also serves as a regular 3rd singular pronoun and as a demonstrative formant (*nikona* 'this', *nokona* 'that').

ARE postposed relative clauses (Paisawa, Pagotto, and Kale 1976)

(24) a. *sebare* *[rabirabi* *i-botu* *kona]* *au-poro i-boai-Ø*
man yesterday 3S-come PRO 1S-pig 3S-kill-3S
'the man who came yesterday killed my pig'

b. *sebare* *[au-poro i-boai* *kona]* *a-kinani-Ø*
man 1S-pig 3S-kill PRO 1S-see-3S
'I saw the man who killed my pig'

c. *sebare* *[au yove-ai* *i-daua* *kona]* *poro* *[kaire* *i-kani kona]*
man 1S house-LOC 3S-come PRO pig sweet.potato 3S-eat PRO

i-boai-Ø
3s-kill-3s
'the man who came to my house killed the pig that ate the sweet potato'

7 Languages with internally headed clauses

Other Papuan Tip languages have developed different strategies to mark relative clauses. In Tawala, the verbs of relative clauses are usually distinguished by being marked for durative aspect, even when such aspect marking is semantically inappropriate, and there is no other marker of subordination or dependence (Ezard 1997:207). Moreover, Tawala relative clauses are internally headed, so that a head noun and its subordinate clause otherwise resemble an independent clause. However, when the head and its clause are topicalised, it is followed by a topic marker, and the head noun is often repeated in the matrix clause. This provides a pathway toward the eventual development of a structure with a relative clause preceding its external head.

Relativisation strategies appear similar in Gumawana, another Papuan Tip language with OV word order. Clauses are internally headed and under certain conditions require marking for 'imperfective' aspect (Olson 1992:321). However, when the relative-clause head functions as the object of the matrix clause, the relative clause follows the matrix verb, yielding VO word order, as in (26b).

TAWALA internally headed relative clauses (Ezard 1997)

(25) a. *[kwasikwasi-na pom u-gima-gimal'-e-ya] i-tutuma*
machete-DEF yesterday 2S-DUR-buy-TR-3S 3S-blunt
'the bush-knife that you bought yesterday is blunt'

b. *[numa hi-wogo-wogo-hi naka] hi-lata duma*
house 3P-DUR-build-3P TOP 3P-grow very
'the houses they built are very big'

c. *[meyagai noka a-ga-gale-hi naka] meyagai dewadewa duma-na*
village there 1S-DUR-see-3P TOP village good very-3S
'the villages that I saw there are very good villages'

GUMAWANA internally headed relative clauses (Olson 1992)

(26) a. *Kai [Dan-ia-na mone-na i-dalaiba] ka-ma*
1PX Dan-REF-3S wife-3S 3S-drive 1PX-come
'we and Dan's wife, who drove, came'

b. *Kelebi i-modugu-di [kewou-ya-di si-salili]*
K. 3S-make-3P canoe-REF-3P 3P-sink
'Kelebi built the canoes that sank'

8 Languages with preposed clauses and markers in NP-final position

Austronesian languages with preposed relative clauses appear largely confined to Central Province, Papua New Guinea. Central Province languages show two word-order innovations with respect to relative clauses: postpositional marking of the modified NP and prenominal position of the modifying clause. The first innovation, they share with many other Austronesian languages of Papua New Guinea (Bradshaw 1982a), although the morphemes involved are not always cognate. The latter innovation appears unique to Central Province languages. In general, the Central Province languages show all of the more radical innovations in word-order features that can be found in New Guinea Austronesian languages. The prenominal positioning of relative clauses puts them in the forefront even of the most innovative languages.

In Motu, an NP containing a relative clause can be formed in one of three different ways. Each strategy requires that the clause be preposed and that the end of the modifed NP be marked with either *-na*, if the referent is singular, or *-dia*, if the referent is plural. The suffixes *-na* and *-dia*, formally identical to the 3rd person possessive suffixes, are also used to mark NP-final position in NPs modified by postposed adjectives or preposed nouns. Nevertheless, Taylor (1970:51ff.), whose analysis of Motu relativisation I rely on here, considers *-na* and *-dia* to be relativisers in this context. I will adopt Taylor's analysis in glossing the two suffixes as RSG and RPL in the examples below, although I do so primarily in order to make the ends of the NPs containing relative clauses easier to perceive.

To embed a modifying clause within an NP in Motu, it suffices simply to prepose the full clause to the head noun and to attach either *-na* or *-dia* to the end of the head noun. Within the relative clause, the position of the noun coreferential with the head may contain either a copy of the head, as in (27), or a null pronoun, as in (28). (In the Motu examples, barred *ɟ* denotes a voiced velar fricative.)

MOTU relative clauses (Taylor 1970)

With head noun copied in embedded clause:

(27) a. *[sisia ese mero e-kori-a sisia-na] e-heau*
dog ERG boy 3-bite-3S dog-RSG 3-run
'the dog that bit the boy ran away'

b. *[sisia ese mero e-kori-a sisia-na] lau ese na-lulu-a*
dog ERG boy 3-bite-3S dog-RSG 1S ERG 1S-chase-3S
'I chased the dog that bit the boy'

c. *[oi ese gaigai o-ita-ia gaigai-na] lau ese na-pidi-a*
2S ERG snake 2-see-3S snake-RSG 1S ERG 1S-shoot-3S
'I shot the snake that you saw'

With coreferent noun deleted in embedded clause:

(28) a. *[boroma e-ala-ia tau-na] na vada e-ma*
pig 3-kill-3S man-RSG SUBJ PERF 3-come
'the man who killed the pig has come'

b. *[Raka ese huala e-ala-ia sinavai-na] na daudau herea*
R. ERG crocodile 3-kill-3S river-RSG SUBJ far very
'the river where Raka killed the crocodile is very far away'

However, if the embedded noun coreferent with the head of the clause is retained in place, it is much more common to replace the head noun with a generic *ɠau* 'thing', as in (29).

MOTU relative clauses (Taylor 1970)

With generic head noun:

(29) a. *[umui vanaɠi o-kara-ia ɠau-na] tama-gu ese e-hoi-a*
2P canoe 2-make-3S thing-RSG father-1S ERG 3-buy-3S
'my father bought the canoe you made'

b. *Raka ese [boroma kaema e-ani ɠau-na] e-lulu-a*
R. ERG pig sweet.potato 3-eat thing-RSG 3-chase-3S
'Raka chased the pig that ate the sweet potato'

c. *[oi ese sisia ae-na o-ha-kwaidu-a ɠau-na] e-tai-mu*
2S ERG dog leg-3S 2-cause-break-3S thing-RSG 3-cry-CONT
'the dog whose leg you broke is crying'

9 The broader context of relative-clause bracketing in Oceanic

When viewed in the historical context of change in word order typology on the New Guinea mainland, the bracketed relative clauses in the Huon Gulf area do not look very exotic after all. In comparison with Oceanic languages elsewhere, those on the New Guinea mainland and offshore islands show two kinds of innovations with regard to the position and marking of relative clauses.

(1) The most common innovation is the marking of final position in NPs containing relative clauses. This final marker, a pronominal or demonstrative element, signals the end of the modified NP and the resumption of the matrix sentence. (It is thus not surprising that the final marker tends to be omitted just in case the matrix sentence ends at that point.) Note Lichtenberk's (1983) use of the term 'resumptive pro-form' for the clause-final marker in Manam and Wivell's (1981) use of 'phrase summary' for the same function in Kairiru. This innovation is consonant with the tendency to mark final position in NPs containing other kinds of modifiers (see Bradshaw 1982a, Ch.4).

(2) A much rarer innovation is the placement of relative clauses before, rather than after, head nouns. This innovation appears confined to the thoroughly OV languages of Central Province. These languages offer two different strategies for cross-indexing between the postposed head and the embedded coreferential noun:

(a) the embedded noun may be deleted or pronominalised, just as in the languages with postposed relative clauses discussed above; or

(b) the embedded noun may be retained and the position of the postposed head noun occupied by either the same noun or a placeholding generic noun, as in the Papuan Tip languages with OV word order.

Option (a) is shared with Oceanic languages more generally and is very likely reconstructable for Proto Oceanic. Option (b) is, as far as I know, unique to the New Guinea area, and seems to offer an innovative pathway toward changing the position of the head noun from preclausal to postclausal position by allowing it or a placeholder to occur in both places at once. Option (b) suggests that the first step in switching the position of the ancestrally preposed head and postposed clause was to place an anaphor of the head noun in clause-final position. That anaphor could be either a full copy or a placeholding pronoun or empty noun, as in the Papuan Tip languages with OV word order.

References

Bradshaw, Joel. 1982a. Word order change in Papua New Guinea Austronesian languages. PhD dissertation, University of Hawai'i.

—— 1982b. Genitives and relatives in Numbami, a New Guinea Austronesian language. In Rainer Carle, Martina Heinschke, Peter W. Pink, Christel Rost, and Karen Stadtlander, eds *Gava': studies in Austronesian languages and cultures dedicated to Hans Kähler,* 123–139. Veröffentlichungen des Seminars für Indonesische und Südseesprachen der Universität Hamburg, vol.17. Berlin: Dietrich Reimer.

—— 1985. Assessing the typological evidence for New Guinea Oceanic. In Andrew K. Pawley and Lois Carrington, eds *Austronesian Linguistics at the 15th Pacific Science Congress*, 19–30. Canberra: Pacific Linguistics.

—— 2001. Iwal grammar essentials, with comparative notes. In Andrew Pawley, Malcolm Ross and Darrell Tryon, eds *The boy from Bundaberg: studies in Melanesian linguistics in honour of Tom Dutton,* 51–74. Canberra: Pacific Linguistics.

—— 2007. Tok Pisin ia-bracketing: Neither substrate nor syntax. In Jeff Siegel, John Lynch and Diana Eades, eds *Language description, history and development: linguistic indulgence in memory of Terry Crowley*, 159–167. Creole Language Library, vol. 30. Amsterdam: John Benjamins.

Bradshaw, Joel and Francisc Czobor. 2005. *Otto Dempwolff's grammar of the Jabêm language in New Guinea.* Oceanic Linguistics Special Publication No.32. Honolulu: University of Hawai‘i Press.

Capell, Arthur. 1976. Features of Austronesian languages in the New Guinea area in general in contrast with other Austronesian languages of Melanesia. In Stephan A. Wurm, ed. *New Guinea area languages and language study,* vol. 2, *Austronesian languages,* 235–282. Canberra: Pacific Linguistics.

Clark, Stephen A. 1994. Clause combinations in Sio. MS. Summer Institute of Linguistics.

Dempwolff, Otto. 1939. *Grammatik der Jabêm-Sprache auf Neuguinea.* Abhandlungen aus dem Gebiet der Auslandskunde, vol.50. Hamburg: Friederichsen de Gruyter.

—— 1936. Bemerkungen über die Siâ-Sprache. Zu den Aufzeichnungen von Missionar Michael Stolz. (Copied out by L.·Wagner, 15/7/1936.) MS. 13pp.

Eckermann, W. 2008. A descriptive grammar of the Bukawa language of the Morobe Province of Papua New Guinea. Canberra: Pacific Linguistics.

Ezard, Bryan. 1997. *A grammar of Tawala: an Austronesian language of the Milne Bay Area, Papua New Guinea.* Canberra: Pacific Linguistics.

Greenberg, Joseph H. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In J.H. Greenberg, ed. *Universals of language,* 2d ed., 73–113. Cambridge, Mass.: The MIT Press.

Hagège, Claude. 1976. Relative clause, center-embedding and comprehensibility. *Linguistic Inquiry* 7:198–201.

Holzknecht, Susanne Carol. 1980. A morphology and grammar of the Amari language of Morobe Province, Papua New Guinea. MA thesis, University of Papua New Guinea.

—— 1989. *The Markham languages of Papua New Guinea*. Canberra: Pacific Linguistics.

Kähler, Hans. 1974. Relative clause formation in some Austronesian languages. *Oceanic Linguistics* 13:257–277.

Lauck, Linda M. 1976. Patep sentences. *Workpapers in Papua New Guinea Languages* 17:5–122.

—— 1980. Patep grammar sketch. Printout. Summer Institute of Linguistics, Ukarumpa, P.N.G.

Lichtenberk, Frantisek. 1983. *A grammar of Manam.* Oceanic Linguistics Special Publication No.18. Honolulu, University of Hawaii Press.

Olson, Clif. 1992. Gumawana (Amphlett Islands, Papua New Guinea): grammar sketch and texts. In M.D. Ross, ed. *Papers in Austronesian linguistics* No.2. Canberra: Pacific Linguistics.

Paisawa, Elsie, Louise Pagotto and Joan Kale. 1976. *A short sketch of Are (or Mukawa), Milne Bay, Papua New Guinea.* University of Papua New Guinea Department of Language, Occasional Paper No.2. 70pp.

Pawley, Andrew K. 1975. On the development of relative clauses in Oceanic languages. Paper presented at the University of Hawai‘i Oceanic Linguistics Festival.

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

Sankoff, Gillian, and Penelope Brown. 1976. The origins of syntax in discourse: a case study of Tok Pisin relatives. *Language* 52:631–666.

Siegel, Jeff. 1984. Introduction to the Labu language. In Malcolm Ross, Jeff Siegel, Robert Blust, Michael A. Colburn and W. Seiler, eds *Papers in New Guinea Linguistics* No.23, 83–159. Canberra: Pacific Linguistics.

Sohn, Ho-Min. 1973. Relative clause formation in Micronesian languages. *Oceanic Linguistics* 12:353–392.

Wivell, Richard. 1981. Kairiru grammar. MA thesis, University of Auckland.

13 *The history of the Tukang Besi pronominals*

MARK DONOHUE

1 Aims

In this paper I argue that the pronominal paradigm of Tukang Besi contains elements that are best thought of historically as not pronominal, and they are probably also not pronominal in a synchronic modern analysis. While the local (first and second) person forms appear to reflect Proto Austronesian pronouns (albeit with some changes in reference), the third person forms mostly do not. These third person forms most likely represent the fossilised relics of a voice or inverse system. Examining the syntactic distribution of these forms we find traces of their earlier non-pronominal uses, uses that are still intimately bound up in the innovative system of diathesis-monitoring in main clauses. The latter system replaced the earlier functions of the morphemes that more closely reflect Proto Austronesian reconstructions.

2 Tukang Besi pronominals

Tukang Besi is an Austronesian language spoken in Southeast Sulawesi, in central Indonesia, on the border of the area usually described as Western Malayo-Polynesian and that described as Central Malayo-Polynesian. Typologically the language is conservative in some ways, preserving a Philippine-style set of case distinctions, and innovative in others, having an agreement system reminiscent of more easterly Austronesian languages. Many of the innovations described by Blust (1993) for Central Malayo-Polynesian apply to Tukang Besi (Donohue and Grimes 2008), but its phonology also displays many highly conservative features. The position of the languages of Southeast Sulawesi is likely to prove critical for future studies of the position of the languages of eastern Indonesia within the Austronesian family. (For the position of Tukang Besi in Southeast Sulawesi, see van den Berg 2003). Within these languages, Tukang Besi is isolated both geographically and linguistically (see (19)). While most of the Muna-Buton languages are concentrated in

Alexander Adelaar and Andrew Pawley, eds
Austronesian historical linguistics and culture history: a festschrift for Robert Blust, 163–177.
Pacific Linguistics, 2009.
© This edition vested with Pacific Linguistics.

southern Buton, with the large island of Muna to its west being relatively recently settled from the south-west of Buton (to judge by the lack of linguistic differentiation—see van den Berg 1989, 1991a), the Tukang Besi islands are found minimally four hours' sailing off the east coast of Buton, with significant reefs lying between. They form a first-order split off the Muna-Buton subgroup, and show many significant differences to the other members of the family.

The basic pronominal sets for Tukang Besi are shown in Table 1 (for the use of these different pronominals, see Donohue 1999, 2004a).[1]

Table 1: Major pronominal forms in Tukang Besi[2]

Role:	S,A		P	Independent	Genitive
(mood):	realis	irrealis			
1SG	ku-	ku-	=aku	iaku	=su
2SG	'u- / nu-	ko-	=ko	iko'o	='u
3SG	no- / o-	na- / a-	='e	ia	=no
1PA	ko-	ka-	=kami	ikami	=mami
1PL	to-	ta-	=kita	ikita	=nto
2PL	i-	ki-	=komiu	ikomiu	=miu
3PL	no- / o-	na- / a-	='e	amai	=no

A number of obvious relationships exist between the different columns in this table; and we note among others that:

- the P agreement enclitics for local persons are transparently related to the independent pronouns, with only the 2sg form requiring any explanation at all;
- only the S,A prefixes are consistently monosyllabic, and only these forms show a contrast for modality.

Looking across the different rows, we consistently find that resemblance between the independent forms and the bound forms is lowest for the third person cells. For instance, the 1SG forms are consistently coded with a *ku* element (the only exception being in the genitive column, where a *k > s change has applied), and the 1PL forms inevitably involve a *tV*; the 2SG forms always involve a back rounded vowel, and either a *k*, a glottal stop, or both; the 2PL forms involve the sequence *(ko)miu* (with the S,A prefixes being

[1] Tukang Besi examples are transcribed in the phonemic transcription used in Donohue (1999). The conventions are broadly equivalent to IPA standards, with the following exceptions: *b* represents a voiced imploded bilabial stop, *d* represents a voiced imploded dental stop, ' represents a glottal stop, *ng* represents a velar nasal, and *u* represents a high back vowel (with dialectal variation between rounded and unrounded realisationssee Donohue 1999:11). Abbreviations follow Leipzig glossing conventions, with the addition of: AgrA, AgrS, AgrP: agreement for A, S or P, Agt: S or A, CORE: core case, Local: first or second person, P2: 'second P': theme in a trivalent clause, PAn: Proto Austronesian, Pat: P, PERF: perfective, PMP: Proto Malayo-Polynesian, R: realis, RED: reduplication, SI: S,A infix.

[2] There are additionally two 'minor' pronominal sets, the dative (1SG: =*naku*, 2SG: =*nso*, 3SG/PL: =*ne*, 1PA: =*nsami*, 1PL: =*nggita*, 2PL: =*ngkomiu*) and the intradirective (1PA: =*ngkami*, 1PL: =*ngkita*, 2PL: =*ngkomiu*, 3PL: ='*e*), used only when verbs of directed motion are serialised with numeral verbs. These two sets are both derived, with irregular developments, from the P enclitics, and so must be historically subsequent to the development of the P clitics. The absence of any singular forms for the intradirective P is simply explained by the fact that the intradirective construction only appears with nonsingular subjects.

exceptional). The bound third person forms, however, are not phonologically related to either the singular or the plural free pronoun; while the S,A prefixal forms reflect fused genitive pronouns (see Blust 1993; Wolff 1996; Himmelmann 1996 for this development), the P clitic form, which in local persons are startlingly similar to the free forms, is not related. It might be argued that the two vowels of *ia* have been reduced to produce the *e*; however, there is no support for this. Although *ay regularly becomes *e* in Tukang Besi, *ia sequences in general are maintained as *ia*, as shown in Table 2. A reduction from *ay to *e* would therefore be without support anywhere else in the language.

Table 2: The treatment of PAn *ia sequences in Tukang Besi

PAn		Tukang Besi
*qalia	'ginger'	*lo'ia*
*i-aku	'1SG'	*iaku*
*liang	'cave'	*lia*
*beRsay	'paddle'	*bose*
*qatay	'liver'	*ate*

The third person cells in the different paradigms seen in Table 1 are paradigmatically and formally exceptional. With the exception of the independent forms, there is no number distinction for third persons. Furthermore, the 3PL form *amai* which does create a number distinction in the independent series, is not historically Austronesian, not being attested in any other Austronesian language (known to me), nor in more 'archaic' forms of the Tukang Besi language, such as songs, poetry, or proverbs. This means that the singular:plural distinction that is found in the third person free pronouns does not reflect the basic system: it is a contrast that has been reinvented. The second persons show a singular: plural contrast with forms that are easily related to older and widely attested Austronesian pronouns (see Table 3), and first persons show a singular: paucal: plural contrast with similarly ancient pronouns. In terms of feature-density, as well as formally, the third person pronouns are distinct from the local persons (that is, first or second persons) (see also the discussion in Donohue 2006).

3 The origins of the pronominal forms

Examined with a view to their development from Proto Austronesian, we again find unusual behaviour with the third person forms. Table 3 shows the evolution of the pronouns that led to the modern Tukang Besi forms; only the relevant pronominals are shown (drawn from Ross 2002; see also Blust 1977). While the local person pronouns are clearly related to the PAn and PMP forms shown here, the third person plural forms in Tukang Besi show no such relationship, and the P form of the third person, which is neutral as to number, is not obviously related to any reconstructed forms (see Table 2).

Table 3: The origin of pronominal forms in Tukang Besi

Role:	PAn	PMP	Independent	P
1SG	*i-aku	*[i]aku	*iaku*	*=aku*
2SG	*[i-]ka-Su	*ikahu	*iko'o*	*=ko*
3SG	*s(i)-ia	*(s)iya	*ia*	*='e*
1PA / 1PL.EXCL	*[i-]k-ami	*[i]kami	*ikami*	*=kami*
1PL / 1PL.INCL	*[i-]k-ita	*[i]kita	*ikita*	*=kita*
2PL	*[i-]k-amu	*[i]kamu *kamu-ihu	*ikomiu*	*=komiu*
3PL	*si-da	*sida	*amai*	*='e*

Another factor that shows us that the third persons should be treated separately from the local persons lies in the use of the pronominal enclitics for P. While prefixal agreement on verbs is obligatory except in some minor constructions (such as imperatives), only two verbs require the P clitics.[3] A verb with two arguments, such as *manga* 'eat', can appear with or without a clitic, as shown in (1) and (2).

(1) *No-manga te pandola na ana-anabou.*
3R-eat CORE eggplant NOM RED-child
'The children ate the eggplants.'

(2) *No-manga='e te ana-anabou na pandola.*
3R-eat=3P CORE RED-child NOM eggplant
'The children ate the eggplants.'

The variation in marking, whereby the same verb has two coding patterns, corresponds to a selection of voice: the clause in (1) is active/direct, while the clause in (2) is inverse (that is, the subject is a non-agent, but the agent is not demoted to an oblique function). That is, in (1) the grammatical subject is the children, while in (2) it is the eggplants. In (3) and (4) we can see that the scope of a floating quantifier is restricted to the subject of the clause, regardless of whether or not that subject is the A or the P (there are other tests that uniquely identify the subject, as described in Donohue 1999, 2004a).

(3) *Saba'ane no-manga te pandola na ana-anabou.*
all 3R-eat CORE eggplant NOM RED-child
'All the children ate the eggplants.'

(4) *Saba'ane no-manga='e te ana-anabou na pandola.*
all 3R-eat=3P CORE RED-child NOM eggplant
'The children ate all the eggplants.'

[3] The verbs are *molinga* 'forget' and *raho* 'affect'. The analysis of *mbea'e* 'not exist' as being based on the root *mbea(ka)* 'not', is not contentious, but suggesting that it involves a P clitic as its only agreement marker is controversial.

When the P of the clause is a local person, however, this optionality vanishes; while a first person 'searcher' can occur with a third person 'searchee' that either is, or isn't, indexed with a clitic on the verb, the reverse is not true, and a local person P must show indexing.[4]

(5) a. *Ku-laha te mia iso.*
1SG-look.for CORE person yon
'I looked for that person.'

b. *Ku-laha='e na mia iso.*
1SG-look.for=3P NOM person yon
'I looked for that person.'

(6) a. **No-laha te iaku na mia iso.*
3R-look.for CORE 1SG NOM person yon

b. *No-laha=aku te mia iso.*
3R-look.for=1SG.P CORE person yon
'That person looked for me.'

The fact that the alternation is more productive with third person Ps than with local persons, and that it has the effect of changing the identity of the grammatical subject, means that the alternation has the primary characteristics of a voice alternation. If we accept that, then we are several steps closer to understanding the origin of the anomalous third person P enclitics.

Table 4 shows the possibilities for verbal inflection with different As and Ps. The only part of the Table where we observe dynamic alternations are the cells with third person Ps; it is not unreasonable to assume that the variability in this part of the Table, and the lack of variation when the P is local, reflects a preference for the selection of local persons as subject whenever possible (see Aissen 1999 for discussion). While it is certain that pragmatic factors play a major role in the selection of verb, and hence clause, types, with more pragmatically salient arguments being preferentially encoded as subjects, examples such as (6) show that a local person must be indexed on the verb. This condition can be profitably rephrased by noting that *a local person P cannot be encoded as an object*.[5] (7) presents the relative ordering of these constraints, phrased using the terminology of Aissen (1999); the constraints are explicated in (8)–(11).

---

[4] Donohue (1999) reports sentences such as (6a) as grammatical. They are interpretable, and will be accepted by speakers if asked, but a search of all collected textual material (conducted for and partially reported in Donohue 2001) reveals the fact that 100% of local persons in bivalent (or monovalent) clauses are indexed on the verb, indicating that sentences such as (6a) are highly infelicitous, since they are not naturally produced.

[5] The fact that a local person A CAN be encoded as an object reflects the relative salience enjoyed by the highest role in a verb's argument structure, compared to the other arguments.

Table 4: Verb and voice selection for different participant combinations

P: / A:	1	2	3
1	–	a-V-p	a-V-p, a-V
2	a-V-p	–	a-V-p, a-V
3	a-V-p	a-V-p	a-V-p, a-V

(7) *OBJ/Pat/Local » *SUBJ/x » *SUBJ/Pat » *SUBJ/Agt

(8) *OBJ/Local
Local persons should not be encoded in the syntax such that they appear as objects

(9) *SUBJ/x
Do not encode a participant with low pragmatic importance (not salient, topical, focussed, etc.) as the subject of the clause

(10) *SUBJ/Pat
Do not encode a 'Pat' (P, theme, lowest argument) of a verb as the subject of the clause

(11) *OBJ/Local
Do not encode an 'Agt' (A, agent, highest argument) of a verb as the subject of the clause

Table 5 below repeats Table 4, but shows the identity of the subject and object of the clause. Whenever there is a local P, it must be selected as subject, and there is variation only with third person Ps. Note that, unlike the passive alternation in English or in many Oceanic languages, the change in identity of the subject is not associated with the demotion of an argument: both the A and the P are core arguments of the verb, and the alternation is more like the inverse described in Algonquian languages (see, for example Givón 2001, and many others).

Table 5: Verb and voice selection for different participant combinations

P`: / A:	1	2	3
1	–	A: OBJ P: SUBJ: 2	A: SUBJ/OBJ P: SUBJ/OBJ
2	A: OBJ P: SUBJ: 1	–	A: SUBJ/OBJ P: SUBJ/OBJ
3	A: OBJ P: SUBJ: 1	A: OBJ P: SUBJ: 2	A: SUBJ/OBJ P: SUBJ/OBJ

I suggest that the facts just described can best be accounted for diachronically by assuming that, just as in many other languages with inverse systems (Dahlstrom 1991, Rhodes 1990, 1994), the system was dynamic only with third person arguments. This would mean that we posit that at an earlier stage of Tukang Besi agreement for A was obligatory, and that clauses with third person Ps could optionally be marked with an invariant inverse marker, [ʔe]. The etymology of this morpheme is not known, but the similarities to the widely-attested southern Austronesian morpheme *aken, related to the Proto Oceanic long transitive marker *-akin, is suggestive (*k > ʔ is irregular, but not unprecedented, in Tukang Besi; additionally there is some infrequent vowel lowering); alternatively the locative applicative suffix *-((V)C)i* (see (13) below), which serves to mark an increase in the pragmatic status of a non-subject participant (see Donohue 1999:Ch.10; 2001) is a not-too unlikely possibility for the inverse marker.[6]

(12)	Local person P	Third person P	
	AgrA-verbroot	AgrA-verbroot	A = Subject
		AgrA-verbroot-INVERSE	P = Subject

Even though the inverse marker was not a pronominal inflection, it was nonetheless a verbal morpheme appearing after the verb root. This is significant in that all the verbal morphology that Tukang Besi preserves from Proto Austronesian is prefixing; suffixing morphemes from Proto Austronesian are no longer productive in Tukang Besi on verbs (see Donohue 2002). (13) shows the affixal possibilities for verbs in Tukang Besi; as can be seen, the verbs are overwhelmingly prefixal, with only the locational applicative showing any antiquity (the prefixes are shown, from top to bottom, in their relative order; thus passives must prefix causatives).[7]

(13)	Prefixes	‘suffixes’
	nominative prefixes (Table 1)	locative applicative *-((V)C)i*
	S,A infix *-um-*	comitative applicative *-ngkene*
	passive prefixes *to-*, *te-*	
	anticausative prefix *mo-*	
	requestive prefix *hepe-*	
	causative prefixes *hoko-*, *pa-*	
	reciprocal prefixes *po-*, *pada-*	
	valency neutral and valency announcing prefixes *he-*, *hoN-*, *heme-*, *me-*, *heka-*, *homo-*, *hopo-*, *para-*, *wo-*, *ban-*, occupational *pa-*	

While there are aspectual clitics that follow the verb root, as in (14a), these are true second-position clitics, as can be seen in (14b) and (14c).

---

[6] This scenario does not suppose that *aken, or *-akin, must be reconstructed for Proto Austronesian. Since the history of this morpheme is not known and it does not exclusively belong to any subgroup, it might as easily be taken to be an areal innovation, possibly originating in a protolanguage south of the Philippines.

[7] By contrast, noun morphology is predominantly suffixal. The suffixal comitative applicative *-ngkene* appears to be a relatively recent grammaticalisation (related to *kene* ‘with’, and *kene* ‘friend’). The transparency of phonological form between the noun, ‘preposition’, and suffix, and the absence of cognate forms in related languages, suggests a very recent spread, though an exact chronology is not yet known.

(14) a. *No-manga=mo te kuikui.*
3R-eat=PERF CORE cake
'They have eaten the cakes.'

b. *Po'oli=mo no-manga te kuikui.*
finish=PERF 3R-eat CORE cake
'They have already eaten the cakes.'

c. * *po'oli nomangamo te kuikui*

In short, suffixal morphology on verbs was almost unknown in a pre-Tukang Besi stage, but with the introduction of the inverse marker this condition in the language began to break down. The obvious first stage in the development of post-root inflection was a stage in which pronouns could be cliticised on the verb, as in (15). Recall that the historical source of the inverse marker is likely to have involved a high vowel, i. This becomes significant when we notice that the free pronouns all begin with an initial i, meaning that the structure in (15) would be likely to be reanalysed as (17), with the third person inverse construction finding support from the developing local-person cliticisation just as the local person cliticisation developed with the support of the inverse marking.

(15)

Local person P	Third person P	
AgrA-verbroot=Pronoun	AgrA-verbroot	A = Subject
	AgrA-verbroot-INVERSE	P = Subject

Hypothesised pre-Tukang Besi, 1

(16) *No-laha=iaku te mia iso.*
3R-look.for=1SG CORE person yon
'That person looked for me.'

(17)

Local person P	Third person P	
AgrA-verbroot-INVERSE=Pronoun	AgrA-verbroot	A = Subject
	AgrA-verbroot-INVERSE	P = Subject

Hypothesised pre-Tukang Besi, 2

(18) *No-laha-i=aku te mia iso.*
3R-look.for-INVERSE=1SG CORE person yon
'That person looked for me.'

We now have a situation in which local persons must always be subject when they are a P, since their clitic pronoun forms have been reinterpreted as showing inverse marking. The inverse marker is optional only with third person Ps, where there is no clitic pronoun set to differentiate between singular and plural.[8]

---

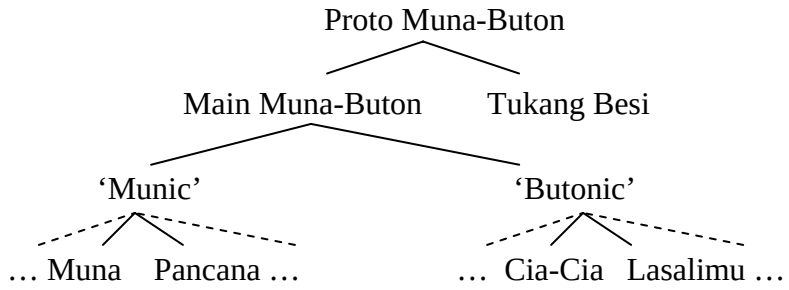
[8] We might hypothesise that the sequence (vowel-final) verb root + inverse suffix + third person singular pronominal, as in (i), created too many vowels in sequence; the maximal number of vowels attested in any word is three, in the loan *daoa* 'market' (a lexeme shared throughout insular Southeast Sulawesi), otherwise only two.

(i) *No-laha-i=ia.*
3R-look.for-INVERSE-3SG
'They looked for her/him.'

We have not yet derived the modern forms; (18) differs from (6b) in the presence of the inverse marker. Is it unreasonable to suppose that this suffix would show two different paths of development, being preserved when there is no pronominal element following, and being lost when there is some? When we examine the other languages of Muna and Buton we find support for this hypothesis.

The languages of Muna and Buton are quite closely related (van den Berg 2003), with the exceptions of Kulisusu and Taloki in the north of Buton (Mead 1998, 1999) and the immigrant languages Wolio and Kamaru (Donohue 2005). The essential details of their relationship are shown in (19); not all languages are shown, only those discussed in this paper.

(19)



An idea of the grammatical differences found in the treatment of pronominal suffixes/enclitics can be gauged by comparing the Tukang Besi and Pancana clauses shown in (20). While Tukang Besi only allows the recipient to show agreement on the verb, (20a), with either double agreement (20c) or agreement for the theme rather than the recipient (20e) being ungrammatical, Pancana allows for both the recipient and theme (in that order) to show agreement on the verb, (20b). This is only grammatical if the suffix indexing the recipient is drawn from the second paradigm of P agreement, as seen in (20d).[9]

(20) Tukang Besi

a. *No-hu'u=ko te atu.*
3R-give=2SG.P CORE that
'They gave that to you.'

c. **No-hu'u=ko='e*
3R-give=2SG.P=3P

e. **No-hu'u=ke te iko'o*
3R-give=3P CORE 2SG

Pancana

b. *No-waa-angko-e.*
3R-give-2SG.P2-3SG.P
'They gave it to you.'

d. **No-waa-ko-e*
3R-give-2SG.P2-3SG.P

[9] Like the dative enclitics in Tukang Besi (see fn.2), the second P agreement paradigm in Pancana resembles the 'primary' paradigm with the addition of an initial nasal element. The forms for the primary P agreement paradigm are: 1SG: *-kanau*, 2SG: *-ko*, 3SG: *-e*, 1PL: *-kaita*, 2PL: *-koomu*, 3PL: *-e*, while the secondary paradigm is: 1SG: *-kanau*, 2SG: *-angko*, 3SG: *-ane*, 1PL: *-kaita*, 2PL: *-angkoomu*, 3PL: *-andai*. Unlike Tukang Besi, the first person forms in Pancana are invariant, and both P suffixal paradigms bear scant resemblance to the modern free pronouns (1SG: *inodi*, 2SG: *ihintu*, 3SG: *anoa*, 1PL.EX: *insaidi*, 1PL.IN: *intaidi*, 2PL: *ihintuomu*, 3PL: *andoa*) (see van den Berg 1991c for the evolution of the pronouns of Muna, which is closely related to Pancana).

Interestingly, while Tukang Besi allows the same P agreement clitic to appear to encode the P of a root-bivalent verb or the second argument of an applicative verb, as seen in (21a) and (21e), Pancana does not. The primary P agreement paradigm cannot be added to an applicative verb to index the applicative argument, (21f) nor may the members of the secondary P agreement paradigm, (21g). Rather, the secondary P agreement suffix alone, without an applicative morpheme, is used, (21h).

(21) Tukang Besi

a. *No-'ita=ko.*
3R-see=2SG.P
'They saw you.'

c. *No-wila(=ako te ia).*
3R-go=APPL CORE 3SG
'They went (for her/him).'

e. *No-wila=ako=ko.*
3R-go=APPL=2SG.P
'They went for you.'

g. **No-kala-a'u-angko*
3R.A-go-APPL-2SG.P2
'They went for you.'

h. *No-kala-angko.*
3R.A-go-2SG.P2
'They went for you.'

Pancana

b. *No-'ondo-ko.*
3R.A-see-2SG.P
'They saw you.'

d. *No-kala(-a'u anoa).*
3R.A-go-APPL 3SG
'They went (for her/him).'

f. **No-kala-a'u-ko*
3R.A-go-APPL-2SG.P
'They went for you.'

It is clear that Pancana does not permit a suffix that marks an increase in the pragmatic status of a clausal participant, such as the applicative, to cooccur with a pronominal agreement marker for that argument; rather, an agreement suffix alone is used. In Cia-Cia, from the Butonic branch of Muna-Buton, and closer to Tukang Besi geographically, the same pattern is found, as seen in (22). In Lasalimu, which is even closer (geographically) to Tukang Besi, the applicative may be followed by P agreement suffixes.

Cia-Cia (van den Berg 1991b, and author's own field notes)

(22) a. *No-'ita-so.*
3R-see=2SG.P
'They saw you.'

b. *No-hangka(-aso ia).*
3R.A-go-APPL 3SG
'They went (for her/him).'

c. **No-hangka-aso-so*
3R-go=APPL=2SG.P
'They went for you.'

Lasalimu (author's own field notes)

(23) a. *A-'ita-so.*
3R-see=2SG.P
'They saw you.'

b. *A-lampa(-aso ia).*
3R.A-go-APPL 3SG
'They went (for her/him).'

c. *A-lampa-aso-ko.*
3R-go=APPL=2SG.P
'They went for you.'

The preceding discussion shows that positing a condition of the form shown in (24) is justified. Here we see that a constraint against the realisation of the inverse marker with a local person P is more highly ranked than the simple realisation of the agreement marker itself. Such a condition is widespread in the languages of Muna and Buton, licensed at least in part by the fact that the languages have more than one set of suffixal markers for Ps. This second set of suffixes is present only in relic form in Tukang Besi, and the condition against the appearance of P agreement with suffixal material is consequently relaxed: regular P clitics may follow an applicative suffix, just as in Lasalimu.

(24) *V-INVERSE-AgrP*local* > *V-AgrP*local*

We assume that the history of these languages involved the following steps.

(25) Stage 1. locative suffix reinterpreted as an inverse marker
(this might have only applied in Tukang Besi; comparative evidence is as yet incomplete, but there are indications that an inverse type is more widespread in the Southeast Sulawesi languages. Donohue (2004b) documents other cases of voice morphology being reanalysed as agreement markers in Austronesian languages)

Stage 2. P agreement markers start to appear in the form of cliticised pronouns
(the fact that only some forms appear to be cognate across the Southeast Sulawesi languages suggest that this was not a uniform process in the area, but subject to repeated innovation)

Stage 3. secondary P agreement markers appear in the west
(the fact that this innovation is found in all Munic languages, and the western Butonic language Cia-Cia, but not in Lasalimu, suggests that this too was a change that spread after the breakup of the main Muna-Buton group. The presence of the archaic dative agreement paradigm in Tukang Besi might suggest that the secondary P agreement markers were an ancient feature lost in the east; more detailed work is required before this question can be answered.)[10]

Stage 4. P agreement markers, having developed from the cliticised pronouns, are judged incompatible with the inverse marker.

Stage 5. P agreement markers are judged incompatible with the applicative morphemes.
(this only applies to languages that developed/preserved the secondary P agreement paradigm)

We have, thus, been able to plausibly derive the two sub-paradigms seen in (26), in which the inverse marker only appears with third persons, and cannot cooccur with the pronominal clitics that both index their P, and mark the clause as having a syntactically inverse syntactic interpretation.

(26)

Local person P	Third person P	
AgrA-verbroot=AgrP	AgrA-verbroot	A = Subject
	AgrA-verbroot-INVERSE	P = Subject

[10] Not enough data is available on the remaining Butonic language, Kumbewaha, spoken near Lasalimu.

4 Consequences

The erratic nature of the third person cells from the P agreement paradigm in modern Tukang Besi has been shown to be due to the non-pronominal nature of their origin. While the local persons are transparently related to the independent pronouns, reflecting a recent grammaticalisation from them, the third person forms bear little resemblance to contemporary Tukang Besi pronouns, or reconstructed Austronesian pronouns, and are best thought of as being non-pronominal in origin, and perhaps in synchronic description as well.

If we accept that pronouns can be represented by pronominal features, then the most parsimonious way to differentiate three persons is with two features, [FIRST] and [SECOND] (this analysis goes back as least to Ingram 1978), as shown in (27).

(27)

	[FIRST]	[SECOND]
first person	+	–
second person	–	+
third person	–	–

There are some phonological correlates to this; in Proto Austronesian, and in Tukang Besi, only the first and second person pronouns, those that have positive features, contain the phoneme *k*. This might be trivial, but it does add to the argument that the local persons are treated differentially to the third persons. I argue that the 'third person' forms are not only unmarked with any positive pronominal features, but they are not pronominal. We have seen that the P clitic can be plausibly described as an inverse marker. The S,A prefixes are suppletive for the realis/irrealis distinction, which is not otherwise marked, and show different patterns of occurrence with time adverbs, as shown in (28); note that the first person singular does not vary along the realis/irrealis dimension, and so in these cases the only realisation of tense is on the adverb, as in (29).[11]

(28) a. *Kehia* *na-mai* *'uka* *la?*
when.FUTURE 3I-come again MASC
'When will they come back?'

b. *Dehia* *no-mai* *'uka* *la?*
when.PAST 3R-come again MASC
'When did they come back?'

(29) a. *Tabea* *ku-mai* *'uka.*
must.FUTURE 1SG-come again
'I will have to come again.'

[11] On a verb root not beginning with *m* the infix <*um*> is likely to appear with the irrealis prefixes, though not completely required. Thus, corresponding to (29), we would find (i) - (iii), in which the behaviour of <*um*> is overt.

(i) *Tabea* *ku-r<um>ato* *'uka.*
must.FUTURE 1SG-arrive<SI> again
'I will have to arrive again.'

(ii) *Tabea* *ku-rato* *'uka.*
must.FUTURE 1SG-arrive again
'I have to arrive again.'

(iii) *Tabeda* *ku-rato* *'uka.*
must.PAST 1SG-arrive again
'I had to arrive again.'

b. *Tabeda ku-mai 'uka.*
must.PAST 1SG-come again
'I had to come again.'

Given that all the prefixes (other than the first person singular) show portmanteau forms incorporating modality, it is not unreasonable to suggest that *na-* and *no-* in (28), while derived from earlier genitive pronouns, are synchronically simply markers of modality, and that a simple principle of morphological blocking operates to restrict their use to contexts in which the portmanteau forms are ruled out due to feature clashes with a third person argument.

What of the independent forms, *ia* '3SG' and *amai* '3PL'? While *ia* is certainly etymologically related to the Austronesian forms seen in Table 3, *amai* is not, and has no known etymology.

In short, Tukang Besi is, then, a language in which the pronominal forms for third persons almost completely lack Austronesian etymologies. This is not a particularly unusual state of affairs, with demonstratives often being pressed into service as third person (or occasionally even local person) personal pronouns. These now apparently third person forms developed through reasonable pathways, but the story that explains their (putative) provenance indicates that a purely synchronic description will lack significant substance if it is formulated without consideration of the historical facts, where they can be determined.

References

Aissen, Judith. 1999. Markedness and subject choice in optimality theory. *Natural Language and Linguistic Theory* 17 (4):673–711.

Blust, Robert. 1977. The Proto Austronesian pronouns and Austronesian subgrouping. *Working Papers in Linguistics, University of Hawaii* 9 (2):1–15.

—— 1993. Central and Central-Eastern Malayo-Polynesian. *Oceanic Linguistics* 32:241–293.

Dahlstrom, Amy. 1991. *Plains Cree morphosyntax*. New York: Garland.

Donohue, Mark. 1999. *A grammar of Tukang Besi*. Berlin: Mouton de Gruyter.

—— 2001. Coding choices in argument structure: Austronesian applicatives in texts. *Studies in Language* 25 (2):217–254.

—— 2002. Voice in Tukang Besi and the Austronesian voice system. In Fay Wouk and Malcolm Ross, eds *The history and typology of western Austronesian voice systems*: 81–99. Canberra: Pacific Linguistics.

—— 2004a. Voice oppositions without voice morphology. In Paul Law, ed. *Proceedings of AFLA 11, ZAS, Berlin 2004*. ZAS Papers in Linguistics Nr. 34 – October 2004:73–88. Berlin: Zentrum für Allgemeine Sprachwissenschaft, Typologie und Universalienforschung, Berlin.

—— 2004b. Voice marking → verbal agreement: reanalysing Austronesian voice morphology. Talk presented to the Department of Linguistics, Max Planck Institute for Evolutionary Anthropology, Leipzig, 20 February 2004.

—— 2005. The pretenders to the Muna-Buton group. In John Bowden and Nikolaus Himmelmann, eds *Papers in Austronesian subgrouping and dialectology,* 21–35. Canberra: Pacific Linguistics.

—— 2006. Pronouns, clitics, orders and grammaticalisation: theoretical issues in Tukang Besi. Paper presented at the tenth international conference on Austronesian linguistics. Linguistic Society of the Philippines and SIL International. Puerto Princesa, Philippines. January 17–20, 2006.

Donohue, Mark, and Charles E. Grimes. 2008. Yet more on the position of the languages of eastern Indonesia and East Timor. *Oceanic Linguistics* 47 (1):115–159.

Givón, Talmy. 2001. *Syntax: an introduction*. Amsterdam: John Benjamins.

Himmelmann, Nickolaus. 1996. Person marking and grammatical relations in Sulawesi. In H. Steinhauer, ed. *Papers in Austronesian linguistics* No.3:115–136. Canberra: Pacific Linguistics.

Ingram, David 1978. Typology and universals of personal pronouns. In Joseph Greenberg, ed. *Universals of Human Language* vol. 3:213–248. Stanford: Stanford Univ. Press

Mead, David E. 1998. Proto-Bungku-Tolaki: reconstruction of its phonology and aspects of its morphosyntax. PhD thesis. Rice University.

—— 1999. *The Bungku-Tolaki languages of South-Eastern Sulawesi, Indonesia.* Canberra: Pacific linguistics.

Rhodes, Richard A. 1990. Obviation, inversion, and topic rank in Ojibwa. In David J. Costa, ed. Procedings of the sixteenth annual meeting of the Berkeley Linguistics Society, February 16–19, 1990: Special session on general topics in American Indian linguistics: 101–115. Berkeley: Berkeley Linguistics Society.

—— 1994. Agency, inversion, and thematic alignment in Ojibwe. In S. Gahl, et al., eds Proceedings of the twentieth annual meeting of the Berkeley Linguistics Society: 431–446. Berkeley: Berkeley Linguistics Society.

Ross, Malcolm. 2002. The history and transitivity of western Austronesian voice and voice-marking. In Fay Wouk and Malcolm Ross, eds *The history and typology of western Austronesian voice systems*: 17–62. Canberra: Pacific Linguistics.

van den Berg, René. 1989. *A grammar of Muna*. Dordrecht: KITLV Press.

—— 1991a. Muna dialects and Munic languages: towards a reconstruction. In Ray Harlow, ed. *VICAL 2: Western Austronesian and contact languages, parts I and II: Papers from the fifth international conference on Austronesian linguistics*: 21–51. Auckland: Linguistic Society of New Zealand.

—— 1991b. 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 KITLV*: 305–324. Leiden: KITLV Uitgeverij.

—— 1991c. Muna historical phonology. In J.N. Sneddon, ed. Studies in Sulawesi linguistics, part II: 1–28. *NUSA* 33.

—— 2003. The place of Tukang Besi and the Muna-Buton languages. In John Lynch, ed. *Issues in Austronesian historical phonology*: 87–113. Canberra: Pacific Linguistics.

Wolff, John. 1996. The development of the passive verb with pronominal prefix in Western Austronesian languages. In Bernd Nothofer, ed. *Reconstructions, classification, description – festschrift in honor of Isidore Dye*n, 15–40. Hamburg: Abera.

14 *Verbal aspect and personal pronouns: the history of aorist markers in north Vanuatu*

ALEXANDRE FRANÇOIS

1 Introduction[1]

The subject and object clitics which are reconstructed for Proto Malayo-Polynesian (Blust 1977) and for Proto Oceanic (Lynch, Ross and Crowley 2002:67) have been replaced in Mwotlap, as in many other languages of north Vanuatu, with a unique set of free pronouns: *nɔ* '1sg'; *nɪk* '2sg'; *kɪ* '3sg'… These may be used both in subject and object positions:[2]

(1) *nɔ* m-ɛtsas *kɪ*, *k* m-ɛtsas *nɔ*.
1sg PRF-see 3sg 3sg PRF-see 1sg
'I saw her and she saw me.'

Most Mwotlap pronouns are morphologically invariant. However, in subject position, the 1sg pronoun shows allomorphic variation between two forms *nɔ* and *nɔk*. This uncommon alternation depends on the tense-aspect-mood (TAM) marking of the verb. Out of the twenty-five TAM categories in Mwotlap (François 2003), eight allow for free variation between the two forms, whereas in the rest of the system, they come in strict complementary distribution: ten markers require *nɔ* as their subject, while seven require *nɔk*. In fact, as we will see below, *nɔk* itself can be described as a portmanteau form indexing both person and aspect.

[1] I am grateful to COOL7 participants for their questions on an oral version of this paper (François 2007), and to Claudia Wegener and Alexis Michaud for their comments on an earlier draft.

[2] All transcriptions use IPA rather than local orthographies, to facilitate comparison. Note that |v| is often bilabial [β]; |j| is the palatal approximant; |t| is a laminal retroflex; all voiced stops are prenasalised (|d| is [ⁿd], |g͡bʷ| is [ⁿg͡bʷ], etc.). Besides abbreviations that follow the Leipzig glossing rules, AO means 'Aorist'; POT 'Potential'; PROSP 'Prospective'; STAT 'Stative'; (P)NCV '(Proto) North-Central Vanuatu'; TAM 'tense-aspect-modality'. The three-letter abbreviations for modern languages are spelled out on Map 1.

Alexander Adelaar and Andrew Pawley, eds
Austronesian historical linguistics and culture history: a festschrift for Robert Blust, 179–196.
Pacific Linguistics, 2009.
© This edition vested with Pacific Linguistics.

This formal variation of the 1sg pronoun depending on the predicate's TAM-marking is typologically original. It also constitutes a morphological puzzle, which I will take as the starting point for this paper. Section 2 will begin with a synchronic approach, by describing the semantic motivation of the *nɔ–nɔk* contrast in Mwotlap; this will lead to the functional definition of an aspect category labeled 'Aorist'.[3] Based on this definition, §3 will investigate the geographical distribution and the formal characteristics of similar aorist markers across the seventeen languages of the Banks and Torres islands. Finally, §4 will take a historical perspective, and attempt to unravel the development of aorist markers in north Vanuatu languages.

2 The Aorist in Mwotlap

The first question I will tackle is the functional distribution of the two allomorphs *nɔ* and *nɔk* in Mwotlap.

2.1 A special pronoun for the Aorist

The word order of constituents in Mwotlap is as follows:

Subject NP – (TAM clitic/prefix) – Predicate – (TAM postclitic) – (Object NP)

As far as the 1sg pronoun is concerned, its unmarked, default form is clearly the shorter allomorph *nɔ*. It is the only one found in non-subject positions—see (1)—as well as for the subject of non-TAM predicates (e.g. *nɔ na-vatɣɔ* 'I'm a teacher', *nɔ itʊk* 'I'm fine'). As for tense-marked predicates, *nɔ* combines with realis (Stative, Perfect, Completive …) as well as irrealis markers (Future, Potential, Counterfactual …):

(2) *nɔ* mɪ-wɪl nu-suk.
1sg PRF-buy ART-sugar
'I've bought some sugar.'

(2′) *nɔ* tɪ-wɪl vɪh nu-suk.
1sg POT_1-buy POT_2 ART-sugar
'I can buy some sugar.'

The seven TAM categories requiring the marked form *nɔk* are the Aorist proper, the Permansive, the Prioritive, the two Presentatives (static and kinetic), the Polite Imperative and the Prospective. Despite their semantic differences, the latter six categories are related, as they are all formally derived from the Aorist, through combination with some secondary morpheme. In other words, the marked form *nɔk* is required whenever the 1sg pronoun is the subject of a tense-marked predicate belonging to the domain of the Aorist, in the wide sense of the term. Therefore I will hereafter gloss it '1sg:AO'.

Crucially, when the predicate is an Aorist strictly speaking—as opposed to one of its derivatives—the pronoun *nɔk* is in fact the only formal TAM marking in the clause. This confirms its interpretation as a portmanteau morpheme, combining person and TAM marking:

[3] About the term 'aorist', see the end of §2.2.2. Note that I use capitalisation, following Comrie (1976:10), whenever a given term is to be understood as a labeling convention for a morphosyntactic category specific of a given language, rather than a typological concept.

(3) *nɔk* ŋ͡mʷʊl
1sg:**AO** return
‘Let me go back!’ (…)

While the first person encodes the Aorist through variation of the pronoun, the 3sg does this with a prefix *ni-* on the verb, in the slot usually devoted to other aspect markers —compare (3′) with (2).

(3′) kɪ *ni*-ŋ͡mʷʊl
3sg **AO**:3sg-return
‘Let him go back!’ (…)

Finally, all persons other than 1sg and 3sg encode the Aorist with a zero:

(3″) ɣɪn (*Ø*-)ŋ͡mʷʊl
1inc:pl (**AO**-)return
‘Let's go back!’ (…)

The Aorist and its derivatives are the only TAM categories of Mwotlap whose marking depends on the person.

2.2 The semantics of the Aorist

2.2.1 The various uses of the Aorist

Like several other TAM categories of Mwotlap, the Aorist is only compatible with semantically dynamic events. Its combination with a stative predicate—whether a stative verb, an adjective or a noun—forces a dynamic interpretation [see (9) and (13) below].

But the semantic information which the Aorist gives about that dynamic event is widely polysemous (François 2003:165–199). For one thing, the Aorist encodes events that come in sequence. This applies equally in past or future contexts:

(4) *nɔk* hajvɛɣ l-ɪŋ͡mʷ nɔnɔn ɛ kɪ *ni*-ɛtsas nɔ.
1sg:AO enter in-house his TOPIC 3sg AO:3sg-see 1sg
[*past context*] ‘I came into his house and (then) he saw me.’
[*future context*] ‘I'll come into his house and (then) he'll see me.’

Crucially, the same sentence (4) may be translated in English either as past or as future. This shows that the Aorist is not a tense, but an aspect, which may attach either to a realis or to an irrealis situation. In itself, (4) says nothing more than ‘(Let there be) my coming into his house, *and then* him seeing me …’. What is relevant here is the relation of sequence or implication between the two successive events, regardless of how they happen to relate to the speech coordinates. The Aorist is commonly found in narratives, for any chain of events:

(5) kɪ *ni*-jɛm haɣ lɛ-vɛt tʊ kɪ *ni*-k͡pʷɪsdi hʊw tʊ *ni*-mat.
3sg AO:3sg-climb up on-stone then 3sg AO:3sg-fall down then AO:3sg-die
‘He climbed up the rock, then he fell down and died.’

This use as a sequential marker in a string of events is ubiquitous in Mwotlap. Yet this form is also required in many other contexts which cannot be reduced to this explanation.

The Aorist is used for generic sentences, such as definitions or procedure descriptions – that is, utterances referring to a timeless event that bears no connection with any specific situation:

(6) “nɛ-ŋ͡mʷjajaj” ɛ, nɪk ɛtɛt hɛjlʊ van aɪ.
STAT-transparent TOPIC 2sg AO:look~IPFV through thither in.it
‘*Transparent* (means that) you see through it.’

(7) na-mtɛ *ni*-jɔj, na-tak͡pʷŋ͡mʷɪ *ni*-jɛjɛj,
ART-your.eyes AO:3sg-sink ART-your.body AO:3sg-shiver

nɪk mat ŋ͡mʷʊl …
2sg AO:die return
‘[with malaria] your eyes sink, your body shivers, you faint …’

Another example where Aorists point to virtual events whose time coordinates are left indefinite, is their use in conditional clauses (note that (8) is identical to (4) above).

(8) nɔk hajvɛɣ l-ɪŋ͡mʷ nɔnɔn ɛ kɪ *ni*-ɛtsas nɔ.
1sg:AO enter in-house his TOPIC 3sg AO:3sg-see 1sg
‘(Suppose) I came into his house (then) he would see me!’

In many cases however, the Aorist does relate to a specific situation, which may be the moment of utterance. This happens, for instance, when it represents an event as imminent:

(9) mahɪ *ni*-k͡pʷʊŋ ɪɣɪn.
place AO:3sg-night now
‘Night is about to fall.’

The imminence of the event is sometimes factual, as in (9), but quite often it is the speaker’s own projection. The Aorist thus takes on modal values, and encodes intent, optative, instructions or commands:[4]

(10) nɪk ɣɛn mɛj nʊk, nɔk ɣɛn mɛj ɣɪn.
2sg AO:eat the.one there 1sg:AO eat the.one there
‘You eat this one, I'll eat that one.’

(11) kɪ *ni*-ŋ͡mʷʊl lɛ-pnʊ nɔnɔn.
3sg AO:3sg-return in-village his
‘Let him return to his village!’ [or: ‘he returned …’, see (5)]

Strictly speaking, the Aorist cannot be said to inherently entail such illocutionary forces as desiderative or imperative, because it is also used in plain declarative sentences. In other words, just as it does not by itself convey any indication of time, it is also underspecified with regard to modality: it is found in statements as much as in hypotheses, commands or optatives. Both the time coordinates and the modal value of the Aorist thus need to be inferred from prosodic clues, and from the discourse context.

This semantic underspecification with regard to time and mood explains why the Aorist (or its derivative the Prospective) is required in modality-bound subordinate clauses: e.g., clause complements of verbs of will or manipulation, as well as purposive and consecutive clauses.

[4] This modal function accounts for the formal links between the Aorist, and the three modal markers derived from it (Prospective; Prioritive; Imperative).

(12) nɔ nɛ-mjʊs sɔ *nɔk* (sɔ) in ni-ti.
1sg STAT-want COMP 1sg:AO (PROSP) drink ART-tea
'I want to drink some tea.' [lit. 'I want that I drink$_{AO}$…']

(13) nɔ mʊ-mʊk nɛ-vɛt l-ɛp tʊ kɪ *ni*-vɛj.
1sg PRF-put ART-stone in-fire then 3sg AO:3sg-red.hot
[*purposive*] 'I laid the stones on the fire so that they become$_{AO}$ red-hot.'
[*consecutive*] 'I laid the stones on the fire so they became$_{AO}$ red-hot.'

The semantic incompleteness of the Aorist thus makes it particularly compatible with certain forms of syntactic dependency, in a way reminiscent of the subjunctive of Indo-European languages.

If a dynamic verb is reduplicated, it acquires imperfective aspectual properties, including when combined with the Aorist. This means (Comrie 1976) it may take either a *habitual* reading or a *progressive* one:

(14) *nɔk* jap hij tita mino.
1sg:AO write to mother my
simple verb: perfective interpretation
[*sequential*] '(then) I wrote to my mother.'
[*intent/optative*] 'Let me write to my mother!' …

(14′) *nɔk* japjap hij tita mino.
1sg:AO write~IPFV to mother my
reduplicated verb: imperfective interpretation
[*habitual*] 'I write to my mother (every day…)'
[*progressive*] 'I'm writing to my mother.'

The absence of reduplication in (14) gives the verb a perfective reading, which makes it compatible with the various aspectual values reviewed so far for the Aorist: sequential, intent, etc. Conversely, reduplication in Mwotlap (François 2004) has the power to convert a perfective into an imperfective, which disrupts the impact of the Aorist marking altogether. Arguably, the latter then functions as a neutral aspect marker, whose role is simply to state the imperfective process (habitual or progressive) in relation to the context.

2.2.2 Defining the underlying mechanism

Despite the impressive polyfunctionality of this TAM category, it is possible to identify a constant aspectual pattern behind the variety of its contextual meanings. In all cases, the Aorist consists in representing a new event considered in itself, that is, regardless of its deictic coordinates in terms of tense or modality.

Precisely because it lacks any inherent deictic reference, this indeterminate event needs to be connected to an external point of reference—its 'anchor'—in order to receive proper pragmatic interpretation. Quite often, the anchoring situation is easy to retrieve from the context. For example, in a string of successive events, it corresponds to the end of the previous event (4, 5). In a subordinate pattern, the dependent event will hook onto the coordinates of the main clause (12, 13). In many cases, the default reference point will be the utterance situation, whether the new event that is supposed to cling to it is presented as a statement of fact (9, 14') or as the speaker's projection (10, 11, 14). Finally, it sometimes happens that this 'orphan' event in search of situational anchoring finds none, and remains

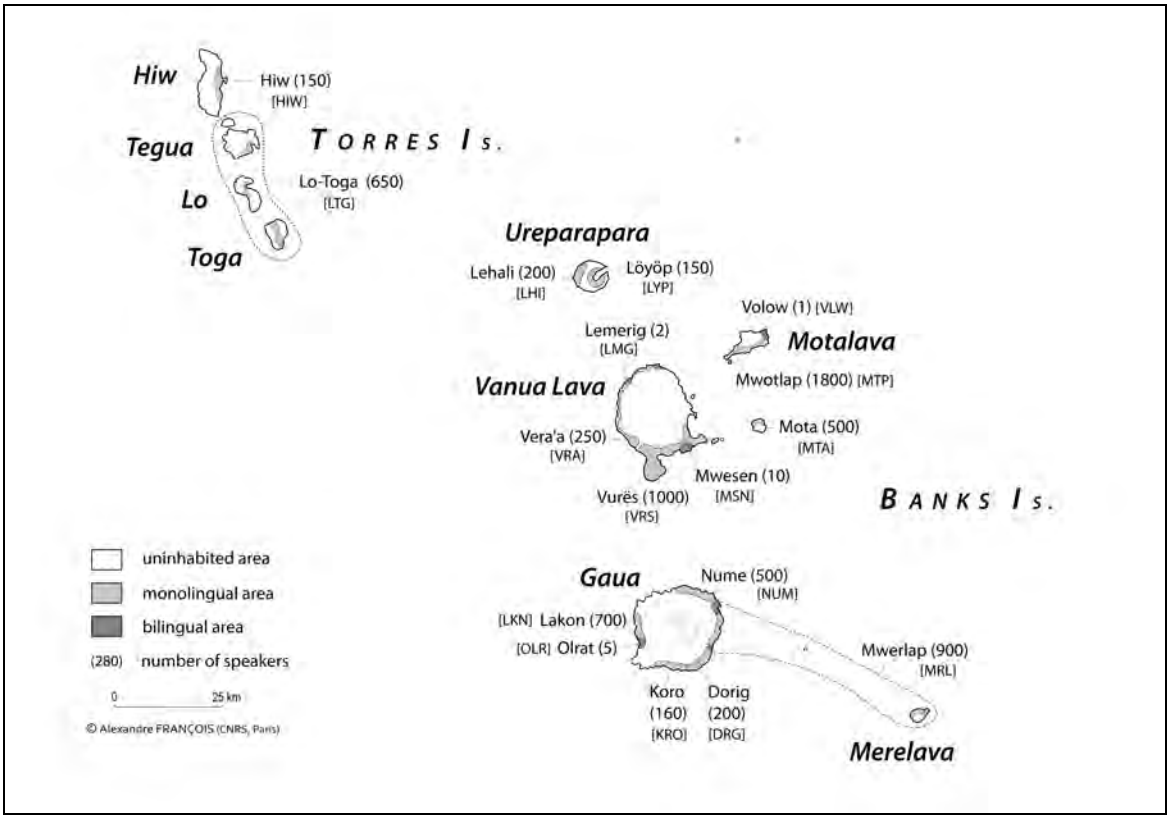
suspended in time: this is what happens when the Aorist points to a timeless event with no connection to any specific situation, as in generic statements (6, 7) or hypotheses (8).

Typologically speaking, the term 'aorist' has been used with various senses, and sometimes inconsistently, across language descriptions. However, the aspectological tradition that has developed, especially in France, after Benveniste (1966) and Culioli (1978), has now solidly established the notions of '*aoriste*' or '*aoristique*', as a verbal aspect whereby the depicted event is disconnected from the situation of utterance. Similar examples of 'aorist' have been described for several languages, such as Coptic (Depuydt 1993), Wolof (Robert 1996) or Berber (Galand 2003). A full typological survey of the aorist aspect still needs to be carried out.

3 The morphology of the Aorist in northern Vanuatu languages

Now that the semantics of the Aorist have been observed for Mwotlap on a synchronic, language-internal basis, it becomes possible to observe whether its neighbors of north Vanuatu possess a similar aspect category, and if so, how they encode it morphologically. This observation might help trace the formal history of Mwotlap's Aorist, and especially of the unusual alternation between the two 1sg pronouns *nɔ* and *nɔk*.

Since 2003, my field investigations have precisely involved the firsthand study of all the languages of the Banks and Torres groups, of which basically nothing was known to date. Map 1 locates these seventeen languages; it indicates their current number of speakers, together with the three-letter abbreviations I propose to use for them. The remainder of this section will summarise the results of this survey with regard to the Aorist aspect.



Map 1: The languages of north Vanuatu

3.1 South Banks

The TAM systems observed in the five languages of Gaua, together with Mwerlap, differ significantly from that of Mwotlap. In particular, the semantic spectrum of Mwotlap's Aorist, instead of being encompassed by a single marker, is divided in these languages into two, three or even four distinct categories, each language showing its own particular distribution (Table 1).

Table 1: Equivalents to Mwotlap's Aorist in the six south Banks languages

<table>
<tr><th></th><th>MTP ex</th><th>MRL</th><th>NUM</th><th>DRG</th><th>KRO</th><th>OLR</th><th>LKN</th></tr>
<tr><td>Sequential</td><td>(4, 5)</td><td>ti</td><td>tɔv</td><td>sɔ</td><td rowspan="2"></td><td>Ø</td><td>tɪ</td></tr>
<tr><td>Generic
(definitions, procedures)</td><td>(6, 7)</td><td rowspan="3"></td><td>vɛ</td><td rowspan="2"></td><td rowspan="2">v-</td><td rowspan="3"></td><td rowspan="3"></td></tr>
<tr><td>Subjunctive
(dependent, conditional)</td><td>(8, 12, 13)</td><td rowspan="2">sV-</td><td rowspan="2">tɛ</td><td rowspan="2">s-</td><td rowspan="2">(ɣ)a</td></tr>
<tr><td>Prospective
(optative, commands …)</td><td>(9, 10, 11, 14)</td><td>s-</td><td>(ɣ)a</td></tr>
<tr><td>Imperfective
(habitual, progressive)</td><td>(14')</td><td></td><td>tɛ… ti</td><td>t-… ti</td><td>t-… ti</td><td>tɪ-… ti</td><td>tɪ-… tɔ</td></tr>
</table>

It would be a matter for complex discussion to decide which of these morphemes should be properly labeled 'Aorist', and which ones should receive a name of their own. For example, in Dorig, it is safe to call *sɔ* a Sequential, and *t-… ti* an Imperfective. As for *s-*, the union of 'generic', 'subjunctive' and 'prospective' could be tagged Aorist, in the sense of 'deictically indeterminate new event', as defined above for Mwotlap. Yet it could as well, and perhaps more accurately, be called *Irrealis* or *Virtual*—a choice impossible in Mwotlap due to both the sequential and the imperfective uses.

In sum, none of these languages possess a proper aorist, in the sense defined for Mwotlap. Furthermore, all markers in Table 1 are invariable prefixes or proclitics, used for all persons. Their forms resemble neither MTP *nɔk* nor *ni-*, the origins of which will have to be sought elsewhere.

3.2 Central Banks

The ten remaining languages of the Banks and Torres are more promising. Indeed, each of these languages possesses a TAM category which essentially matches the Aorist of Mwotlap, encompassing all the functions of Table 1, from 'sequential' to 'imperfective';[5] I shall therefore use the label 'Aorist' everywhere. And, crucially, in each language, its formal marking depends on the person of the subject, in a way reminiscent of Mwotlap.

Let us first observe the three languages located in the central part of the Banks Islands: Mota, Mwesen and Vurës. Taking the verb 'see' (MTA *ilo*, MSN-VRS *ɪl*) as an example, Table 2 illustrates the behavior of subject markers for the Aorist, in comparison with an

[5] In order to demonstrate this, a full set of examples should ideally be provided for each language. Unfortunately, this is impossible here due to considerations of space.

ordinary TAM marker—in this case, the Perfect.[6] The pattern for 1inc:pl, which is given here, exemplifies the twelve non-singular forms.

Table 2: Aorist inflections in three central Banks languages

Language	TAM	1sg	2sg	3sg	1inc:pl
Mota	Perfect	nau *me* ilo	ko *me* ilo	ni *me* ilo	nina *me* ilo
	Aorist	**na** ilo	**ka** ilo	**ni** ilo	nina **a** ilo
Mwesen	Perfect	na *mɛ* ɪl	nɪk *mɛ* ɪl	nɪ *mɛ* ɪl	nin *mɛ* ɪl
	Aorist	na **na** ɪl	nɪk **a** ɪl	nɪ **ni** ɪl	nin **a** ɪl
Vurës	Perfect	nɔ *mɪ*-ɪl	nɪk *mɪ*-ɪl	nɪ *mɪ*-ɪl	nɪn *mɪ*-ɪl
	Aorist	**na** ɪl	nɪk **i** ɪl	nɪ **ni** ɪl	nɪn **a** ɪl

Taking only the non-singular forms, we would simply have an invariant clitic *a* 'Aorist' behaving like other TAM markers. But the singular makes the description more complex, because the marking of the Aorist differs according to the person of the subject. To be precise, two distinct cases are attested:

a. The pronoun itself remains unchanged, but the Aorist clitic presents allomorphic variation according to the person of the subject. Thus for Mwesen, the Aorist is *a* for all persons, but *na* for 'AO:1sg' and *ni* for 'AO:3sg'.

b. The sequence {pronoun + TAM marker} found with other tenses is replaced by a single portmanteau clitic that incorporates person- and TAM-marking. Thus in Mota, *na* should be properly glossed '1sg:AO', and *ka* '2sg:AO'.[7]

Vurës combines the two patterns: (a) for 2sg and 3sg, but (b) for 1sg. In fact the same complexity was found in Mwotlap, where *nɔk* was to be analyzed as an aspect-indexed pronoun ('1sg:AO'), but *ni-* as a person-indexed aspect prefix ('AO:3sg').

Now, MTP *ni-* is clearly the same morpheme as *ni* in these three languages. Furthermore, a connection can be drawn between that *ni* 'AO:3sg' and the form of the free pronoun for 3sg *ni* in Mota, Nume, Dorig and Koro. In several languages of north Vanuatu, the 3sg pronoun (*nɪ, niə*…) reflects an earlier form *nia '3sg', itself connected with *ni*. This formal connection has been blurred in Mwotlap, where the 3sg pronoun is now an innovative *kɪ*.

These first findings thus shed light on our initial puzzle. Yet still nothing can be said about the strange form *nɔk* in Mwotlap: where does this /k/ come from? The answer will appear as we continue our survey further north.

3.3 North Banks

Not surprisingly, a system much closer to Mwotlap can be found in Volow, an extinct dialect formerly spoken on the same island, and passively remembered by a handful of people. The structures of the two dialects are so parallel that the only differences lie in the phonological forms of the markers: to the alternation between MTP *nɔ* and *nɔk* corresponds a

---

6 For each language, the first row translates as 'I have seen' …; the second row as 'Let me see', etc.

7 For 3sg in Mota, the Aorist can be identified either as a *Ø* clitic (commuting with *a*) or as incorporated in *ni* (commuting with *ka*).

pair of forms *nɛ* '1sg' versus *nɛŋ* '1sg:AO' (see Table 3 below).[8] But precisely because Volow is so close to Mwotlap, it is of little help in our investigation.

More instructive findings come from the four languages of the northwest Banks area: Lehali, Löyöp, Lemerig and Vera'a. Unlike the three languages of Table 2, they do possess a trace of the velar /k/ which is found in the 1sg pronoun of Mwotlap. But, interestingly, instead of being part of the pronoun itself, the consonant /k/ is separable from it, and prefixed to the verb. This becomes obvious when the pronoun and the verb are separated by another morpheme. Compare the Prospective of Mwotlap with its form in Lehali and Vera'a:

(15) MTP *nɔ**k*** sɔ mitij.
1sg:AO PROSP sleep

LHI nɒ dɛ **k-** mutuj.
1sg PROSP AO:1sg- sleep

VRA nɔ sɔ **k-** miʔir.
1sg PROSP AO:1sg- sleep
'I'd like to sleep.'

This syntactic test makes it easy to define the boundary between the personal pronoun proper and the (person-conditioned) TAM-marker. Unlike Mwotlap, these northwest Banks languages thus treat the 1sg Aorist marker in the same way as their 3sg, as a prefix to the verb:

(16) MTP *kɪ* sɔ ***ni-*** mtij.
3sg PROSP AO:3sg- sleep

LHI *kɛ* dɛ ***n-*** mutuj.
3sg PROSP AO:3sg- sleep

VRA *di* sɔ ***nɛ-*** miʔir.
3sg PROSP AO:3sg sleep
'He'd like to sleep.'

Lemerig does not allow any element between the pronoun and the (inflected) verb, which makes it impossible to conduct the test illustrated by (15). For example, 'Let me sleep' will take the ambiguous surface form /nœ**k**miʔir/, which could be parsed *nœk miʔir* <1sg:AO | sleep> as in Mwotlap, or *nœ k-miʔir* <1sg | AO:1sg-sleep> as in Vera'a. My Lemerig corpus shows 38 instances of a 1sg Aorist, out of which 37 show this ambiguity. Luckily, one sentence has two Aorists chained together, a context where the personal pronoun may be dropped. This single example gives the solution to the puzzle, and highlights the structural difference between Lemerig and Mwotlap:

(17) LMG *nœ* ***k***-œn sur ɛ (*nœ*) ***k***-miʔir.
1sg AO:1sg-lie down LNK (1sg) AO:1sg-sleep

MTP *nɔk* ɛn hij ɛ (*nɔk*) mitij.
1sg:AO lie down LNK (1sg:AO) (AO:)sleep
'Let me lie down and sleep.'

[8] Note that the correspondence between MTP /k/ and VLW /ŋ/ syllable-finally is regular, and reflects a former prenasalised voiced stop [ᵑg] (noted **g*).

Finally, the case of Löyöp is slightly tricky, because it is a hybrid of both patterns. On the one hand, just like its neighbours but unlike Mwotlap, Löyöp has kept a 1sg Aorist prefix of the form *k*-. This form appears when the clause lacks the free pronoun *nø*:

(18) LYP nø nɣɛ møjøs sɛ ***k***-sʊwjɪ n-k͡pʷɛŋ.
1sg STAT want SUB AO:1sg-cast ART-net
'I want to go net-casting.' [lit. I want that I_{AO} cast the net]

However, Löyöp has also taken the same path as Mwotlap, in that the combination of the free pronoun *nø* and of the *k*- prefix has been resegmented, giving rise to a new, unanalysable pronoun *nʊk*, with an unpredictable vowel:

(19) LYP *nø* tɛ pɪl ʧɛ, ***nʊk*** dɔn sɛ n-jø-k.
1sg NEG_1 steal NEG_2 1sg:AO think SUB ART-CLPOS-1sg
'I didn't steal it, I thought it was mine.'

Taking the verb meaning 'see' again as an example,[9] Table 3 shows the Aorist morphology for the six languages under comparison here—including Mwotlap and Volow—thereby covering the whole 'north Banks' area. The hybrid case of Löyöp appears in the middle.

Table 3: Aorist inflections in six north Banks languages

Language	1sg	2sg	3sg	1inc:pl
Lemerig	nœ **k**-ʔɛt	næk (ø-) ʔɛt	ti **n**-ʔɛt	ɣæt (ø-) ʔɛt
Vera'a	nɔ **k**-ʔɪn	nikɪ (ø-) ʔɪn	di **nɛ**-ʔɪn	ɣidɪ **k**-ʔɪn
Lehali	nɒ **k**-ɛt	nɛk (ø-) ɛt	kɛ **n**-ɛt	ɣɛn (ø-) ɛt
Löyöp	**k**-ɛt / nʊ**k** ɛt	niŋ (ø-) ɛt	kjɛ **n**-ɛt	jɛn (ø-) ɛt
Mwotlap	nɔ**k** (ø-) ɛt	nɪk (ø-) ɛt	kɪ **ni**-ɛt	ɣɪn (ø-) ɛt
Volow	nɛ**ŋ** (ø-) ɛt	nɪŋ (ø-) ɛt	gɪ **n**-ɛt	ɣɪn (ø-) ɛt

The problem raised by *k*- in Vera'a non-singular forms will be addressed in §4.2.

3.4 The Torres Islands

3.4.1 Two sets of personal markers

I will end this survey of Aorist markers in north Vanuatu with the two languages of the Torres group. In comparison with the Banks languages, the Aorist inflection in Hiw and Lo-Toga is morphologically richer. Not only are there specific (non-zero) morphemes for each person in the singular, but also for non-singular subjects, including different forms for the dual and for the plural (the Torres languages have lost the trial).

Table 4 lists the complete sets of personal subject markers for the two languages. On the left are given the free pronouns; on the right, the set of person-indexed Aorist clitics.

[9] Once again, the case of '1inc:pl' is meant to illustrate all non-singular forms.

Table 4: Full pronouns versus Aorist clitics in the two Torres languages

	Hiw		Lo-Toga	
	full pronoun	Aor. clitic	full pronoun	Aor. clitic
1sg	nɔkə	kə	nɔkə$_{(LO)}$/ nɛkə$_{(TGA)}$	kə
2sg	ikə	wət ~ wək	nikə	wə
3sg	ninə	nə	niə	ni
1in:du	tə^{g}$_{L}$ə		ṭor	
1ex:du	kama^{g}$_{L}$ə	^{g}$_{L}$ə	kəmər	or
2du	kimi^{g}$_{L}$ə		kəmor	
3du	sə^{g}$_{L}$ə		hor	
1in:pl	titə	tə	ṭəɣə$_{(LO)}$/ ɣiṭə$_{(TGA)}$	
1ex:pl	kama	–	kəmɛ(m)	ɣə
2pl	kimi	–	kəmi	
3pl	sisə	sə	nihə	

The first obvious observation is that these two languages possess interesting clues for our study: the form of the 3sg clitic (*nə*/*ni*) recalls the prefixes *n-* or *ni-* we saw in the Banks languages; and the 1sg clitic *kə* is reminiscent of the prefix *k-* shown in Table 3 above. Finally, the 1sg pronoun *nɔkə* strikingly resembles Mwotlap *nɔk*, a point which warrants a discussion of its own (see §4.2.2).

3.4.2 TAM markers or light pronouns?

Just as in Banks languages, the function of the clitics of Table 4 is essentially to encode a TAM category, the Aorist. This status is proven by the comparison of (20) and (20´). Both sentences show serial verb constructions, one with the Potential, the second with the Aorist.

(20) LTG kəmi ***si*** ŋ^{w}ulə ***si*** mətur.
2pl POT return POT sleep
[*Potential*] 'You (pl) may go back and sleep.'

(20´) LTG kəmi ***ɣə*** ŋ^{w}ulə ***ɣə*** mətur.
2pl AO:PL return AO:PL sleep
[*Aorist*] 'You (pl) go back and sleep!'

But there is further complexity. Amongst the clitics of Table 4, only two (*ni* and *ɣə* in Lo-Toga, none in Hiw) may be immediately preceded by a free pronoun, as in (20') *kəmi ɣə*. All other clitics must be deleted in presence of the free pronoun, in which case the latter is directly followed by the verb. As a result, most Aorist sentences, when they include the free pronoun, appear to be unmarked (or zero-marked) for TAM. Conversely, the clitics are restricted to those clauses that lack a free pronoun. This happens typically in a string of clauses, when the pronoun is mentioned only with the first verb [see (17)]:

(21) LTG *nikə* (Ø) ŋ^{w}ulə ***wə*** mətur.
2sg (AO) return AO:2sg sleep

HIW *ikə* (Ø) ŋ^{w}ujə ***wət*** miti^{g}$_{L}$.
2sg (AO) return AO:2sg sleep
'You (sg) go back and sleep!'

A superficial analysis of (21) would probably have posited only one marker for the Aorist (*zero*), and then two sets of pronouns: 'heavy' pronouns for the main or first clause in a string, followed by 'light' pronouns in secondary and other dependent clauses. In that framework, it would have made sense to label these clitics 'secondary subject pronouns'. This interpretation is appealing, and could perhaps be proposed for Hiw; but in Lo-Toga, it seems to be contradicted by (20'). For the sake of consistency, it is thus safer to analyze *wə* in (21) not as a personal pronoun, but as a (person-indexed) aspect clitic. For most subjects,[10] deletion rules must be formulated, whereby the sequence {pronoun+clitic} simplifies to {pronoun}, e.g. **nikə wə* V ⇒ *nikə* V.

Because Hiw operates this deletion rule for all its pronouns, its Aorist clitics seldom show up in fluent speech, as they are restricted to subordinate or secondary clauses; and even in that case they are optional, being often replaced by the full pronouns. The situation is very different in Lo-Toga, where the clitics are extremely productive, and massively represented in my corpus. This productivity of Lo-Toga clitics has two reasons: first, the two clitics *ni* and *ɣə* cannot be deleted, and are pervasive in speech; second, each clitic also appears as a constituent element in three compound TAM markers historically derived from the Aorist: Prospective ⟨*tɛ* + Cl.⟩, Time Focus ⟨Cl. + *akə*⟩, and Future ⟨*tɛ* + Cl. + *akə*⟩ – e.g., *nikə* ***tɛ w' akə*** *mətur* 'you will sleep'.

In sum, in Hiw and Lo-Toga, one identifies a clause as Aorist either because it displays an Aorist clitic, or because it consists of {free pronoun + zero-marked verb}. For example, the Lo-Toga sentence (21) shows two Aorist predicates: ***nikə*** *ŋʷ̃ulə* 'you return$_{[AO]}$' and ***wə*** *mətur* 'you sleep$_{[AO]}$'.

4 The historical perspective

The eleven languages endowed with a genuine aorist (§3.2 to §3.4) show such solid formal similarities that they obviously share a common history. In this section I will endeavor to reconstruct a set of aorist markers for their common (post-POc) ancestor language. Logically, this protolanguage should be Proto North-Central Vanuatu, the proposed ancestor for the majority of Vanuatu languages (Clark 1985). But since my reconstruction is reflected only in the northernmost languages of this family—those spoken in the Banks and Torres groups—it could well represent a subgroup within NCV, the precise limits of which would need to be confirmed by further diagnostic evidence.

4.1 Reconstructing the set of Aorist proclitics

First, this pre-modern system possessed a set of free pronouns. They can be reconstructed for these northern Vanuatu languages (Clark 1985; Lynch and Ozanne-Rivierre 2001:38): **nau* '1sg'; **nigo* '2sg'; **n(a)ia* '3sg'; **kida* '1inc:pl'; **ga(ma)mi* '1exc:pl'; **gamuyu* '2pl'; **n(a)ira* '3pl'.

The system can also be reconstructed with a set of person-indexed aorist markers, consisting of proclitics preceding the verb. Taking into account the phonological history of this area—especially the phenomenon of vowel reduction (François 2005)—the most plausible reconstructions are as follows:

---

[10] In Lo-Toga, this deletion rule applies to 1sg, 2sg, and dual forms. To this list, one must add *ʈəɣə* '1inc:pl' in the Lo dialect of Lo-Toga: compare Toga *ɣiʈə ɣə ŋʷ̃ulə* with Lo *ʈəɣə* (Ø) *ŋʷ̃ulə* 'Let's go back'.

- 1sg: all languages point to the 'nasal grade' *g [ⁿg] (see fn.8). As for the vowel that followed this consonant, comparative evidence (see below) suggests it must have been /u/, hence a protoform ***gu**.
- 2sg: a single vowel, probably ***u** (reflected as /w-/ in the Torres; /i/ in Vurës; *zero* in most languages).
- 3sg: *nV, probably ***ni** (reflected as /ni/, /nə/, /nɛ/ or /n-/).
- non-singular: the modern forms /ɣə/, /a/ and Ø suggest a reconstruction ***(k)a**.[11] The dual forms of the Torres may result from a local innovation, perhaps **(k)a-ru (?).

The semantic array reconstructible for this set of aorist clitics most probably coincided with the observations made for modern Mwotlap (§2.2) and its neighboring languages. Their function was to construe a 'deictically indeterminate new event'—a definition which encompasses the functions of *sequential*, *generic*, *subjunctive*, *prospective* and (with reduplication) *imperfective*.

Syntactically, these proclitics {*gu, *u, *ni, *(k)a} occupied the same slot as other TAM markers. They were preceded by the free pronoun in main clauses, or in the first clause of a chain (serialised verbs, narratives); yet they appeared on their own in dependent or secondary clauses (e.g., same-subject sequential clauses). It is thus possible to reconstruct sentences such as (22):

(22)

*nau	*gu*=	mule	*gu*=	maturu
1sg	AO:1sg=	return	AO:1sg=	sleep

'So I went back and slept.' ~ 'Let me go back and sleep!' …

*nigo	*u*=	mule	*u*=	maturu
2sg	AO:2sg=	return	AO:2sg=	sleep

'So you went back and slept.' ~ 'You go back and sleep!' …

*nia	*ni*=	mule	*ni*=	maturu
3sg	AO:3sg=	return	AO:3sg=	sleep

'So he went back and slept.' ~ 'Let him go back and sleep!' …

*kida	(*k*)*a*=	mule	(*k*)*a*=	maturu
1inc:pl	AO:non.sg=	return	AO:non.sg=	sleep

'So we went back and slept.' ~ 'Let's go back and sleep!' …

As far as their origin is concerned, the singular forms {**gu*, **u*, **ni*} are reminiscent of two sets of personal markers:[12]

- the 1st, 2nd and 3rd person singular forms of the possessive suffixes, whose protoforms in north Vanuatu are {**-gu*, **-u*[13]/*-mu*, **-na*}, from POc {**-gu*, **-mu*, **-ña*};

---

11 The Sungwadia language of north Maewo also has a series of aorist prefixes (Agnès Henri pers. comm.). The plural form is *ɣe*, which tends to confirm the consonant in our reconstruction *(k)a.

12 A formally very similar set of subject pronouns 1sg *gu*, 2sg *mu*, 3sg *ni*… is attested in modern Marovo, a Northwest Solomonic language spoken in the Solomon islands (Evans 2008:404).

13 For 2sg **-u*, see François (2005:486). This local variant of the suffix may have played some influence upon the 2sg proclitic **u*.

- one of the sets reconstructed for the POc subject proclitics, namely {**ku*, **mu*, *(*y*)*a*/*ña*}, which ultimately reflect PMP genitive pronouns (Blust 1977; 2003). As for the non-singular prefix *(*k*)*a*, it is reminiscent of a subject clitic **ka[i]* '1exc:pl' reconstructed for some Oceanic interstage languages (Lynch et al. 2002:68).

Given the nature of the Aorist clitics, the subject clitics are a more likely source than the possessive suffixes. There is still some debate about the precise function of these POc proclitics. According to Kikusawa (2005), they retained in POc their earlier function as ergative subjects (i.e., 'A' in divalent clauses) as opposed to intransitive subjects ('S' in monovalent clauses). Lynch et al. (2002:68) suggest that this function was probably 'being lost when POc broke up'.

But one still has to explain how a set of ordinary subject pronouns should have evolved into TAM-marking clitics. A tentative hypothesis would suggest these subject markers once became specialised in subordinate or other dependent clauses while main declarative clauses eventually generalised the use of free pronouns. As a result, what were once genuine subject pronouns in clauses otherwise unmarked for TAM, eventually grammaticalised into subjunctive-like TAM markers. This hypothesis would account for the affinities of Aorist predicates with syntactic dependency, discourse backgroundedness, and TAM indeterminacy. But this is mainly speculation at this stage; the functional connection at stake here definitely warrants further investigation.

4.2 From the protosystem to modern languages

After tentatively reconstructing the protosystem of Aorist clitics in the protolanguage ancestral to the Torres and Banks groups (either PNCV or one of its branches), I will end this study with an overview of the various paths of evolution that historically led to the modern systems.

4.2.1 Phonological attrition and affixation

The phonological process of *unstressed vowel deletion*, which massively affected the languages of the whole Banks and Torres area (François 2005), explains why former *CV clitics are generally reflected as a single consonant in modern languages (*gu > /k/ or /ŋ/; *ni > /n/…), and also why the two vowel-only clitics are so often reflected as *zero*. Lemerig shows the expected reflexes in this regard:

	Protosystem	>	Lemerig
'let me go back'	*ˈnau *gu*=ˈmule	>	nœ **k**-mʊl
'you go back'	*ˈnigo *u*=ˈmule	>	næk Ø-mʊl
'let him go back'	*(ˈnia) *ni*=ˈmule	>	(ti) **n**-mʊl
'let's go back'	*ˈkida *a*=ˈmule	>	ɣæt Ø-mʊl

This vowel reduction process explains why the former clitics generally became prefixes. It also accounts for the formal convergence between 2sg **u* and the non-singular clitic **a*—variant of *(*k*)*a* in the form of *zero*. As for Mwesen and Mota (Table 2), this convergence rather results from the spread of the non-singular clitic *a* to 2sg (MTA *ka* <**ko a*); see also below for the case of 1sg.

Among the four clitics {**gu*, **u*, **ni*, **(k)a*}, only 3sg **ni* is preserved in the eleven languages endowed with a genuine 'aorist', from Hiw all the way down to Mota. On the other hand, **(k)a* seems to show the widest historical extension, as its reflexes are scattered in various places from Lo-Toga *ɣə* down to Lakon/Olrat *(ɣ)a* (Table 1).

4.2.2 The intricate destiny of 1sg **gu*

As for 1sg **gu*, it has survived in the eight (or nine: see below for Vurës) northernmost languages of the area, but has left little trace in any other language further south. To the best of my knowledge, the only other NCV language with a reflex of the subject marker **gu* is Tamabo, with *ku* '1sg' (Jauncey 2002:610).

The detailed evolution of this form **gu* in north Vanuatu languages is intricate. The former proclitic **gu=* is regularly reflected as a proclitic *kə=* in the Torres languages, or a prefix *k-* in northwest Banks languages (Table 3). The 1sg pronoun alternation (*nɔ*/*nɔk*) of Mwotlap—the starting point of this study—results from a process of reanalysis: the sequence /*nɔ* + *k*-/, with no intervening element, was so overwhelmingly frequent in speech, that it was eventually resegmented as *nɔk*; the former verbal clitic became accreted to the preceding pronoun, while the verb itself appeared in its bare form. The zero form taken by the Aorist with other persons (Table 3) probably added to the pressure towards morphological leveling. Volow followed a similar path, leading to an allomorphic variation between *nɛ* and *nɛŋ*; as well as Löyöp, with the forms *nɔ* and *nʊk*.

So far, the accreted form *nɔk* in Mwotlap has remained restricted to its original function—namely, 1sg subject of an Aorist predicate—without much affecting the regular form of the 1sg pronoun *nɔ* in other contexts. However, a slight tendency towards the expansion of *nɔk* is perceptible with other TAM categories. Besides the six markers that are formally derived from the Aorist (§2.1), the use of *nɔk* is increasingly frequent, albeit optional, for as many as ten TAM markers, historically unrelated to the Aorist—for example, the negative markers.

This tendency for the accreted form to gain ground over the original 1sg pronoun is only incipient in Mwotlap, but has reached its final stage in the two Torres languages. Due to their high frequency in discourse, the augmented forms—*nɔkə* in Hiw and Lo, *nɛkə* in Toga—have now become the ordinary 1sg pronoun regardless of the predicate's TAM-marking, and indeed for all functions (subject, object, etc.).[14] In other words, the pronoun resulting from the coalescence of **nau gu* has replaced **nau* in all positions.

Unlike Mwotlap, the coalescence of **nau* and **gu* in the Torres languages did not eliminate **gu* as an independent morpheme (*kə*) in the system. But the fact it was historically incorporated into the free pronoun explains why the two forms are incompatible (**nɔkə kə* is ungrammatical). Crucially, this is the key to many of the so-called 'deletion rules' which are required in the synchronic description of the Torres languages (§3.4.2). A similar process of accretion is the origin of the form *ṭəɣə* used in the Lo dialect, the only plural form in Lo-Toga to be incompatible with *ɣə* (see fn.10): *ṭəɣə* < *(ɣi)ṭə ɣə* < **kida ka*. Likewise, in Hiw, *ninə* has incorporated the clitic *nə* (*ninə* < **ni(ə) nə* < **nia ni*); *titə* has incorporated *tə*; *sisə* has incorporated *sə*; and so on. Hiw is the language where the contamination of the pronoun system with Aorist clitics has been maximal.

---

[14] The earlier forms **nɔ* and **nɛ* only survive vestigially, in the 1sg possessive of Lo (*mi-nɔ*) and Toga (*mi-nɛ*)—etymologically 'with me'.

Interestingly, the 1sg Aorist clitic **gu* has also undergone another reanalysis of a completely different kind. In modern Vera'a, the prefix *k-* encodes the Aorist not only for 1sg, but for all non-singular persons as well (Table 3). It looks as if this prefix were in the process of being reanalyzed as the general marker for Aorist, thereby losing its original connection with the first person; ironically, this is the exact opposite of the Torres evolution. Even 2sg and 3sg seem threatened by the expansion of VRA *k-* to all persons, judging by the attestation of *nik sɔ* ***k****-van* and *di sɔ* ***k****-van* as alternatives to the regular *nik sɔ* ***Ø****-van* and *di sɔ* ***nɛ****-van* (respectively 'you/he should go').

Finally, this generalisation of /k/ to non-singular persons in Vera'a possibly provides the explanation for a non-etymological /k/ found in several non-singular pronouns in Vera'a's closest neighbor, Vurës: *kɛmɛ****k*** '1exc.pl'; *kʊmʊrʊ****k*** '1exc.du'; *dʊrʊ****k*** '1inc.du'. If this hypothesis is right, then Vurës would combine the evolution patterns of Vera'a (spreading of **gu* to non-singular persons) and of the Torres languages (accretion of **gu* to the preceding pronoun, and generalisation of the augmented form to all functions, regardless of the verb's aspect). If this hypothesis is true, then these three non-etymological /k/ are the only trace left by **gu* in Vurës. Indeed, this language—like Mota and Mwesen—encodes its 1sg Aorist subjects with an innovative form *na* (Table 2), surely the result of the coalescence of *nɔ* <**nau* and *a* <*(*k*)*a*.

5 Conclusion

Out of the seventeen languages of north Vanuatu, eleven share an aspect category labeled 'Aorist', whose function is to represent a new event regardless of its deictic coordinates in terms of tense or modality. Despite their differences, the various morphosyntactic patterns that encode this Aorist can be shown to derive ultimately from a single protosystem: a set of four portmanteau proclitics {*gu-*u-*ni-*(k)a} combining aspect marking and subject agreement.

The reconstruction proposed in this paper not only helps to explain the development of these languages' TAM systems, but also proves indispensable in unravelling the historical morphology of personal pronouns in this part of Oceania. The next challenge is now to define the precise link—both formal and functional—that connects these four clitics to the set of personal pronouns which Robert Blust reconstructed for the remote ancestors of these languages.

References

Blust, Robert. 1977. The Proto-Austronesian pronouns and Austronesian subgrouping: a preliminary report. *University of Hawai'i Working Papers in Linguistics*, 9/2:1–15.

—— 2003. Three notes on early Austronesian morphology. *Oceanic Linguistics* 42–2:438–478.

Benveniste, Émile. 1966 [1959]. Les relations de temps dans le verbe français. In *Problèmes de Linguistique Générale*, vol. 1, 237–250. Tel. Paris: Gallimard.

Clark, Ross. 1985. Languages of North and Central Vanuatu: groups, chains, clusters and waves. In Andrew Pawley and Lois Carrington, eds *Austronesian linguistics at the 15th Pacific Science Congress*, 199–236. Canberra: Pacific Linguistics.

Comrie, Bernard. 1976. *Aspect. An introduction to the study of verbal aspects and related problems*. Cambridge: Cambridge University Press.

Culioli, Antoine. 1978. Valeurs aspectuelles et opérations énonciatives : l'aoristique. In J. David and Robert Martin, eds *Actes du colloque sur la notion d'aspect*, 182–193. Metz: Klincksieck.

Depuydt, Leo. 1993. *Conjunction, contiguity, contingency: on relationships between events in the Egyptian and Coptic verbal systems*. New York: Oxford University Press.

Evans, Bethwyn. 2008. Subject agreement in Marovo: diachronic explanations of synchronic conditions. *Oceanic Linguistics* 47–42, 383–408.

François, Alexandre. 2003. *La sémantique du prédicat en mwotlap (Vanuatu).* Collection Linguistique de la Société de Linguistique de Paris, 84. Paris, Louvain: Peeters.

—— 2004. La réduplication en mwotlap : les paradoxes du fractionnement. In E. Zeitoun, ed. *Les langues austronésiennes*. *Faits de langues* n°23–24:177–194.

—— 2005. Unraveling the history of the vowels of seventeen northern Vanuatu languages. *Oceanic Linguistics* 44–2:443–504.

—— 2007. Verbal aspect and personal pronouns: the history of aorist markers in north Vanuatu. Paper presented at *7th International Congress on Oceanic Linguistics* (COOL7), Université de Nouvelle-Calédonie, Nouméa, July 2007.

Galand, Lionel. 2003. L'aoriste berbère, l'aspect et les valeurs modales. In Jérôme Lentin and Antoine Lonnet, eds *Mélanges David Cohen*. Paris: Maisonneuve and Larose, 235–246.

Jauncey, Dorothy. 2002. Tamabo. In Lynch, Ross and Crowley, eds 2002:608–625.

Kikusawa, Ritsuko. 2003. A new view of the Proto Oceanic pronominal system. *Oceanic Linguistics* 42–1:161–186.

Lynch, John and Françoise Ozanne-Rivierre. 2001. Some shared developments in pronouns in languages of Southern Oceania. *Oceanic Linguistics* 40–1:33–66.

Lynch, John; Malcolm Ross and Terry Crowley. 2002. *The Oceanic languages*. Richmond: Curzon Press.

Robert, Stéphane. 1996. Aspect zéro et dépendance situationnelle: L'exemple du Wolof. In C. Müller, ed. *Dépendance et intégration syntaxique (Subordination, coordination, connexion)*. Linguistische Arbeiten. Tübingen: Niemeyer, 153–161.

15 *Austronesian typology and the nominalist hypothesis*

DANIEL KAUFMAN

1 Background[1]

The idea that a deep connection exists between verbs and nouns in Austronesian languages was already present among early Austronesianist grammarians such as van der Tuuk (1864–67) and Adriani (1893) (Blust 2002). Among twentieth century scholars, this view was echoed by Lopez (1928:51) concerning Tagalog: 'the quasi verb is not a real verb, for it is treated like a nomen in the sentence and the enlargements, according to their forms, are considered as attributes and not as objects'. More recently, the link between notional verbs and nouns has yet again been underscored by Capell (1964), Naylor (1975, 1980, 1995), Starosta, Pawley and Reid (1982), De Wolf (1988) and Himmelmann (1987, 1991, 2008), among others.

One primary basis for this is the identity in case marking between possessors and agents of non-actor focus verbs. In this article, I argue that there is in fact far more to recommend the nominal view of Austronesian verbs than the simple case syncretism. I extend arguments that several defining features of morphosyntactically conservative Austronesian languages[2] are intimately connected on the basis of a historical reanalysis of nominalisations to verbal

[1] This paper represents one side of on-going work that I have presented at the Zentrum für Allgemeine Sprachwissenschaft, the CUNY Graduate School, and AFLA XV. I am indebted to those audiences and especially to Nikolaus Himmelmann for detailed comments on a previous draft. It is truly an honor to present it to Bob Blust who has been a tremendous source of both inspiration and encouragement to me.

[2] I am purposefully vague in employing this designation for the following reason. Most of the features I discuss here for '(morphosyntactically) conservative languages' apply throughout the Philippine languages and are found in certain Formosan languages as well, suggesting a reconstruction higher than Proto-Malayo-Polynesian (PMP). Nonetheless, there do appear to be some significant differences in certain other Formosan languages and it is not clear whether these differences should be treated as innovations or retentions. Ross (this volume) argues convincingly that a set of morphosyntactic differences in Puyuma, Rukai and Tsouic represent retentions with the consequence that the features referred to here as conservative may have been post-PAn innovations. Thus, at this stage, I refrain from associating the conservative features discussed here with a particular subgroup.

Alexander Adelaar and Andrew Pawley, eds
Austronesian historical linguistics and culture history: a festschrift for Robert Blust, 197–226.
Pacific Linguistics, 2009.
© This edition vested with Pacific Linguistics.

categories as first suggested by Pawley (1977) and Starosta, Pawley and Reid (1982) (henceforth SPR). The contribution of the present paper is to show that many of the synchronic properties of morphosyntactically conservative languages can be explained if we take apparent verbs to still be underlyingly nominal. Moreover, we can explain what appears to be a large scale convergence of typological features among Indonesian languages which do not comprise an exclusive genetic subgroup by understanding these features to be the natural outcome of the reemergence of the verbal category.

The paper is organised as follows: in §2, I discuss the place of Philippine languages within the typology of case syncretisms and alignment types. In §3, I explore the idea of reinterpreting apparent verbal predication as nominal predication showing how this accounts for distributional facts, extraction asymmetries, coordination facts, and a curious asymmetry between two kinds of imperatives. In §4, I show how the primary cues for nominal oriented syntax eroded in Indonesian languages leading to the reemergence of a truly verbal category. In §5, I discuss some problems for the nominalist hypothesis and I conclude in §6 with suggestions for further research.

2 Alignment systems

Austronesian languages are probably best known for their rich voice system, referred to in the earlier literature as the 'focus system' (see Blust 2002 for the history of this terminology). Whereas many language families of the world possess rich case systems (Uralic, Kartvelian, Indo-European, to name a few), a similar richness in the voice system, as seen in Austronesian, is exceedingly rare. What makes this type of system even more remarkable is the fact that all voices in Philippine languages tend to be equally marked, morphologically speaking (Ross 1995a:737). In other words, the typologically unusual voices (i.e., the instrumental/conveyance voice and locative voice) do not appear to take one of the ostensibly more basic voices (i.e., patient voice and actor voice) as their base. The modern reflexes of the PAn voice morphemes shown in (1) (following Ross 1995a) typically do not co-occur with each other, and thus appear to form a paradigm of sorts.

(1)

PAn Form	Function
*<um>	Actor voice
*-en	Patient voice
*-an	Locative voice
*Si-	Instrumental/Conveyance voice

Aldridge (2004) and Ross (2006), however, do argue for an applicative analysis of the locative and conveyance voices. On this view, there are only two true voices, the patient/undergoer voice which forms canonical transitive clauses and the actor voice which is employed for intransitive and antipassive type clauses. The locative and instrumental/ conveyance morphemes are analyzed as applicatives which are added to the undergoer voice to promote adjuncts to subject. There are, however, several difficulties with such an analysis which can be noted here.

First, we do not expect that an applicative affix (i.e., PAn *-*an*, **Si*-) would replace a transitive voice affix (i.e., *-*en*), but this clearly appears to have been the situation from the beginning in Austronesian. Second, the two putative applicatives cannot create new objects, but are rather restricted to creating new subjects. As noted by Ross (this volume, fn.4) and argued for by Aldridge (2004), it may be possible that applicatives in ergative languages

behave differently in promoting applicative objects directly to subject/absolutive. Nonetheless, it is odd for there to be a ban on applicatives co-occurring with the actor voice/antipassive, as this is seen to occur in other robustly ergative languages. Third, the two putative applicatives cannot cooccur with each other, a common possibility afforded to applicatives cross-linguistically. Finally, it is not clear that reflexes of **-an* and **Si-* can be considered any more valency-increasing than reflexes of *-en*. Exemplifying with Tagalog, notionally monovalent roots can typically become bivalent simply by the addition of *-in* (< PAn ***-en*** PV), as shown in (2)–(4). This appears to be problematic for the applicative analysis as increasing valency should be an applicative feature and not a voice feature.

(2) *lakad* *lakar-in*
'walk' 'to walk to x'

(3) *langoy* *languy-in*
'swim' 'swim in x, swim for x duration'

(4) *init* *init-in*[3]
heat 'to heat x'

In any case, we do not, as of yet, have any clear non-Austronesian functional analogues of the Austronesian morphemes in question. These morphemes, as emphasised by Blust (2002), appear to have features of voice and case, as well as bearing certain resemblances to applicatives. The issue is thus not terminological, but rather can only be resolved by a plausible syntactic scenario which can account for their mixed behaviour, a point to which we return later.

Because of the unusual status of the above paradigm, identifying the alignment system of Philippine-type languages has been at the center of several syntactic and typological debates, most of which have centered on whether Tagalog and other Philippine type languages are best analyzed as ergative or accusative.[4] Out of languages which morphologically distinguish the two arguments of a transitive clause, there are those which treat intransitive and transitive subjects alike in the accusative pattern, and those which mark intransitive subjects similar to transitive objects in the ergative pattern. This is shown schematically in (5), where A represents the transitive proto-Agent, P the transitive proto-Patient, and S, the sole argument of the intransitive clause. The indices 1 and 2 represent morphological case marking.

(5) **A-P differentiation:** $\{A\}_1\ \{P\}_2$
Accusative: $\{A, S\}_1\ \{P\}_2$
Ergative: $\{A\}_1\ \{S, P\}_2$

The syntactic and interpretive limitations on actor voice objects in Philippine languages suggest that the patient voice is a canonical transitive and therefore that transitive patients are case marked similarly to intransitive subjects in accordance with the ergative pattern

---

[3] For many Tagalog property denting roots, a causative affix is required to obtain the meaning 'to make *x* PROPERTY' but in several cases (e.g. *initin*) this morphology is not obligatory.

[4] De Guzman (1988), Gerdts (1988), Aldridge (2004 et seq), Liao (2004), Reid and Liao (2004) among others, argue that the actor voice is a marked choice for expressing transitives and is identifiable with an anti-passive. In a different camp, Kroeger (1993), Richards (2000), Rackowski (2002) argue for a more accusative type syntax. Himmelmann (1996) and Foley (1998) questions the entre utility of 'ergative type language' as a meaningful typological category for Austronesian languages. Because of the primarily typological focus of the present paper, particular syntactic proposals will not be discussed in any detail.

(De Guzman 1988; Gerdts 1988; Aldridge 2004; Liao 2004 among others). But in order to situate the position of these languages more meaningfully within an alignment typology, it is necessary to further articulate our conception of the ergative pattern. Morphological ergativity is defined minimally as a case syncretism between the intransitive subject and the transitive patient, but in the vast majority of ergative languages, there exist other syncretisms among the core and peripheral cases which are not taken into account. These further syncretisms provide important clues as to the historical origins of the pattern with syncretisms between the transitive agent and other cases being particularly revealing (cf. Palancar 2002). In a large number of ergative languages, the ergative argument is marked as an instrumental or ablative. In another group, the ergative argument is marked in the same manner as possessors, i.e., with the genitive case. This is shown schematically in (6), where we identify genitive and instrumental types as subtypes of the ergative alignment.

(6) **Ergative:** $\{A\}_1\ \{S, P\}_2$
Instrumental: $\{A, \text{Instrumental}\}_1\ \{S, P\}_2$
Genitive: $\{A, \text{Possessor}\}_1\ \{S, P\}_2$

As discussed by Plank (1979), Garrett (1990) and Dixon (1994), the Instrumental subtype is typically the outcome of a historical reanalysis in which passives or middles are reinterpreted as canonical transitives. Because adjunct agents are generally introduced by the instrumental (or directional cases) a homophony comes into being between the instrumental and the case of transitive agents after reanalysis. The genitive type, on the other hand, comes about from the reanalysis of nominalisations as canonical predicates. The agent of the event predicate is thus expressed as the possessor of the nominalisation.[5]

Reanalysis of nominalisations is precisely the type of event posited by SPR for PAn, which they describe as, 'a strongly noun-oriented language, with a high percentage of nominalisation strategies' (SPR:149). Similar scenarios have also been posited for a number of other language families on the basis of the genitive case marking pattern and independent supporting evidence, for example, Gildea (1998) for Cariban, Johns (1992) for Eskimoan, Bricker (1981) for Mayan. Nominalisation is a broad category which can refer to a number of related constructions and Austronesian appears to differ from some of the other language families with an ostensibly similar history in the type of nominalisations which were relevant in the reanalysis. In Austronesian, the nominalisations must have been of the thematic type (e.g., employer, employee) and not of the event-type (e.g., employment, destruction,

[5] Obviously, a historical reanalysis cannot be adduced for all cases of syncretism. The ergative-instrumental syncretism is very widespread in Australian languages, for instance, but no evidence of a historical reanalysis have been found. Either the earlier transitive patterns have been replaced without a trace in such languages or the syncretism did not arise through reanalysis at all but is rather a direct reflex of the semantic similarities between agent and instrument. The ergative-genitive syncretism is more likely to have a purely diachronic source as the semantic similarity between possessors and agents is more obscure. Baerman et al. (2005:52) express a similar opinion:

> With the ergative, type 2 syncretism in our sample most typically joins it with the genitive, as in the Tacanan language Araona, the isolate Burushaski, Lak and the Tibeto-Burman language Limbu. It is likely that this is not a random choice, in that there are languages which have cases which inherently combine the functions of ergative and genitive (e.g. the relative case of the Eskimoan languages). Such constructions may have their origin in nominalizations, with the agent expressed by the genitive. However, although diachronic explanations may be found, it is unlikely that a direct, synchronic motivation can be demonstrated for most type 2 pattern.

cooking).[6] Accordingly, SPR analyze the voice paradigm shown above in (1) as derivational nominalisation morphology, as in (7).[7]

(7)

PAn Form	Function
*<um>	Agent nominalisation
*-en	Patient nominalisation
*-an	Locative nominalisation
*Si-	Instrumental nominalisation

Beyond the historical reanalysis, SPR (p.148) further claim that these nominalisers, 'have in fact retained this function to a previously unrecognised extent even within the Philippine group'. Relatedly, the synchronic consequences of historical change are discussed by Manning (1996:21), who suggests that divergent origins of the ergative pattern can lead to differing varieties of synchronic ergativity:

> I believe that historical origin could be a good guide in subdividing the types of ergative languages, although the matter would require much further investigation. Making an initial cut between ergativity arising from a perfective or passive origin (reinterpreting an oblique instrumental or agent as the ergative NP) seems promising. (...) I am suggesting that many languages where ergativity arises from nominalisation are syntactically ergative (whereas the ergativity in the Indic Indo-European languages, for example, seems superficial from the point of view of syntactic behavior).

It is this connection between the putative historical reanalysis of nominalisations and the synchronic syntactic typology of Philippine languages which is the primary point of interest here and it is this topic which we begin to tackle next.

3 Austronesian languages as nominal predicate languages

3.1 The status of aspect morphology

If the genitive-ergative syncretism really has deep roots in nominalisation, we expect to find that event-type predicates, i.e., apparent verbs, display nominal characteristics. This may, at first blush, be a surprising claim considering that one of the most typical hallmarks of verbs cross-linguistically is tense/aspect marking and this is an integral part of event predicates in conservative Austronesian languages. Just as in more familiar languages, event predicates, but not arguments, are the canonical bearers of aspect morphology. As seen in (8), the event predicate *nag-bitaw* 'resign' is marked with perfective aspect (compositionally via the BEGUN affix and the lack of INCOMPLETIVE reduplication) while the subject *pangulo* 'president' lacks aspect morphology.

(8) *nag-bitaw* *ang=pangulo*
AV.BEG-resign NOM=president
'The president resigned.'

---

[6] These types are referred to as action nominalisations and argument nominalisations, respectively, in the terminology of Comrie and Thompson (1985:347).

[7] The forms have been updated to reflect the now commonly accepted reconstructions. The PAn infix *<*in*>, which SPR analyzed as a voice marker is now generally accepted to have belonged to the aspectual paradigm, indicating the perfective or realis. The instrumental, which was reconstructed by Wolff (1973) as PAn **i-* and by SPR as **iSi-* was later revised to **Si-* by Dahl (1986) and is now the generally agreed upon form.

On the face of it, then, there is nothing radically different in the canonical distribution of tense/aspect marking in such languages when compared to English. Nevertheless, I will argue that in conservative Austronesian languages, event predication, as expressed informally in (9a), is syntactically more analogous to (9b) and (9c) than it is to (9d).[8]

(9) a. [Fred employs the students]
b. The students are Fred's employees (patient nominal predicate)
c. Fred is an employer of the students (agent nominal predicate)
d. Fred employs the students (verbal predication)

The fact that these putatively nominal event-denoting predicates are marked with aspect only shows that aspect morphology is promiscuous in its selection of lexical hosts. Aspect appears to have been marked in Proto Austronesian by the use of two morphemes and the combination thereof: **Ca-/CV-* reduplication (PROGRESSIVE or INCOMPLETIVE), and the infix <in> (PERFECTIVE or BEGUN, cf. Reid 1992; Ross 2002). That these morphemes also attach to unambiguous lexical nominals in many languages is clear. Both are found abundantly on lexicalised, entity-denoting forms throughout Austronesian. The PAn **<in>* affix marked aspect on event-denoting predicates but its reflexes are also very commonly found on lexicalised referent-denoting words, so much so that many have interpreted this affix as a nominaliser in its own right. Reid (1992:68), for instance, echoing an idea proposed earlier in SPR, states that **<in>* was used on 'derived nouns that were the result of the action of the verb'. Some examples of lexicalised formations with *<in>* in Tagalog are shown in (10). Although (10a-b) can be considered lexicalised, they all have quite transparent event-denoting counterparts (e.g., *harap-in* front-PV 'to face'). The words in (10c-d), on the other hand, have no event-denoting counterparts in the modern language.

(10) a. *k<in>a~kapatid-Ø*
<BEG>INCM~sibling-PV
'relation between the sponsor and sponsored in a baptism, marriage, etc.'
b. *h<in>a~harap-Ø*
<BEG>INCM~front-PV
'future'
c. *b<in>abae-Ø*
<BEG>woman-PV
'hermaphrodite'

[8] This idea, too, has a long pedigree in Austronesian studies. Among twentieth century authors, we find it expressed by Bloomfield (1917), Scheerer (1924), Lopez (1937/1977), Capell (1964), Schachter and Otanes (1972), Lemaréchal (1991), Naylor (1995), De Wolf (1988) and Himmelmann (1991), among others. Note, however, that Ross (this volume) offers evidence from Puyuma showing that nominal and verbal predication may not have been symmetrical in PAn. As opposed to verbal predicates, Puyuma requires that nominal predicates are preceded by a determiner.

It is not clear to me at this point whether there exists a significant connection between the reanalysis of nominalisations as canonical predicates and equational type syntax. It seems that, while a symmetric treatment of verbal and nominal predicates may not be a necessary correlate of this reanalysis, it would certainly facilitate it by removing one of the more salient differences between nouns and verbs, the need for a copular element in non-verbal predication.

d. *s<in>ulid-Ø*
<BEG>flax-PV
'thread'

At least one referent-denoting etymon with this affix can also be traced to PAn: the word for intestines, PAn **C<in>aqi* 'intestines' from PAn **Caqi* 'feces'. The unpredictable relationship between these aspectual derivations with *<in>* and their stems in addition to the lack of a productive aspectual paradigm for many of them underscores the fact that they should be treated as lexicalised entity-denoting words, i.e., canonical nouns.[9] This strongly suggests that aspect marking was never the sole provenance of a particular lexical class in Austronesian. This having being established, aspect marking cannot be taken as evidence for equating event-denoting predicates with the lexical category verb.

3.2 The distribution of voice marked words

We now turn to the syntactic distribution of words marked with voice and aspect, showing that, in addition to playing a canonical 'verbal role', they also pattern with nouns cross-linguistically. The most obvious place in which this holds true is in the use of voice marked words as arguments, as exemplified in (11).[10, 11]

(11) a. *ang=b<um>ili*
NOM=<AV:BEG>buy
'the one who bought'

b. *ang=b<in>ili-Ø*
NOM=<BEG>buy-PV
'the (thing) bought'

c. *ang=b<in>il-han*
NOM=<BEG>buy-LV
'the (place) bought at'

d. *ang=i-b<in>ili*
NOM=CV-<BEG>-buy
'the one bought for'

Under analyses which impose a traditional verb/noun distinction on Tagalog and other Philippine languages these are treated as headless relatives. Although headless relatives are attested widely throughout the languages of the world, it is of interest that no morphosyntactically conservative Austronesian language requires constructions as in (11)

---

9 Discussion of aspectual reduplication is also relevant here but must be postponed until §5. For the moment, we may simply note that it is also well attested in both an apparent nominal and verbal function.

10 As in so much earlier work, I employ Tagalog to exemplify some typical characteristics of conservative MP languages. It should thus be kept in mind that the features under discussion here apply far more widely than Tagalog.

11 Abbreviations: ABS – absolutive, ACC – accusative, ADJ – adjectival, APPL – applicative, AV – actor voice, BEG – begun, CLASS – classifier, CV – conveyance voice, DEF – definite, DEP – dependent, DER – derivational morphology, ERG – ergative, EXT – existential, GEN – genitive case, INCM – incomplete, LNK – linker, LV – locative voice, NAV – non-actor voice, NEG – negation, NOM – nominative, NVL – non-volitional, OBL – oblique case, P – personal, PASS – passive, PL – plural, PREP – preposition, PRT – particle, PST – past, PV – patient voice, RELT – relative marker, RL – realis, STA – stative, TOP – topic marker, TR – transitive.

to be 'headed' by an unambiguous entity-denoting word (i.e., as by 'one', 'thing', 'place' in the English translations above).[12]

Relatedly, we typically find no indefinite pronouns in conservative languages. Instead, indefinite unspecified arguments (e.g., 'something', 'someone', 'somewhere', etc.) are expressed by the combination of the existential and the appropriate nominalisation. An indefinite agent must be expressed with the agent nominalisation as in (12); an indefinite patient with the patient nominalisation as in (13); an indefinite locative argument with the locative nominalisation as in (14) and so forth. The nominative case in the existential sentences below is assigned by the existential predicate itself and not the aspect marked predicate. The fact that existential predicates take aspect marked complements is predicted if these complements are in fact nominals. Note that this situation differs markedly from that found in mainland East Asian languages, where wh- words typically double as indefinite pronouns.

(12) *May bi~bili nang=uling*
EXT AV:INCM~buy GEN=charcoal
'Someone will buy charcoal.'

(13) *May bi~bil-hin si=Obet*
EXT INCM~buy-PV P.NOM=Obet
'Obet will buy something.'

(14) *May pu~punta-han si=Liwayway*
EXT INCM~go-LV P.NOM =Liwayway
'Liwayway has somewhere to go.'

Again related here is the requirement that content questions must be formed as cleft-like constructions in conservative languages. The notional predicate must be preceded by the nominative marker, as shown in (15a). Marking the interrogative phrase rather than the notional predicate with the nominative, as would be expected by a traditional extraction account, is ungrammatical, as shown in (15b). This is expected if all basic sentences, including interrogatives, are essentially copular clauses with PRED-SUBJ order. In content questions, then, it is the interrogative element which is in the predicate position and the aspect marked word which is in the subject position (cf. Keenan 1995; Gerassimova and Sells 2008).

---

[12] Note, however, that the same possibilities for such apparently headless nominal phrases are also afforded to prepositional oblique phrases, as seen in (i).

(i) *ang=para sa=bata*
NOM=for OBL=child
'the one that is for the child.' (Lemaréchal 1982:21 via Reid 2002:301)

As Reid (2002:301) notes, this is problematic for a theory which treats the complements of the case marking determiners as nominals, as it would require analyzing a prepositional phrase as a nominalisation. Reid (2002) treats the determiners themselves as head nouns and thus derives the productivity of headless relatives from the fact that the apparent case marker is a nominal head with the following complement as something akin to a relative clause. However, this cannot derive the nominal characteristics of verbs even when they are undetermined (e.g., the genitive-ergative syncretism and extraction facts to be discussed below). Nonetheless, this is an important point which unfortunately must be left open here.

(15) a. *Ano ang=b<in>ili-Ø=mo?*
wha NOM=<BEG>buy-PV=2S.GEN
'What did you buy?'

b. **Ang=ano b<in>ili-Ø=mo?*
NOM=what <BEG>buy-PV=2S.GEN

3.3 Coordination and constituency

Also predicted by the nominalist hypothesis is the fact that the predicate and the transitive agent form a constituent to the exclusion of the nominative argument. As possessor and possessum (genitive agent and predicate, respectively), they must constitute a larger nominal type phrase, represented schematically in (16).

(16) [[Pred Gen] Nom]

If coordination respects this constituency, we expect that the predicate plus genitive argument can be coordinated under a nominative argument as in (17a) but that the predicate and nominative constituent could not be coordinated under a genitive argument. This turns out to be correct, as shown by Kroeger (1993) who exemplifies with the coordinated constructions in (18) and (19), representing (17a) and (17b), respectively. In (18), the nominative phrase in final position is an argument of both coordinated predicates but in (19), the final genitive phrase can only serve as an argument to both preceding predicates with difficulty because it involves coordination of non-constituents.

(17) a. [[Pred Gen] and [Pred Gen] Nom]

b. *[[Pred Nom] *and* [Pred Nom] Gen]

(18) *hu~hugas-an=ko at pu~punas-an=mo ang=manga=pinggan*
INCM~wash-LV=1s.GEN and INCM~wipe-LV=2s.GEN NOM=PL=plate
'I'll wash and you dry the dishes.' (Kroeger 1993:34)

(19) *?*Ni-luto-Ø ang=pagkain at h<in>ugas-an ang=manga=pinggan*
BEG-cook-PV NOM=food and <BEG>wash-LV NOM=PL=plate
ni=Josie
P.GEN=Josie
(For, 'Josie will cook the food and wash the dishes.') (Kroeger 1993:34)

3.4 Case and extraction

The feature which has been lavished with the most attention in the syntax literature is the 'subjects only' restriction on extraction. Extraction (or apparent extraction) of arguments in question formation, relativisation and topicalisation have been described for many Austronesian languages as being restricted to the subject, i.e., the nominative argument. We can illustrate this with topicalisation, as topicalisation does not require altering the subject-predicate structure of the sentence. Taking a base sentence such as (20a), the nominative argument can be topicalised, as in (20b), but the genitive argument cannot, as shown in (20c).

(20) a. *B<in>ili-**Ø*** *nang=babae* *ang=libro* *kahapon*
<BEG>buy-PV GEN=woman NOM=book yesterday
'The woman bought a book yesterday.'

b. *Ang=libro ay b<in>ili-**Ø*** *nang=babae* *kahapon*
NOM=book TOP <BEG>buy-PV GEN=woman yesterday
'The book, the woman bought yesterday.'

c. **Nang=babae* *ay* *b<in>ili-**Ø*** *ang=libro* *kahapon*
GEN=woman TOP <BEG>buy-PV NOM=book yesterday

The generalisation 'subjects only', however, is a misnomer, as many other types of phrases can be extracted. For instance, the dative/oblique argument in (21), the bare temporal adverb in (22), the genitive marked temporal adjunct in (23), and the genitive marked clausal adjunct in (24).

(21) *Sa=paaralan* *ay* *nag-abuloy=sila* *nang=kotse*
OBL=school TOP AV.BEG-donate=3p.NOM GEN=car
'To the school, they donated a car.'

(22) *Kahapon* *ay* *b<in>ili-**Ø*** *nang=babae* *ang=libro*
yesterday TOP <BEG>buy-PV GEN=woman NOM=book
'Yesterday, the woman bought the book.'

(23) *Nang=ala-una* *ay* *<um>alis=sila*
GEN=o'clock-one TOP <AV.BEG>leave=3p.NOM
'At one o'clock, they left.'

(24) *Nang=hindi=niya* *na-malay-an* *ay* *na-hulog-**Ø**=siya*
GEN=NEG=3s.GEN NVL.BEG-conscious-LV TOP NVL.BEG-fall-PV=3s.NOM
'Without noticing it, he fell.'

Note that the case marking on a phrase does not help much in determining its potential for extraction. Although several types of adverbs and adjuncts are introduced with genitive case in Philippine languages, not all of them are unextractable. Instead, the restriction is properly described as applying to direct dependents of the predicate, that is, agents of non-actor voice predicates and objects of actor voice predicates. These types of arguments, in addition to certain 'inner adverbials' form a larger constituent with the predicate which cannot be extracted from. Note that these phrases are also dependent in the sense that they cannot stand alone and are unable to function as predicates in typical conservative languages, as exemplified by (25).[13]

[13] Typically, possessor predicates are expressed in the oblique case. Some languages, like Amis, are exceptional in allowing genitive phrases to fill the predicate position.

Extraction out of the predicate phrase should not be confused with genitive phrases preposed within the predicate phrase, as found in many languages. The difference is typically visible in that preposed genitives, unlike postposed genitives, are connected to the following material with the linker. This can be seen with the first person agent in the Pazeh sentence in (i) (compare the genitive agent of *kinan* without the linker). The alternation between a postposed clitic possessor and a preposed linked possessor is shown for Timugun Murut in (ii) and (iii). See also (39) and (40) below for similar examples.

(25) **Hindi nang=mundo ang=guro=ng iyon*
NEG GEN=world NOM=teacher=LNK that
(For, 'That teacher is not of the world.')

This restriction is typically presented as an exotic feature of Austronesian languages, especially in light of independent findings suggesting that subjects are more difficult to extract than objects cross-linguistically. However, given the nominalist hypothesis, the restriction is anything but exotic as the extraction of genitive arguments is equivalent to extraction from NP, a decidedly marked operation cross-linguistically.[14] To exemplify with non-Austronesian languages, we can observe the ungrammaticality of possessor extraction in Semitic and English. In (26), from Modern Hebrew, and (27), from Levantine Arabic, we see that in order to question a possessor, the entire NP within which it is contained must be fronted, as in the (a) sentences. Fronting of just the possessor while stranding the rest of the NP as in the (b) sentences is ungrammatical.

(26) a. *[et=ha=bayt šel mi]*$_i$ *raita* t$_i$ *?*
ACC=DEF=house of who see.PST.2s
'Whose house did you see?'

b. **[šel mi raita]*$_i$ *[et=ha=bayt* t$_i$ *]?*
Of who saw.PST.2s ACC=DEF=house

(27) a. *[be:t mi:n]*$_i$ *šuft* t$_i$*?*
house who see.PST.2S
'Whose house did you see?'

b. **mi:n*$_i$ *šuft [be:t* t$_i$*] ?*
who see.PST.2S house

This can be compared with the similar English facts in (28). In (28a), only a possessor is questioned but the entire containing NP must be fronted. The ungrammaticality of extracting just a possessor from this type of NP is shown in (28b) and (28c).[15]

(28) a. [Whose house] did you see t_i?
b. *[Whose]$_i$ did you see [t_i house]?
c. *[Of whom]$_i$ did you see a house t_i

(i) *naki a t<in>alek-Ø a alaw ka k<in>an-Ø ni=Asilu*
1S.GEN LNK <PRF>cook-PV LNK fish TOP <PRF>eat-PV P.GEN=Asilu
'the fish that I cooked was eaten by Asilu.' (Blust 1999a:351)

(ii) *baloy=min*
house=2S.GEN
'your house' (Prentice 1971:181)

(iii) *akaw ra baloy*
2s LNK house
'YOUR house' (Prentice 1971:205)

[14] As far as I am aware, Naylor (1980:42) is the only one to have made the connection between the unextractability of the *nang* phrase in Tagalog and its modifier/attribute status, although the basis is not made entirely clear: 'Obviously, structures that are bound retrogressively to the preceding constituent cannot precede the constituent. Thus, *nang*-NPs never occur initially in the clause (nor in the phrase).'

[15] The conditions on extraction from NP in English are notoriously difficult. Unlike the more categorial case of Semitic, the specificity of the containing NP and the nature of the predicate y play a large role (Erteschik-Shir 1973; Horn 1974).

The restriction on the extraction of non-actor voice agents and actor voice objects can now be reduced to whatever it is which blocks extraction of possessors from NP cross-linguistically (see Kaufman 2008 for some ideas on a formal implementation of this).[16] This general approach has the advantage of putting the famous Austronesian restriction on extraction on more universal grounds rather than treating it as an exotic case of a 'subjects-only' constraint on syntactic movement.

3.5 The syntax of independent and dependent imperatives

The final feature to be mentioned in this section relates to the syntax of imperatives. The imperative, as a speech act category, belongs entirely to the verbal realm. Accordingly, illocutionary/speech act categories are observed to be the last types of functional elements to be included in clausal nominalisation (Malchukov 2004). One of the defining syntactic features of imperatives cross-linguistically is the omission of the imperative addressee. Interestingly, imperative addressees are rarely omitted with the 'voice' derivations in (7). It is in fact, ungrammatical to omit the addressee in many contexts, as in the simple Tagalog imperative in (29).[17]

(29) *Sige, kain-in(*=mo)!*
alright eat-PV=2S.GEN
'Alright, eat (it)!'

However, if these forms are nominalisations, we may expect that predicates such as those in (29), which are often treated as imperatives, are not imperatives at all. In fact, the forms used for imperatives are not dedicated for this purpose but rather have a more general non-aspectual function also found in infinitive contexts. This helps account for the fact that these apparent imperatives cannot license omission of the addressee as imperatives in other languages typically do.[18] The most convincing evidence for this however comes from the behaviour of the forms in the so-called dependent paradigm, which we turn to next.

---

[16] This incidentally resolves a difficulty noted for Keenan and Comrie's (1977) Accessibility Hierarchy, shown in (i), in which relations further to the left were posited to be easier to extract.

(i) Accessibility Hierarchy (Keenan and Comrie 1977)
SU > DO > IO > OBL > GEN > OCOMP

Aldridge (2005) points out that the positioning of subjects as more extractable than objects is problematically based on data from Austronesian languages in which Keenan and Comrie equated transitive patients with subjects. Accepting the nominalist hypothesis, we can still maintain the cross-linguistic generalisation that objects may be easier to extract than subjects cross-linguistically (Rizzi 1990 and the references therein). What appears to be a preference for extracting SU in (i) is in fact a ban on the extraction of GEN. Asymmetries between SU and DO are thus essentially irrelevant here.

[17] Were the person marking in question to be understood as subject agreement, this would not be so surprising as imperatives can be marked for agreement with a null second person subject. But in light of the fact that the person markers in question are second position clitics, such an analysis is ruled out under the assumption that subject agreement must be marked on a verbal category (V or AUX) and can cooccur with full NP arguments. Neither of these criteria hold true in Tagalog nor in the majority of conservative Austronesian languages.

[18] Note that the indication of aspect in nominalisations, while not very widespread, is attested (see for instance Noonan 1992:213 for Lango).

Alongside the voice forms discussed above, Wolff (1973) also reconstructs another set which he terms the dependent forms, shown in (30).[19] As Wolff shows, the independent voice paradigm appears to have been used in matrix declarative clauses while the dependent paradigm was used in special contexts such as imperatives and negated clauses, a situation which is still maintained in many Austronesian languages today. SPR argue that the reanalyzed nominalisations represented by the independent forms supplanted the dependent forms, which were the original verbs of PAn.

(30)

	Independent	Dependent
Agent	*<*um*>	Ø
Patient	**-en*	**-a*
Locative	**-an*	**-i*
Conveyance	**Si-*	**-an*

Ross (2002a:46) makes the important observation that the interchangeability between notional nouns and verbs in Austronesian only applies to the independent forms. The similarity in distribution breaks down when we examine the dependent forms, which are restricted in appearing only in predicate and not in argument position. Although Ross (1995a:758 fn.24) correctly states that this difference is not visible in standard Tagalog, which has lost the dependent forms, we can see the distinction clearly in certain provincial dialects of Tagalog, such as that of Batangas, which preserve the dependent forms in imperatives.[20] In (31a) and (b) we see the imperative use of the independent and dependent forms in predicate position in Batangas Tagalog. In (32), we see the same imperatives in argument position with the patient in predicate position. But here, only the independent form is grammatical.

(31) a. *Buks-an=mo* *ang=pintuan!*
open=LV=2S.GEN NOM=door
'Open the door!'

b. *Buks-i* *ang=pintuan!*
open=LV.DEP NOM=door
'Open the door!'

---

[19] Ross (2002) reconstructs the dependent form (his 'non-indicative') of the conveyance voice as alternatively *án-i* V or V+*án-i*. He also reconstructs projective forms which include a suffix *-a* preceding the dependent suffixes in (30). Because these differences are not directly relevant here they will not be discussed further.

[20] Note that, as with most generalisations, exceptions can be found. The Ilonggo sentence in (i) shows the use of the locative dependent form embedded under a nominative case marker. This is highly unusual, however, if at all really permissible. Out of some 490 hits on the Google search engine of 'hatagi' (give-LV.DEP in several Bisayan languages), only two attestations of 'ang hatagi' could be found. Compare this to the 70 hits of the embedded independent form 'ang hatagan' (NOM=give-LV) out of 464 hits for 'hatagan' (give-LV) more generally.

(i) *Ako* *ang=hatag-i* *sang=pabo=mo!*
1S.NOM NOM=give=LV.DEP GEN=turkey=2S.GEN
'Just give your turkey to ME!'
<www.sabong.com.ph/forum/ showthread.php?t=10889&page=208?>

(32) a. *Pintuan ang=buks-an=mo!*
door NOM=open-LV=2S.GEN
'Open the WINDOW! Not the door' ('Window is your one to open!)

b. **Pintuan ang=buks-i!*
door NOM=open-DEP.LV

This shows that the noun-like nature of the aspect marked lexemes in Tagalog and other Philippine languages is not necessarily an across the board phenomenon which is basic to the syntax (*pace* Gil 1993). Rather, the syntax is capable of distinguishing nouns and verbs but nominals have simply subsumed verbs for the expression of most event-type predicates. The verbal nature of the dependent forms is further supported by their behaviour in imperatives. Although dependent forms license agents in negation and other auxiliary contexts, they require omission of the imperative addressee in all languages for which they are attested, as exemplified again by Batangas Tagalog in (33).[21] The contrast between the obligatory omission of the addressee of dependent form imperatives and their near obligatory inclusion in independent form imperatives underscores the verb like nature of the former set and the non-verbal nature of the latter.

(33) *Buks-i(*=mo)* *ang=pintuan!*
open=LV.DEP=2S.GEN NOM=door
'Open the door!'

The rest of this paper will focus on the disintegration of nominalism in Indonesian languages and the consequent reemergence of a robust verb-noun distinction in the morphosyntax of these languages. It is argued that many of the diverse changes which characterise the MP languages outside of the Philippines may be traced to the redeployment of true verbs as event-type predicates.

4 The disintegration of nominalism in Indonesia

The breakdown of the complex voice/nominalisation system in Indonesian languages has long been a topic of some interest among Austronesianists (see the papers in Wouk and Ross 2002 among others).[22] As emphasised in the recent literature (Blust 1985; Pawley and Ross 1993; Ross 1995b; Blust 1999b), there is no good comparative evidence suggesting that the MP languages outside of Central-East-Malayo-Polynesian form a single subgroup. Furthermore, there appears to be no large scale subgroup that includes a majority of extra-Philippine MP languages while excluding the Philippine languages. It is somewhat unexpected then that the MP languages outside of the Philippines show recurring morphosyntactic characteristics which set them apart from the more conservative Austronesian languages. Blust (2002:68) remarks on the rather uniform attrition of the voice system outside the Philippine area: '… [W]ith a few notable exceptions, languages closer to the probable Austronesian homeland in Taiwan have preserved more of the original focus system than languages at a greater distance from it. It is an intriguing

[21] Some languages only disallow inclusion of the imperative adressee when the addressee is singular, e.g., **Buksi=mo!* 'Open it!' but *Buks-i=ninyo!* '(You pl.) open it!'.

[22] It has only been relatively recently that we can identify with certainty the features of Indonesian voice systems as historical simplifications and not retentions. This is due to the distribution of the more complex systems across Formosan languages, which comprise several primary subgroups of Austronesian and thus demand reconstruction to PAn.

question why this should be the case.' If there was no exclusively shared common ancestor for the relevant languages, then there are only two alternatives left, as identified by Ross (2002:52): '… [T]heir similarities are at least in part the results of independent parallel developments and of language contact.' Here, I suggest that the cues for the nominal organisation of the syntax were lost by natural erosion and that this led to the re-emergence of a true verbal category as canonical event-denoting predicates. This is best seen as a consequence of the universal tendency towards treating event-denoting predicates as a separate lexical category with a privileged link to Tense-Aspect-Mood marking.[23]

Two of the most salient morphological cues to the nominal-based predication of conservative Austronesian languages are the nasal linker and the distribution of genitive case on both possessors and the agents. We review below how both of these cues eroded in Indonesian languages thereby setting the stage for the re-emergence of canonical verbs.

4.1 The loss of the linker

The linker connects all elements within a domain of modification. This can be seen by the position of the linkers in the Tagalog determiner phrase in (34).

(34) *Ito=ng* *dalawa=ng* *ma-laki=ng* *aso=ng* *ito*
NOM.this=LNK two=LNK ADJ-big=LNK dog=LNK NOM.this
'These two big dogs'

The linker is a ubiquitous feature of Philippine languages and is also present in Formosan languages.[24] Observe its obligatory presence in a simple adjective-noun constituent in Itbayat, Ilokano, Ibanag, Kapampangan, Tagalog, Waray (from Yamada and Tsuchida 1975:1) Maranao, Western Bukidnon and Amis as seen in (35a-i), respectively.

(35) a. *mahilid a* *raraxan*
narrow LNK road

b. *akikid na* *dalan*
narrow LNK road

c. *atazzi' nga* *dalan*
narrow LNK road

d. *makitid a* *dalan*
narrow LNK road

e. *makitid na* *daan*
narrow LNK road

f. *haligot nga* *dalan*
narrow LNK road

g. *maroni a* *wata'*
small LNK child

h. *madagway ha* *bulat*
beautiful LNK flower
(Post and Gardner 1992:64)

i. *miming-ay* *a siri*
small-FAC LNK goat
(Wu 2006:72)

The linker is nearly invisible in Indonesian languages, as can be gleaned from the sample in (36a-f), from Manuk Mangkaw Sama, Belait, Karo Batak, Makassarese, Mori Bawah and Kambera, respectively (Adelaar and Himmelmann 2005).

---

[23] Ultimately, I believe, the trigger for morphological erosion in the Indonesian languages was contact, as already suggested by Ross. Note that it was not only the linker, but a substantial part of the derivational morphology which was jettisoned in the Austronesian migration out of the Philippines. The complex derivational morphologies found in certain Indonesian languages (e.g., Tukang Besi, [Donohue 1999], Kambera [Klamer 2005], Acehnese [Durie 1985]) have been in large part reinnovated, although this deserves more in-depth research.

[24] The function and reconstruction of the linker is discussed by Blust (1974), Dempwolff (1934-38) and Ross (2006).

(36) a. *lansa heya*
boat large
'large boat'
(Akamine 2005)

b. *berejin ma'ang*
durian red
'red durian'
(Clyne 2005:436)

c. *telu wari*
three day
'three days'
(Woolams 2005:549)

d. *tau rua=yya*
person two=DET
'those two people'
(Jukes 2005:669)

e. *ambau opaa*
carabao four
'four carabao'
(Mead 2005:687)

f. *tau wàu*
personto.smell
'smelly person'
(Klamer 2005:714)

The correspondence between reduction in the voice system and the absence of linkers is not perfect. Philippine languages like Sarangani Manobo and Central Subanen have lost the linker, as seen in (37a-b), but preserved the voice system. Conversely, Indonesian languages like Toba Batak have made certain simplifications to the voice system but retain the linker, as seen in (38):

(37) a. *mepiya otaw*
good person
(Dubois 1976:97)

b. *lima basu*
five glass

(38) *dalan na soppit*
road LNK narrow
(Blust 1974:10)

Nonetheless, I believe the connection between loss of the linker and simplification of the voice system is highly significant. It is very plausible that the basis of this connection is that the domain of linking also includes genitive marked dependents of the predicate. The positioning of genitive phrases in PAn was probably much like it is today in modern Tagalog and Paiwan, as seen in (39) and (40). When postposed as in the (a) sentences, the genitive phrase was introduced by a simple case marker (or case marked clitic). However, when preposed, it was followed by the linker.

(39) a. *bahay=niya*
house=2s.GEN
'His/her house'

b. *kanya=ng bahay*
1s.GEN=LNK house
'His/her house'

(40) a. *umaq ni=maju*
house GEN=3s
'his/her house'

b. *ni=maju a umaq*
GEN=3s LNK house
'his/her house'

(Egli 1990:155; Himmelmann 2005:164)

In languages which retain the linker, good evidence that genitive agents are a type of possessor within the modificational domain is always available: preposed genitive agents are overtly linked to their predicates just like adjectives and other types of modifiers, as shown in (41).

(41) a. *P<in>unta-han=niya ang=bahay*
<BEG>go-LV=3s.GEN NOM=house
'He went to the house.'

b. *Kanya=ng p<in>unta-han ang=bahay*
3S.GEN=LNK <BEG>go-LV NOM=house
'He went to the house.'

The loss of the linker leads to the creation of a real (i.e. category particular) relative marker in many Indonesian languages. This can be illustrated by the comparison between Tagalog and Indonesian. In (42a) and (43a), we see that determiner phrase internal modification requires the linker in Tagalog but not in Indonesian. As seen by (42b), Tagalog treats modification of a noun by an event predicate just as it treats determiner phrase internal modification, that is, with the intermediation of the linker. On the other hand, the example in (43b), shows that Indonesian displays an asymmetry between determiner phrase modification and relativisation, requiring the marker *yang* for the latter.

(42) a. *ang=malaki=ng aso=ng iyon*
NOM=big=LNK dog=LNK that
'that big dog'

b. *ang=na-kita-Ø=ko=ng aso*
NOM=NVL.BEG-see-PV=1S.GEN=LNK dog
'the dog I see'

(43) a. *anjing besar itu*
dog big that
'that big dog'

b. *anjing *(yang) ku=lihat*
dog RELT 1s=see
'the dog that I see'

The loss of the linker thus leads to a categorial difference in the treatment of modification by canonical determiner phrase internal elements (e.g., adjectives, numerals, determiners) and modification by verbs. This obviously underscores the status of the verb as a *syntactically* distinct category in Indonesian languages and shows how the natural morphological erosion of the linker in Indonesian languages could have led to a significant reorganisation of the grammar.[25]

4.2 The loss of ergative-genitive syncretism

The other pervasive cue for nominalism mentioned above is the syncretism between non-actor voice agents and possessors. Alieva (1980) claims that the similarity is still significant for languages like Indonesian, offering the examples in (44) and (45) to demonstrate the similar marking of patient voice agents and possessors.

(44) *buku ini sudah di-baca anak-anak*
book this already PV-read child-child
'the children already read this book.'

(45) *ini buku-buku anak kami*
this book-book child 1P.EXCL
'these are our children's books.'

---

[25] Interestingly, we can see the development of the relative marker in Indonesian from the earliest attestations in Old Malay. Mahdi (2005:195) notes that the marker *yaṃ* in Old Malay (Classical Malay *yang*) was often omitted where it was obligatory in later Classical Malay. For instance, in introducing the relative clause in (i).

(i) *ni-vunuḥ kāmu sumpaḥ ni-minu[m]=māmu*
PV-kill 2p curse PV-drink=2P.GEN
'you will be killed by the curse which is drunk by you.'

Alieva (1980:421) further suggests that differences between possessors and agents in Indonesian are perhaps due to foreign influence. This is, however, incorrect, as several significant differences can already be seen in the earliest attestations of Malay. Firstly, agents, but not possessors, could already be introduced by the preposition *oleh* in early Classical Malay texts. Secondly, possessors were regularly doubled by genitive pronouns in early Classical Malay, as shown in (46), a situation which probably obtained even in Old Malay, as suggested by (47). A genitive pronoun following a patient voice verb in Classical Malay was never co-referent with a following (unmarked) agent. A following NP is always interpreted as the patient, as shown by the interpretation of (48).[26]

(46) *Apakah* ***dosa-nya*** ***anak-ku***, *maka engkau bunuh-kan dia?*
what sin-3s.GEN child-1s.GEN thus 2s kill-APPL 3s
'What was my child's sin that you killed him?'
(Hikayat Bayan Budiman 202:16, from the Malay Concordance Project http://mcp.anu.edu.au/)

(47) *sthāna-ña* *śatru-ṇku*
residence-3S.GEN enemy-1S.GEN
'the position/residence of my enemy' (SKN 12, Mahdi 2005:194)

(48) *Ia men-cari damar, di-pasang-nya di-suluh-nya* ***di-lihat-nya***
3s AV-search resin PV-put-3S.GEN PV-torch-3S.GEN PV-see-3S.GEN
anak-nya
child-3S.GEN
'He searched for resin, he put it (in its place), he lit it, and **he saw his child**.'
(Hikayat Bayan Budiman 132:23, from the Malay Concordance Project http://mcp.anu.edu.au/)

Once the noun-verb distinction took on this new syntactic significance, the verbal category was free to develop along cross-linguistically familiar lines. In particular, it developed person agreement and a true passive, lacking in earlier stages. Person agreement on verbs, first documented systematically by Haaksma (1933), is a feature which sharply distinguishes Indonesian languages from Philippine and Formosan ones (Wolff 1996; Zobel 2002; Kikusawa 2003; Himmelmann 2005:149–151). Crucially, it cleaves apart the two functions of the inherited genitive pronouns as it treats non-actor voice agents differently from possessors.[27] We can see how this typically works from the sample of Sulawesi languages discussed by Noorduyn (1991:148–149) and shown in Table 1. In all the languages shown, the agent marker attaches to the verb as an agreement prefix while the possessor is a phrase final pronominal clitic. Note that, as may be gleaned from the

[26] Full NP agents could be doubled by a genitive pronoun but only when introduced by the agent marker *oleh*, and even then, typically only when dislocated, as in (i).

(i) *Maka oléh Bedawi itu pun di-beri-kan-nya air kepada Hasanah*
so by B. that even PV-give-APPL-3S.GEN water to H.
'So that Bedawi, he gave water to Hasanah.' (Hikayat Bayan 198:7)

[27] In what at first appears to be a parallel development, genitive proclitics were also innovated in Paiwan and Puyuma. A major difference exists, however, between the Formosan proclitics and their Indonesian counterparts. In both Paiwan and Puyuma these proclitics mark both non-actor voice agents and possessors whereas in Indonesian languages they only have an agent marking function.

largely unrelated basic lexemes, not all of these languages are closely related to each other and the development in question was quite clearly not inherited from a common ancestor, even on the single island of Sulawesi (Mead 2002).

Table 1: Ergative agents versus possessors in Sulawesi languages

Transitive clause	Noun phrase	Language
na-*hilo=a*	*tomi=**ku***	Uma
na-*cini=ka'*	*ballak=**ku***	Makassarese
na-*kita=na'*	*banua=**ngku***	Sad'an Toraja
la-*longa-aku*	*sapo=**ku***	Barang-barang
a-*kamata-aku*	*banua=**ku***	Wolio
no-*toa-aku*	*laika-**nggu***	Tolaki
3S.ERG-see=1S.ABS	house=1S.GEN	
'S/he sees me'	'my house'	

4.3 The development of a passive

The development of a true passive voice is another widespread innovation among Indonesian languages that is unknown in Philippine languages.[28] One of the correlates of the passive is the introduction of the agent as an oblique prepositional phrase rather than a genitive phrase. This can be seen in the diverse group of Indonesian languages in (49)–(53).[29]

(49) *Ni-kokko'=a'* ***ri*** *meong=ku* — Makassarese
PASS-bite=1s.NOM PREP cat=1s.GEN
'I was bitten by my cat' (Jukes 2006:254)

(50) *Mbe'e ede ra-nduku* ***ba*** *ompu sia* — Bima
goat that PASS-hit by grandfather 3s
'The goat was hit by his/her grandfather' (Arka 2002)

(51) *Tu' da-kerja* ***ulih*** *dua iku' nsia* — Mualang
this PASS-work by two CLASS human
'This is done (later) by two persons.' (Tjia 2007:152)

---

[28] Some authors treat stative forms in conservative Austronesian languages as passives due to the fact their undergoers are typically mapped to the nominative argument and their agents are freely omitted (see Reid and Liao 2004:462). In Central Philippine languages like Tagalog there is no evidence for this, as stative agents also display genitive marking and show similar syntactic behaviour to dynamic non-actor voice agents, the only interpretative difference being in the realm of volitionality and ability. In certain other Northern Philippine and Formosan languages, however, stative forms appear to resist licensing a genitive agent. If Ross (1995a:741) is correct in reconstructing the PAn stative **ma-* as derived by the combination of **ka-* STATIVE and **<um>* ACTOR VOICE (i.e. **k<um>a-* > **ma-* via loss of the first syllable) then statives may be best treated as a variety of actor voice predicate rather than a variety of passive, unlike Indonesian passives which have no connection at all to the actor voice.

[29] There is additional evidence for the independent but parallel nature of this development in the fact that several widespread languages appear to have made an early borrowing of the agent introducing preposition *oleh* from Malay in some form. Among others, a form *le* introduces passive agents in Manggarai, Acehnese and various Sama languages, a rather motley distribution which does not subgroup closely.

(52) *Ami ongga **le** hia* Manggarai
1p.EXCL hit by 3s
'We were hit by him/her.' (Arka and Kosmas 2005)

(53) *Lôn ka geu-côm **lé**-gopnyan* Acehnese
1p IN 3-kiss OBL-she
'I was kissed by her.' (Durie 1988)

We predict that when agents are introduced obliquely as in the above languages they should not be bound by the restrictions on possessor extraction discussed earlier, as they are no longer contained within a predicate noun phrase. This is confirmed by Mualang and Sundanese below (see also the Classical Malay example in fn.25) which both allow topicalisation of passive agents.

(54) *Ulih dua iku' nsia tu' da-kerja* Mualang
by two CLASS human this PASS-work
'This is done (later) by two persons.' (Tjia 2007:152)

(55) *Ku bapa=na bade di-pang-meser-keun motor* Sundanese
By father=3S.GEN will PASS-DER-buy-TR motorbike
'His father will buy him a motorbike.' (Müller-Gotama 2001:33)

4.4 The development of canonical applicatives

Another consequence of the development of canonical verbs in Indonesian languages is the licensing of true applicatives. Recall that the various nominal 'voices' in more conservative languages do not exactly serve to increase the valency of the predicate. In these languages, when a benefactee is selected by the conveyance morphology on the predicate, all other arguments are typically expressed in the genitive case. It is a much discussed fact that similar morphology in Indonesian languages has the ability to create new objects on actor voice verbs, as seen in (56)–(58).

(56) *Aku men-ulis-**kan** kamu sajak* Indonesian
1s AV-write-APPL 2 poem
'I wrote a poem for you'

(57) *Bib n=pun-**ak** kolay peda* Taba
Bib 3s=kill-APPL snake machete
'Bib killed the snake with a machete' (Bowden 2001:122)

(58) *Ia meli-**ang** Nyoman umah* Balinese
3 AV.buy-APPL name house
'(S)he bought a house for Nyoman' (Arka 2002)

Note that applicatives are a nearly uniquely verbal category, rarely attested in the nominal domain (Malchukov 2004). It is thus unsurprising that with the loss of the cues for nominal predicates and the consequent development of a true verbal category, we should also find the parallel development of applicatives, a morphological category which is virtually unknown in the more conservative languages.[30]

[30] Recall from above, however, that Aldridge (2004) and Ross (2006), among others, do argue for an interpretation of CV and LV morphology as applicatives.

4.5 The status of the imperative addressee

Finally, we can note that verbs are naturally suited for a wider variety of illocutionary acts than are nominals, whose proto-typical functions are referential in nature. We expect that the new verbal category of Indonesian languages should be similar to the old dependent forms examined earlier in requiring null imperative addressees. That this is generally correct can be seen from the representative examples in (59), from Malay and Selayarese, respectively. The imperative addressee is obligatorily omitted in Selayarese and typically omitted in Malay.

(59) a. *Masak sayur=nya!*
cook vegetable=3S.GEN
'Cook the vegetables!'

b. *Keo=a!*
call=1S.NOM
'Call me!'

We have seen in this section how several typological traits shared among Indonesian languages of various MP subgroups can be related to the redevelopment of canonical verbs. We also saw how this development could have been aided by morphological attrition, as this has removed some of the most salient morphological clues for the nominal oriented syntax so characteristic of the more conservative languages. In the next two sections we review three problems for the nominalist hypothesis and offer directions for further research.

5 Residue

5.1 **Ca-* reduplication and the noun-verb distinction

Blust (1998) brings to light an intriguing problem for SPR's nominalist hypothesis. He shows that PAn appears to have had an independent method for forming instrumental nominalisations, **Ca-* reduplication, which was unrelated to the voice system. Words with **Ca-* reduplication form unambiguous entity-denoting words and appear not to take aspect marking. Problematically, this appears to be nearly identical to the function of **Si-* according to SPR, except that the latter clearly formed event-denoting predicates, typically taking aspect morphology. If the voice forms were indeed nominalisations, it is difficult to explain why these two forms share the same basic semantics but differ syntactically in modern languages along apparent noun-verb lines.

A plausible reanalysis of the facts involves taking the reduplication in question not as a marker of instrumental formation per se, but rather as an instantiation of the very same morpheme which has been reconstructed for the durative aspect (Ross 1995a:750–751, Blust 1998:34–35). As is common cross-linguistically, the durative would have also been used to denote habitual action. In the case of instrumentals, then, the reduplication would indicate the habitual use of the root but not the instrumental semantics itself, which must consequently be considered as the product of a zero-derivation. This would explain why **Ca-* instrumental forms could not take additional aspectual morphology, as they would have already been marked for aspect. Doubtlessly, lexicalisation has occurred in a large number of these forms as their interpretations are not always transparently derivable from the roots (see Blust 1998 for extensive discussion), but the reanalysis suggests that there is no deep categorical difference between such instrumentals and their aspectually productive counterparts with **Si-*.

Although defending this proposal properly requires far more space than is available here, some pieces of supporting evidence can be briefly brought to bear on the problem. If

**Ca-* reduplication indicated habitual aspect we would expect to find it not only on instrumentals but also on agentive and locative nominalisations, among others types. As Blust notes, many forms which can be interpreted as instrumentals take the PAn locative suffix **-an*. Evidence from several languages further confirms this prediction. As discussed by Huang (2002), Mayrinax Atayal employs *Ca-* reduplication for both what she describes as non-actor irrealis forms and nominalisations.[31] In the actor voice, both of these functions happen to be subsumed by *pa-* instead of *Ca-* (see Huang 2001:56–57 for an analysis of this). Crucially, however, the same form functions as an aspectual marker and as an apparent nominalisation marker, as can be seen from (60) and (61), respectively.

(60) ***pa~Ø-paquwas*** *kuʔ=irawiŋ=mu*
IRR-AF-sing NOM=friend=1S.GEN
'My friend will sing' (Huang 2002:211)

(61) *βaq-un=mu* *kuʔ=**papaquwas*** *kaʔ=hacaʔ*
know-PF=1S.GEN NOM=singer LNK=that
'I know that singer (who will sing there)' (Huang 2002:211)

The same pattern is found not only with agents but with patients, locatives and instruments, as well. What are translated as canonical nominalisations consistently take *Ca-* reduplication in Mayrinax Atayal. Note that this is also the case with many lexicalised nominalisations in Philippine languages. In Tagalog, for instance, we find reduplication (of the *CV-* variety) regularly in forms such as those in (62), just as we find it in the aspectual paradigm.[32]

(62) a. *la~lamun-an*
INCM~swallow-LV
'throat'

b. *mang~ang-awit*
AV.DIST~INCM-sing
'singer'

Intriguingly, although Mayrinax Atayal possesses a reflex of PAn **Si-* in the form of *si-*, used for the instrumental/benefactive voice, it does not employ this prefix either for the irrealis verb or for instrumental nominalisations (Huang 2002:219–220). Thus, the *Ca-* reduplicant forms the instrumental *pa~patiq* 'pen' from the root *patiq* 'write' but is also used without overt indication of voice in the aspectually productive paradigm of the instrumental/benefactive voice, as can be seen from the difference between realis (63) and irrealis (64).

(63) *si-ʔaɣal=miʔ* *cuʔ=pilaʔ* *kiʔ=sayun* *kuʔ=naβakis*
BF-take=1S.GEN ACC.NRF=money OBL=sayun NOM.RF=old.man
'I took money from Sayun for the old man.' (Huang 2001:54)

(64) *ʔa~ʔaɣal-∅=miʔ* *cuʔ=pilaʔ* *kiʔ=sayun* *kuʔ=naβakis*
IRR-take-BF=1s.GEN ACC.NRF=money OBL=sayun NOM.RF=old.man
'I'll take money from Sayun for the old man.' (Huang 2001:54)

---

[31] Although the gloss irrealis is not congruent with a habitual interpretation, the function of aspectual reduplication has clearly undergone major changes in many languages. On the shifting aspectual semantics of **Ca-*/**CV-* reduplication see Reid (1992) and Ross (1995a:750–752, 2002).

[32] Lexicalised nominalisations are often further distinguished from aspectually productive ones in Tagalog by a process of length flip (more commonly referred to as 'stress shift' in the literature) by which vowel length is removed from roots with a long penult and added to those without one.

Note that in (64), Huang marks the benefactive voice as being zero-derived in the irrealis. If this irregularity is a retention from PAn, then it follows that what appear to be instrumental nominalisations with **Ca-* in various Austronesian languages are better described as lexicalised etymons marked with proto-durative aspect (to indicate the habitual) and a null morpheme indicating instrumental voice.

5.2 The case of dependent form agents

A perhaps more serious problem for the SPR analysis which appears to have gone generally undiscussed is the case frame of the dependent forms. Recall that one of the primary motivations for the nominalisation hypothesis was that genitive case was used both to mark possessors and non-actor voice agents. Nominalisation explains this pattern as arguments of nominalisations are typically expressed as possessors. However, if the dependent forms were the original verbs—having been ousted from matrix clauses by nominalisations—we should certainly not expect that they would also mark agents with the genitive, but this is in fact what we find throughout. Observe the marking of the agents in Samarenyo (65)–(66) and Atayal (67).

(65) *Wara'=**ku** balik-a a=sirbisa*
NEG=1S.GEN return-PV NOM=beer
'I didn't return for the beer.' (Wolff 1973:76)

(66) *Wara'=**ku** hingalimt-i a=isturya*
NEG=1S.GEN forget-LV.DEP NOM=story
'I didn't forget the story.' (Wolff 1973:78)

(67) *nanu' kina ini' gngi'-i **na' Asang** pi qu' pqziuan*
and perhaps NEG forget-LV.DEP GEN A. PRT NOM legend
mrhuuraral ga'
forefather PRT
'Maybe Asang has not forgotten the legends of our forefathers.' (Wolff 1973:78)

In both languages, as in every other language for which the dependent forms are attested, the non-actor voice forms require the genitive case on the agent. Thus, the spread of nominalisation meant to explain the distribution of the genitive case has to apply before nominalisation even enters the scene. I take this to be the most daunting challenge facing the nominalist hypothesis at present.[33]

5.3 The syntax of roots

Another puzzle concerning the nominalist hypothesis comes to light when we examine the syntax of bare roots in certain Philippine languages. Just as we unexpectedly find genitive agents with dependent forms where we do not expect them, we also commonly find genitive agents of bare roots in languages like Tagalog. Observe the sentences in (68) where no voice or aspect marking is found on the predicates. In both cases the agent is obligatorily assigned genitive case and the patient is assigned nominative case. It is not

[33] Ross (this volume) argues on the basis of data from Puyuma, Rukai and Tsouic that this state of affairs characterises what he terms Proto Nuclear-Austronesian but not PAn proper, which did not show this syncretism.

plausible to derive such bare root predicates by simply taking them to be reductions of their corresponding voice and aspect inflected counterparts (e.g. *d<in>a~dala-Ø* <BEG>INCM~carry-PV for (68a) and *na-ki~kita-Ø* NVL.BEG-INCM~see-PV for (68b)) because they do not have the same aspectual interpretation. Aspect inflected forms describe events while bare root forms describe states as indicated by the stative translations below. Thus, the sentence in (68a) is an infelicitous answer to the question 'What is he doing?' unlike its voice and aspect inflected counterpart. Note also that the patient oriented nature of bare roots in Tagalog is present even without the presence of a genitive agent, as shown by (69) where the nominative argument can only be interpreted as the patient of 'see' (see also Himmelmann 2008).

(68) a. *dala=niya* *ang=niyog*
carry=3S.GEN NOM=coconut
'The coconut is his carried thing.' (i.e. 'He carries the coconut.')

b. *kita=niya* *ang=bangka*
see=3S.GEN NOM=boat
'The boat is his visible thing.' (i.e. 'He sees the boat.')

(69) *kita=ka*
see=2S.NOM
'You're visible' (not, 'You see')

Again we are faced with the embarrassment of nominal characteristics without the presence of nominalizing morphology (i.e. the independent voice forms). If event-denoting roots in PAn were inherently nominal to begin with then the independent voice forms could not be understood as nominalisers. However, it is not at all clear that the nominal properties of roots seen in Tagalog can be reconstructed all the way to PAn although the work required to ascertain this one way or the other has yet to be carried out systematically.

6 Conclusion

In this paper, I have argued that several morphosyntactic features broadly associated with conservative Austronesian languages result from their nominal syntax and that the features characteristic of many MP subgroups outside of the Philippines result from the re-emergence of a canonical verb. The side of the story which can be told through ordinary genetic inheritance is summarised in Table 2. The function of the dependent paradigm was presumably purely predicational in PAn (or pre-PAn) but was marginalised in PMP to imperative, negative and narrative contexts. Presumably, the independent paradigm was primarily used for arguments in PAn (or pre-PAn) but came to be used for canonical event-denoting predicates in PMP.

Table 2: Ordinary genetic inheritance

	(Pre-)Pan functions	**PMP functions**
Dependent paradigm	canonical predicate	(imperative, negative, narrative)
Independent paradigm	canonical argument	canonical predicate

Subsequent developments appear to have been the result of convergence and parallel change among those groups which migrated further south. In these languages, the dependent and independent paradigms were merged to a large extent to create a robustly verbal category which often allowed applicatives and person agreement. This helps to explain certain puzzling typological similarities between disparate genetic subgroups of Indonesian languages. The erosion of the cues for nominality—perhaps due to simplification via heavier contact with speakers of non-Austronesian languages south of the Philippines—led to the redevelopment of verbs as a morphosyntactic category. The redevelopment of verbs gives way to a constellation of properties commonly found in Indonesian language, among which we find unique relative markers, verbal agreement and object creating applicatives. Further work on the PAn dependent paradigm and on the syntax and semantics of bare roots should ultimately help elucidate the validity of the nominalist hypothesis.

References

Adelaar, Alexander and Nikolaus P. Himmelmann, eds. 2005. *The Austronesian languages of Asia and Madagascar*. London: Routledge.

Adriani, Nicolaus. 1893. *Sangireesche spraakkunst*. Leiden: Nederlandsch Bijbelgenootschap.

Akamine, Jun. 2005. Sama (Bajau). In Adelaar and Himmelmann, eds 2005:377–392.

Aldridge, Edith. 2004. Ergativity and word order in Austronesian languages. PhD thesis, Cornell University.

—— 2005. Two fallacies relating to the accessibility hierarchy. Paper presented at the LSA conference, Oakland.

Alieva, Natalia. 1980. The possessive syntactical type and the Malay language. *Bijdragen tot de Taal-, Land- en Volkenkunde* 136.4:417–425.

Arka, I Wayan and Jeladu Kosmas. 2005. Passive without passive morphology? Evidence from Manggarai. In I Wayan Arka and Malcolm Ross, eds *The many voice of Austronesian systems: some new empirical studies*, 87–117. Canberra: Pacific Linguistics.

Arka, I Wayan. 2002. Voice and being core: evidence from (Eastern) Indonesian Languages. Paper presented at Austronesian Formal Linguistics Association 7.

Baermann, Matthew, Dunstan Brown and Greville G. Corbett. 2005. *The syntax-morphology interface: a study of syncretism*. Cambridge: Cambridge University Press.

Bloomfield, Leonard. 1917. *Tagalog texts with grammatical analysis*. 3 vols. Urbana, Ill.: University of Illinois.

Blust, Robert. 1974. Proto-Austronesian: the first step. *Oceanic Linguistics* 13.1/2:1–15.

—— 1985. The Austronesian homeland: a linguistic perspective. *Asian Perspectives* 26.1:45–67.

—— 1998. Ca- reduplication and Proto-Austronesian grammar. *Oceanic Linguistics* 37.1:29–64.

—— 1999a. Notes on Pazeh phonology and morphology. *Oceanic Linguistics* 38/2:321–365.

—— 1999b. Subgrouping, circularity and extinction: some issues in Austronesian comparative linguistics. In E. Zeitoun and P. Jen-kuei Li, eds *Selected Papers from the Eighth International Conference on Austronesian Linguistics*, 31-94. Taipei: Academia Sinica.

—— 2002. Notes on the history of 'focus' in Austronesian languages. In Wouk and Ross, eds 2002:63–78.

Bowden, John. 2001. *Taba: description of a South Halmaheran language*. Canberra: Pacific Linguistics.

Bricker, Victoria. 1981. The source of the ergative split in Yucatec Maya. *Journal of Mayan Linguistics* 2.2:83–127.

Capell, Arthur. 1964. Verbal systems in Philippine languages. *Philippine Journal of Science* 93:231–249.

Clynes, Adrian. 2005. Belait. In Adelaar and Himmelmann, eds 2005:429–455.

Comrie, Bernard and Sandra A. Thompson. 1985. Lexical nominalization. In Timothy Shopen, ed. *Language typology and syntactic description*, vol. III: *grammatical categories and the lexicon*, 349–398. Cambridge, New York: Cambridge University Press.

Dahl, Otto Christian. 1986. Focus in Malagasy and Proto-Austronesian. In P. Geraghty, L. Carrington and S.A. Wurm, eds *FOCAL II: Papers from the Fourth International Conference on Austronesian Linguistics*, 21–42. Canberra: Pacific Linguistics.

De Guzman, Videa. 1988. Ergative analysis for Philippine languages: an analysis. In R. McGinn, ed. *Studies in Austronesian linguistics*, 323–345. Athens, Ohio: Ohio University Monographs in International Studies.

De Wolf, Charles. 1988. Voice in Austronesian languages of Philippine type: passive, ergative, or neither? In Shibatani, ed. *Passive and voice*, 143–193. Amsterdam: Benjamin.

Dempwolff, Otto. 1934–38. *Vergleichende Lautlehre des austronesischen Wortschatzes* (3 vols), Beihefte Zeitschrift für Eingeborenensprachen 15, 17, 19. Berlin: Dietrich Reimer Verlag.

Dixon, Robert. 1994. *Ergativity*. Cambridge: Cambridge University Press.

Donohue, Mark. 1999. *A grammar of Tukang Besi*. Amsterdam: Mouton de Gruyter.

DuBois, Carl D. 1976. *Sarangani Manobo*. Manila: Summer Institute of Linguistics.

Durie, Mark. 1988. *A grammar of Acehnese on the basis of a dialect of North Aceh*. Dordrecht: Foris.

Egli, Hans. 1990. *Paiwangrammatik*. Wiesbaden: Harrassowitz

Erteschik-Shir, Nomi. 1973. On the nature of island constraints. PhD dissertation, Massachusetts Institute of Technology.

Foley, William A. 1998. Symmetrical voice systems and precategoriality in Philippine languages. Paper presented at the 3rd LFG conference, Brisbane.

Garrett, Andrew. 1990. The origin of split NP ergativity. *Language* 66:2.261–296.

Gerassimova, Veronica and Peter Sells. 2008. Long-distance dependencies in Tagalog: the case for raising. In Charles B. Chang and Hannah J. Haynie, eds *Proceedings of the 26th West Coast Conference on Formal Linguistics*, 190–198. Somerville, MA: Cascadilla Proceedings Project.

Gerdts, Donna. 1988. Antipassives and causatives in Ilokano: evidence for an ergative analysis. In Richard McGinn, ed. *Studies in Austronesian linguistics*. Athens, Ohio: Ohio University.

Gil, David. 1993. Tagalog semantics. *Proceedings of BLS 19*, 390–403.

Gildea, Spike. 1998. *On reconstructing grammar: comparative Cariban morphosyntax*. Oxford: Oxford University Press.

Haaksma, Rémy. 1933. *Inleiding tot de Studie der Vervoegde Vormen in de Indonesische Talen*. Leiden: Brill.

Himmelmann, Nikolaus. 1987. *Morphosyntax und morphologie – Die Ausrichtungsaffixe im Tagalog*. München: Fink.

—— 1991. The Philippine challenge to universal grammar, *Arbeitspapier 15*. Köln: Institut für Sprachwissenschaft.

—— 1996. Person marking and grammatical relations in Sulawesi. In H. Steinhauer, ed. *Papers in Austronesian linguistics No.3*. Canberra: Pacific Linguistics.

—— 2005. Typological characteristics. In Adelaar and Himmelmann, eds.

—— 2008. Lexical categories and voice in Tagalog. In Peter K. Austin and Simon Musgrave, eds *Voice and grammatical functions in Austronesian languages*. Stanford: CSLI.

Horn, G.M. 1974. The noun phrase constraint. PhD thesis, Massachusetts Institute of Technology.

Huang, Lillian, M. 2001. Focus system of Mayrinax Atayal: a syntactic, semantic and pragmatic perspective. *Journal of Taiwan Normal University: Humanities and Social Science* 46 (1–2):51–69.

—— 2002. Nominalization in Mayrinax Atayal. *Language and Linguistics* 3 (2):197–225.

Johns, Alana. 1992. Deriving ergativity. *Linguistic Inquiry*, 23.1:58–87.

Jukes, Anthony. 2006. Makassar. In Adelaar and Himmelmann, eds 2006:649–682.

Kaufman, Daniel. 2008. Austronesian voice as thematic nominalization. Paper presented at the 15th meeting of the Austronesian Formal Linguistics Association, Sydney. July 1, 2008.

Keenan, Edward and Comrie, Bernard. 1977. Noun phrase accessiblity and universal grammar. *Linguistic Inquiry* 8:63–69.

Keenan, Edward. 1995. Predicate-argument structure in Malagasy. In C.S. Burgess, ed. *Grammatical relations: theoretical approaches to Empirical questions*. Stanford: CSLI Publications.

Kikusawa, Ritsuko. 2003. The development of some Indonesian pronominal systems. In Barry J. Blake and Kate Burridge (eds, with Jo Taylor) *Historical linguistics 2001: selected papers from the 15th International Conference on Historical Linguistics, Melbourne, 13–17 August 2001*. Current Issues in Linguistics Theory, 237–268. Amsterdam: John Benjamins.

Klamer, Marian. 2005. Kambera. In Adelaar and Himmelmann, eds 2005:709–727.

Kroeger, Paul R. 1993. *Phrase structure and grammatical relations in Tagalog*. Stanford: Stanford University Press.

Lemaréchal, Alain. 1991. Dérivation et orientation dans les langues des Philippines, *Bulletin de la Société de Linguistique de Paris* 86:317–358.

Liao, Hsiu-chuan. 2004. Transitivity and ergativity in Formosan and Philippine languages. PhD thesis, University of Hawai‘i at Mānoa.

Lopez, Cecilio. 1928. Comparison of Tagalog and Iloko. PhD thesis, University of Hamburg.

Lopez, Cecilio. 1937/1977. Preliminary study of affixes in Tagalog. In Ernesto Constatino, ed. *Selected writings in Philippine linguistics*, 28–104. Diliman: University of the Philippines Press.

Mahdi, Waruno. 2005. Old Malay. In Adelaar and Himmelmann, eds, 2005:182–201.

Malchukov, Andrej. 2004. *Nominalization/Verbalization*. LINCOM Studies in Language Typology. Munich: Lincom Europa.

Manning, Christopher. 1996. *Ergativity: argument structure and grammatical relations*. Stanford: CSLI.

Mead, David. 2002. Proto-Celebic focus revisited. In Wouk and Ross, eds 2002:143–177.

—— 2005. Mori Bawah. In Adelaar and Himmelmann, eds 2005:683–708.

Müller-Gotama, Franz. 2001. *Sundanese*. Languages of the World/Materials 369. München: Lincolm

Naylor, Paz B. 1975. Topic, focus, and emphasis in the Tagalog verbal clause. *Oceanic Linguistics*. 14:1.

—— 1980. Linking, relation-marking, and Tagalog syntax. In Paz B. Naylor, ed. *Austronesian Studies, papers from the 2. Eastern Conference on Austronesian Languages*, 33–49. Ann Arbor: The University of Michigan.

—— 1995. Subject, Topic, and Tagalog syntax. In: David Benett, Theodora Bynon and George B. Hewitt, eds *Subject, voice and ergativity*, 161–201. London: SOAS.

Noonan, Michael. 1992. *A grammar of Lango*. Berlin: Mouton de Gruyter.

Noorduyn, J. 1991. The languages of Sulawesi. In H. Steinhauer, ed. *Papers in Austronesian Linguistics* No.1. Canberra: Pacific Linguistics.

Palancar, Enrique L. 2002. *The origin of Agent Markers*. Studia Typologica 5. Berlin: Akademie Verlag.

Pawley, Andrew. 1977. Lecture given at the Linguistic Society of America's Summer Institute.

Pawley, Andrew and Malcolm Ross. 1993. Austronesian historical linguistics and culture history. *Annual Review of Anthropology* 22. 425–459.

Plank, Frans. 1979. Ergativity, syntactic typology and universal grammar: some past and present viewpoints. In Frans Plank, ed. *Ergativity: towards a theory of grammatical relations*. London: Academic Press.

Prentice, D.J. 1971. *The Murut languages of Sabah*. Canberra: Pacific Linguistics.

Post, Ursula and Mary Jane Gardner (compilers). 1992. *Binukid dictionary*. Studies in Philippine Linguistics, 9:2. Manila: Linguistic Society of the Philippines and Summer Institute of Linguistics.

Rackowski, Andrea. 2002. The structure of Tagalog: specificity, voice, and the distribution of arguments. PhD thesis, Massachusetts Institute of Technology.

Reid, Lawrence. 1992. On the development of the aspect system in some Philippine languages. *Oceanic Linguistics* 31.1:65-91.

—— 2002. Determiners, nouns, or what? Problems in the analysis of some commonly occurring forms in Philippine languages. *Oceanic Linguistics* 41.2:295–309.

Reid, Lawrence and Hsiu-Chuan Liao. 2004. A brief syntactic typology of Philippine languages. *Language and Linguistics* 5.2:433-490.

Richards, Norvin. 2000. Another look at Tagalog subjects. In Ileana Paul, Vivianne Phillips and Lisa Travis, eds *Formal issues in Austronesian linguistics*, 105–116. Dordrecht: Kluwer.

Rizzi, Luigi. 1990. *Relativized minimality*. Cambridge: Massachusetts Institute of Technology Press.

Ross, Malcolm. 1995a. Reconstructing Proto-Austronesian verbal morphology: evidence from Taiwan. In Paul Jen-kuei Li, Dah-an Ho, Ying-kuei Huang, Cheng-hwa Tsang and Chiu-yu Tseng, eds *Austronesian Studies Relating to Taiwan*. Symposium Series of the Institute of History and Philology, Academica Sinica 4, 727–791. Taipei: Institute of History and Philology, Academica Sinica.

—— 1995b. Some current issues in Austronesian linguistics. In Darrell Tryon, ed. *Comparative Austronesian dictionary: an introduction to Austronesian studies. Part 1: Fascicle 1*, 45–120. Berlin: Mouton de Gruyter.

—— 2002. The history and transitivity of western Austronesian voice marking. In Wouk and Ross, eds 2002:63–78.

—— 2006. The argument structure of undergoer voice clauses in Formosan and other Philippine type languages. MS. Research School of Pacific and Asian Studies and Centre for Research on Language Change, The Australian National University.

Schachter, Paul and Fe T. Otanes. 1972. *Tagalog reference grammar*. Berkeley: University of California Press.

Scheerer, Otto. 1924. On the essential difference between the verbs of the European and the Philippine languages. *Philippine Journal of Education* 7:1–10.

Starosta, S. Pawley, A. and Reid L. 1982. The evolution of focus in Austronesian. In Amran Halim, Lois Carrington and S.A. Wurm, eds *Papers from TICAL*, vol. 2, 201–216. Canberra: Pacific Linguistics.

Tjia, Johnny. 2007. A grammar of Mualang, an Ibanic language of Western Kalimantan, Indonesia. PhD thesis, University of Leiden. LOT Publications.

Tuuk, H.N. van der. 1864-1867. *Tobasche Sprakkunst.* 2 vols. Amsterdam: Muller.

Wolff, John U. 1973. Verbal inflection in Proto-Austronesian. In A.B. Gonzalez, ed. *Parangal Kay Cecilio Lopez*, 71–91. Quezon City: Philippine Journal of Linguistics (Special monograph 4).

—— 1996. The development of the passive verb with pronominal prefix in Western Austronesian languages. In Bernd Nothofer, ed. *Reconstruction, classification, description – festschrift in honor of Isidore Dyen*, (Asia-Pacific 3), 15–40. Hamburg: Abera.

Woollams, Geoff. 2005. Karo Batak. In Adelaar and Himmelmann, eds 2005:534–561.

Wouk, Faye and Malcolm Ross, eds. 2002. *The history and typology of Western Austronesian voice systems*. Canberra: Pacific Linguistics.

Wu, Jing-lan Joy. 2006. Verb classification, case marking, and grammatical relations in Amis. PhD thesis, University of Buffalo, New York.

Yamada, Yukihira and Shigeru Tsuchida. 1975. *Philippine languages, Asian and African grammatical manual No. 15b*. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa.

Zobel, Erik. 2002. The position of Chamorro and Palauan in the Austronesian family tree: evidence from verb morphosyntax. In Wouk and Ross, eds 2002:405–436.

16 *Start and finish: some grammatical changes in Toqabaqita*

FRANTISEK LICHTENBERK

1 Introduction[1]

In this paper I want to consider some grammatical changes, at least some of which involve grammaticalisation, in the Toqabaqita language. Specifically, I will consider two phasal verbs, one that means 'end, finish; be finished' (§3) and one that means 'start, begin' (§4), and one grammatical element that can function as a conjunctive coordinator 'and' (§5). The histories of these three elements bear on various issues that are relevant to our understanding of grammatical change, specifically grammaticalisation. The issues involved include functional splits without formal differentiation (here absence of phonological erosion), syntactic restructuring without a (significant) semantic difference, structural ambiguity, extension based on analogy, and communicative need/problem solving and its presumed role in grammaticalisation. Some of these issues are revisited in the concluding section, §6. There grammaticalisation is characterised as a development of a grammatical element from a lexical element or from another, less grammatical element, a type of grammatical change that is unidirectional, by definition.

However, before we look at the grammatical phenomena, it is necessary to have some information on the relevant parts of the grammar of the language. These are discussed in §2.

2 Grammatical background

Toqabaqita is an Oceanic language spoken in the Solomon Islands. Within Oceanic, it is a member of the Southeast Solomonic subgroup. The basic constituent orders in verbal clauses are shown in (1):

[1] It gives me great pleasure to contribute to this volume honouring Bob Blust. An earlier version of this paper was delivered at the Symposium on Language Change and Linguistic Variation, held at Ludwig-Maximilian University of Munich, 8–9 November 2007. I am grateful to a number of participants at the Symposium for their comments on that version. I also wish to thank the referees and Andrew Pawley for valuable comments on an earlier version of this paper.

Alexander Adelaar and Andrew Pawley, eds
Austronesian historical linguistics and culture history: a festschrift for Robert Blust, 227–245.
Pacific Linguistics, 2009.
© This edition vested with Pacific Linguistics.

(1) a. intransitive clauses: S V
b. transitive clauses: A V O

Verbs may be accompanied by a variety of particles, some of which come before the verb, and some after the verb. The particles express notions such as Aktionsart, directionality, aspect, intensification, and several others. Example (2), with an intransitive verb, contains the immediate-past/immediate-future preverbal particle *biqi* and the postverbal assertive particle *boqo*.

(2) *Nau ku biqi fula boqo*[2]
1SG 1SG.NFUT IMM arrive ASRT
'I have just arrived.' 'I arrived just now.'

Among the preverbal particles are subject-tense/aspect/etc. markers, which simultaneously index the subject with respect to person, number and clusivity (inclusive/exclusive) and express tense, aspect, sequentiality, negation or dehortation. (The dehortative set is archaic.) The subject markers occur before the other preverbal particles. In (2) above, the subject marker is first person singular nonfuture tense *ku*; in (10) in §3.1 the subject marker is third person singular sequential *ka*; and in (11), also in §3.1, the subject marker is first person singular future tense *kwai*.

Lexical direct objects follow the postverbal particles. This is shown in (3), where the object noun phrase *wela baa* 'the child' follows the combination of the perfect and the andative directional particles:

(3) *Kera tole-a na=kau*[3] *wela baa.*
3PL.NFUT lead-3.OBJ PRF=AND child that
'They have led the child away.'

Pronominal direct objects come immediately after the verb, before any of the postverbal particles. In example (4) the third person singular independent pronoun *nia* precedes the combination of the perfect and the andative directional particles:

(4) *Kera tole nia na=kau.*
3PL.NFUT lead 3SG PRF=AND
'They have led her away.'

---

[2] The following abbreviations are used in glossing the examples: 1, 2, 3 – first, second, third person, ABL – ablative, ALL – allative, AND – andative, ASRT – assertive, ASSOC – associative, ATTN – attenuative, CAUS – causative, CLF – classifier, COM – comitative, COMP – complementiser, COMPL – completive, DETR – detransitive, DIM – diminutive, DU – dual, DVN – deverbal noun, EXCL – exclusive, EXHST – exhaustive, FADD – foregrounding-additive, FOC – focus, FUT – future, GENP – general preposition, IMM – immediate past/immediate future, IN – inessive, INCEP – inceptive, INCL – inclusive, INTS – intensifier, IPFV – imperfective, LIG – ligature, LIM – limiter, LIP – low individuation of participants, LOC – (general) locative, NEG – negative, NEGV – negative verb, NFUT – nonfuture, NMLZ – nominaliser, OBJ – object, PERS – personal, PERS.MKR – person marker, PL – plural, PRF – perfect, PROFORE – pronominal foregrounder, PROL – prolative, PRTT – partitive, PURP – purpose, RDP – reduplication, RECBEN – recipient-benefactive, REL – relative, SEQ – sequential, SG – singular, TR – transitive, VENT – ventive, VIVID – vivid.

The Toqabaqita data come from my field notes. The sources of the other data are as acknowledged in the text.

[3] The equal sign = designates phonological fusion of two or even three words, one or both/all of which occur in a reduced form. The full form of the perfect marker is *naqa*; see (10) in §3.1.

There are two basic classes of transitive verbs in the language, Class 1 and Class 2, distinguished by a number of properties having to do with object indexing. Class 2 transitive verbs have object-indexing suffixes for all grammatical persons and numbers. On the other hand, Class 1 transitive verbs have object-indexing suffixes only for the third person, singular, dual, and plural. It is Class 1 transitive verbs that will be relevant here. With Class 1 transitive verbs, objects other than third person can be expressed only by means of independent personal pronouns. In (5) the direct object is second person singular:

(5) *Nau ku riki qoe …*
1SG 1SG.NFUT see 2SG
'I saw you …'

In the case of third person objects, there are two options. One is to use an independent pronoun, as in (6a), and the other is to use an object suffix on the verb, as in (6b), except that the independent pronouns are not normally used with reference to inanimates or lower animates.

(6) a. *Qo riki nia?*
2SG.NFUT see 3SG
'Did you see him/her?'

b. *Qo riki-a?*
2SG.NFUT see-3SG.OBJ
'Did you see him/her/it?'

The example in (7) shows the verb 'see' with the third person plural object suffix *-da*:

(7) *Qo riki-da?*
2SG.NFUT see-3PL.OBJ
'Did you see them?

The third person singular object suffix *-a*, shown in (6b), is also used to index lexical direct objects, regardless of their grammatical number. Since with lexical objects the suffix does not specify the grammatical number, there is no specification of number in the glosses. In (8) the object is singular, and in (9) plural:

(8) *Nau ku rongo-a kini qeri.*
1SG 1SG.NFUT hear-3.OBJ woman that
'I heard the woman.'

(9) *Nau ku rongo-a kini qe=ki.*
1SG 1SG.NFUT hear-3.OBJ woman that=PL
'I heard the women.'

The suffix *-a*, used to index lexical direct objects, also indexes object complement clauses. This object suffix and the existence of preverbal and postverbal particles will be relevant to some of the discussion that follows.

I will start my discussion of the grammatical developments in Toqabaqita with the element that can function as the verb 'end, finish; be finished'.

3 *Sui* ‘end, finish; be finished’, etc.

The form *sui* has a variety of meanings and functions, lexical and grammatical. These are discussed in §3.1 and §3.2 and are summarised in Table 1 in §3.3.

3.1 *Sui* as a verb

Sui can function as an intransitive verb ‘end, finish; be finished’:

(10) *Firu-a* *naqi* *ka* *sui* *naqa.*
fight-DVN this 3SG.SEQ end PRF
‘This fight should/must end now.’

Like many other intransitive verbs, *sui* can serve as the input into a causative derivation. That is, there is a causative transitive verb ‘finish’:

(11) *Nau* *kwai* *faqa-sui-a* *raa* *naqi* *qi* *taraqena.*
1SG 1SG.FUT CAUS-be.finished-3.OBJ work this LOC today
‘I will finish this work today.’

And like any other verb, *sui* can be nominalised:

(12) *sui-la-na* *fa* *ngali*
be.finished-NMLZ-3.PERS CLF year
‘the end of the year’

As a verb, *sui* ‘end, finish; be finished’ is often used to signal the end of a state of affairs expressed in another, preceding clause:

(13) *Kamiliqa* *mili* *kwaqe-a* *masuqu* *loo* *[e*
1PL(EXCL) 1PL(EXCL).NFUT cut-3.OBJ bush upward 3SG.NFUT
sui *naqa].*
be.finished PRF
‘We have finished cutting down the bush up there.’ (lit. ‘We cut down the bush up there, it is finished.’)

Although in this function *sui* is a verb and heads the predicate of a clause, that clause is severely restricted in terms of what it can contain. For convenience, the type of clause in which *sui* occurs in cases like (13) will be referred to as a ‘mini-clause’. In (13) above and (14) below, the mini-clauses are in square brackets. A mini-clause can contain, in addition to the verb, only a third person singular subject marker which is either nonfuture (as in [13]) or sequential (as in [14]). In addition, *sui* mini-clauses often contain the perfect marker and may contain an assertive marker/intensifier (for emphasis). In (13) only the perfect marker is present; in (14) both the perfect marker and the assertive marker are.

(14) *Iu,* *si* *uqunu* *qeri* *qe* *fula* *qi* *kula*
OK DIM story this 3SG.NFUT arrive LOC place
qeri *[ka* *sui* *bo=naqa].*
this 3SG.SEQ be.finished ASRT=PRF
‘OK, this is the end of this little story.’ (lit. ‘OK, this little story (has) arrived at this place, it is finished.’)

Even though constructions with *sui* mini-clauses are biclausal in the sense of containing two predicates, the two clauses are tightly linked prosodically: unlike elsewhere in coordination, there is normally no rise in intonation at the end of the first clause, and there is normally no pause between the two clauses. Nevertheless, this construction does not involve verb serialisation: not just direct objects, but also oblique objects can come between the two verbs; see (13) for the former, and (14) above and (15) below for the latter. In (15), the prepositional phrase between the two verbs is in square brackets:

(15) *Qe* *alu=[i* *gwau-na* *bete-qe* *duqu* *ba]=e*
3SG.NFUT put:3SG.OBJ=LOC top-3.PERS post-ASSOC ridge that=3SG.NFUT

sui, *nia* *ka* *gale-a* *qato, ….*
be.finished 3SG 3SG.SEQ make.V.shaped.cut.in-3.OBJ rafter
'When he [a man building a house] has finished putting it [a ridgepole] on top of the kingpost, he makes a V-shaped cut in the rafter, ….' (lit. 'He has put it on top of the kingpost, it is finished, he makes a V-shaped cut in the rafter, ….')

As (15) shows, *sui* mini-clauses can also be used to express sequences of states of affairs, especially to highlight the fact that the first state of affairs reaches completion before the second one obtains. The first state of affairs is expressed in the first clause, preceding the *sui* mini-clause, and the second state of affairs is expressed in the clause following the *sui* mini-clause. Here too, the *sui* mini-clause is prosodically integrated with the preceding clause. On the other hand, there is normally a prosodic break between the *sui* mini-clause and the following clause: a rise in intonation at the end of the *sui* mini-clause and usually a slight pause. The basic structure of the overall construction is given schematically in (16), where SoA stands for 'state of affairs':

(16) [[clause of 1st SoA] [*sui* mini-clause]] [clause of 2nd SoA]

Another example is given in (17):

(17) *Kera* *thau-ngi-a* *ulu* *wela* *qe=ki* *qe* *sui,*
3PL.NFUT kill-TR-3.OBJ three child that=PL 3SG.NFUT be.finished

keka *lae* *na-da.*
3PL.SEQ go PRF-3PL.PERS
'After they killed the three children, they went (away).' (lit. 'They killed the three children, it was finished, they went.')

3.2 Other functions of *sui*

Besides being a verb, the form *sui* has several other functions. First, it can function as a sequencing marker 'then'. The sequencer occurs clause-initially. When it functions as a sequencer, *sui* is not a verb. It can never have its own subject marker and does not form a clause of its own. Prosodically, it is part of the clause that signals the second state of affairs. The basic structure of this construction is shown schematically in (18):

(18) [clause of 1st SoA] [*sui* rest of clause of 2nd SoA]

For example:

(19) *Saqu-a qaba-mu, [sui feteqi qoko biqi fanga].*
wash-3.OBJ hand-2SG.PERS then INTS 2SG.SEQ IMM eat
'Wash your hands and then (and only then) eat.' 'Wash your hands before you eat.'

The sequencer can occur sentence-initially, without a clause expressing the first state of affairs in the same sentence:

(20) *Kukeqe nia bo=naqa na kai qegwe-a,*
wife 3SG ASRT=PRF FOC 3SG.IPFV extract.kernel.out.of-3SG.OBJ

bia wela nia ki. Sui keka ngali-a
COM child 3SG PL then 3PL.SEQ take-3.OBJ

bii-na ngali kera qeri,
bamboo.container-3.PERS canarium.nut 3PL that

keka alu-a bo=naq=i luma, …
3PL.SEQ put-3SG.OBJ ASRT=PRF=LOC house
[When the quantity of canarium nuts is relatively small, the man will crack them by himself, without the help of other men.] 'It will be his wife that will extract the kernels out of them, (together) with his children. Then they will take the bamboo-containerfuls of their canarium nuts into the house, …'

The etymon *sui* has some other functions, all of which are grammatical. In one, it functions as a postverbal completive particle, to signal that the state of affairs expressed in the clause has been completed. There is no necessary implication of another state of affairs following. The completive particle is typically accompanied by the perfect marker, as in (21):

(21) *Nia qe thau-ngani-a sui naqa luma nia.*
3SG 3SG.NFUT build-TR-3.OBJ COMPL PRF house 3SG
'He has finished building his house.'

Even though as a completive marker *sui* can directly follow a verb, as in (21), it is not the case that the verb and *sui* form a serial verb construction. Toqabaqita does have serial verb constructions of a compounding type, with the two (or more) verbs obligatorily occurring next to each other. However, if in a serial verb construction the first/non-final verb is Class 1 transitive (§2), it cannot carry an object suffix, even though outside of serialisation such verbs do take third person object suffixes. If the second/final verb in such a serial verb construction is intransitive, the whole serial construction is intransitive. The participant that would otherwise be expressed as the direct object is instead expressed as an oblique object. In (22) the verb 'build, make' does not form part of a serial verb construction. It has a direct object, which is indexed on the verb by the object suffix *-a*:

(22) *Kini kai thau-ngani-a qa-kuqa teqe teeteru.*
woman 3SG.FUT make-TR-3.OBJ RECBEN-1SG.PERS one fan
'The woman will make me a fan.'

However, when the verb 'build, make' occurs as the non-final verb in a serial construction and the final verb is intransitive, the patient participant is expressed as an oblique object, not as a direct object. The verb 'build, make' has no object suffix, and, furthermore, it has the form *thau-ngaqi*, rather that *thau-ngani*:

(23) *Kini qe thau-ngaqi qaliqali qana teeteru.*
woman 3SG.NFUT make-TR be.fast GENP fan
'The woman made a/the fan quickly.'

In (21) further above, the patient participant is expressed as a direct object, the verb 'build, make' carries the object suffix *-a*, and has the form *thau-ngani*. There is no serial verb construction in (21); *sui* is a particle, not a verb.

Another grammatical function that the form *sui* has is as a particle in noun phrases. As a noun-phrase internal particle, *sui* has an 'exhaustive-marking' function: it signals that the state of affairs expressed in the clause applies to all the relevant participants.

(24) *Wela nau ki sui boqo kera sukulu qi manga qeri.*
child 1SG PL EXHST ASRT 3PL.NFUT attend.school LOC time this
'All of my children attend school at this time.'

The link between the completive function of *sui* with verbs and the exhaustive function in noun phrases is not difficult to see. In many cases, a state of affairs that involves multiple participants in the same role can be seen as being completed only when it applies to all those participants.

There is one more grammatical function that the etymon *sui* has: it can serve as a coordinator to connect clauses when the relation between two propositions is one of unexpectedness or contrast. In this function *sui* is glossed 'but'. In (25) the relation between the propositions is one of unexpectedness: the man was very big but/even though still young:

(25) *Wane qeri wan=daraa n=naqa, thaama-na* [4]
man that man=be.young.and.single LIG=PRF father-3.PERS

wane baqita naqa, sui nia wan=daraa.
man be.big PRF but 3SG man=be.young.and.single
'The man was already a *daraa* man, a very big man, but (still) only a *daraa* man.'
(*Daraa* is used about males who are no longer considered children, but are still relatively young and not yet married.)

On the other hand, in (26) the relation between the propositions is one of contrast: the people spoken of in the first clause watched out for an imaginary enemy, while the American soldiers watched out for a real enemy:

(26) *Keka lio, maqasi-a maqalimae uri-a gaetemu-la-na*
3PL.SEQ look wait.for-3.OBJ enemy PURP-3.OBJ guide-NMLZ-3.PERS

figu-a kera ki, sui Merika qe=ki kera lio,
gather-DVN 3PL PL but America that=PL 3PL.NFUT look

maqasi-a maqalimae mamana.
wait.for-3.OBJ enemy be.real
'[During the Marching Rule movement] they [the adherents to the movement] watched, waiting for an [imaginary] enemy in order to lead their groups, but/while/whereas the Americans watched, waiting for a real enemy.'

---

[4] *Thaama-na*, lit. 'his father', is used here with the sense 'big'.

In its function as a coordinator, *sui* is often reinforced by means of one of two other elements. In (27) it is reinforced by means of *mena*, which also functions as a concessive marker (‘although’), also glossed ‘but’ here:

(27) *Nau ku dora qana wane naqi, sui mena nau*
1SG 1SG.NFUT not.know GENP man this but but 1SG

ku toqomatafa-na wane leqa ni bana.
1SG.NFUT have.feeling.that-3.OBJ man be.good LIG LIM
‘I don’t know this man, but/still/nevertheless I have a feeling (he is) just a good man.’ (There is nothing bad about him.)

And in (28) *sui* is reinforced by the element *taa*, also glossed ‘but’, which elsewhere functions as an interrogative noun ‘what?’ and also as an interjection ‘lo!, behold’. It is probably the latter use of *taa* with its sense of surprise, unexpectedness that is more closely related to the coordinating function:

(28) *“Qo laa-lae neri, sui taa qe aqi qosi*
2SG.NFUT RDP-go VIVID but but 3SG.NFUT NEGV 2SG.NEG

ngali-a ma=ta si doo.”
take-3.OBJ VENT=some PRTT thing
‘[One man got angry, and said], “You went (lit. had gone) [to that place], but you didn’t bring anything back [as gifts]”.’ (One is normally expected to bring back gifts from a foreign place.)

3.3 Comparative evidence

Table 1 summarises the various meanings and functions of the etymon *sui* in Toqabaqita. Cognates of *sui* are found in languages closely related to Toqabaqita, and in some of them the cognates have grammatical functions. The meanings and functions of cognates of *sui* are given in Table 2.

None of the sources specifically mentions the function of *sui* as a completive marker used with verbs. It is not possible to tell whether this means that *sui* does not have this function in those languages, or whether it is, or is considered to be, a verb when it has that function.

At least some of the polysemy/heterosemy goes back some way, even though the possibility of some parallel independent innovations cannot be discounted. (The term ‘heterosemy’ designates a situation where the reflexes of a proto-form belong in two or more different morphosyntactic categories; see Lichtenberk 1991a.) Inakona, Tolo and Bugotu are members of the Guadalcanal-Nggelic subgroup of Southeast Solomonic, while all the other languages, including Toqabaqita, are members of the Cristobal-Malaitan subgroup. The exhaustive-marker function (‘all’) and possibly also the sequencer function (‘then, afterwards’) are reconstructible to the Proto Southeast Solomonic stage.

Table 1: Meanings and functions of *sui* in Toqabaqita

Verb:	'end, finish; be finished'; also used in a mini-clause to highlight completion of a state of affairs
Grammatical functions:	clause/sentence-initial sequencer postverbal completive marker noun-phrase internal exhaustive marker clause coordinator to express unexpectedness or contrast

Table 2: Cognates of Toqabaqita *sui*

Language	Meanings/functions
Lau	*sui* verb 'be finished'; sequencer 'then, thereupon, after that'; exhaustive marker 'all, the whole, both'; *suitā* contrast/ unexpectedness coordinator? 'well then, that being so, however, but then' (*tā* 'what, whatever') (cf. Toqabaqita *sui taa*) (Fox 1974:179)
Kwara'ae	*sui* verb 'finish; be finished' (pp.36, 65); *sui*, *suita* sequencer 'then' (p.81) (*ta* 'what?'); exhaustive marker 'all, every' (p.31); *sui bore* 'adversitive, antithetical' conjunction 'but' (p.83) (*bore* 'although') (Deck 1934)
Kwaio	*sui* verb 'finished, completed, used up' (Keesing 1975:189); *kee sui* 'and then' ('or simply to indicate that the action of a preceding verb finally finished') (*kee* temporal particle) (Keesing 1985:121)
'Āre'āre	*sui* verb? 'finished, ended'; exhaustive marker 'all, everybody, everything' (Geerts 1970:113)
Inakona	*sui* verb 'finished'; exhaustive marker 'all' (p.134): *gira sui* 'all those' (*gira* 3pl pronoun) (p.118) (Capell 1930)
Tolo	*sui* verb 'be finished, completed, be over', *suilania* (transitive) 'finish, complete' (*-a* 3sg object suffix); *sui* sequencer? 'afterward(s) (when that is finished)' (Crowley 1986:46)
Bugotu	*hui*[5] verb 'take down', 'let down', 'cease', 'finish', 'redeem'; *vahuihui* 'last, final' (*va-* causative prefix) (Ivens 1940:20)

[5] Ivens calls *hui* 'v[erb] t[ransitive]', but in the sole example of use he gives (p.20) *hui* appears to be an intransitive verb: *hui gohi na tarai* 'are prayers over?' (*gohi* 'already', *na* indefinite [?; F.L.] article, *tarai* 'pray', 'prayer').

Grammaticalisation is sometimes said to be characterised by phonological erosion. However, there is no erosion in any of the languages that have reflexes of *sui. Both the lexical and the various grammatical elements have the same form, *sui*, even though at least the exhaustive-marker function is of considerable antiquity. In Toqabaqita, *sui* is one of a large number of postverbal particles. These are listed in Table 3, arranged according to their position relative to the verb.

Table 3: Toqabaqita postverbal particles

Verb	
fasi, si	precedentive
quu	anterior, anterior-continuing
sui, danga	completive (*danga* is rare)
laqu	additive, restitutive
qasia	intensifier, only together with *boqo* or *naqa*
toqo	obligation
boqo	assertive, intensifier
mena	concessive
qa-	self-benefactive, recipient-benefactive
ba-	limiter
naqa	perfect, intensifier in combination with *qasia*
feteqi	intensifier
mai	ventive
kau	andative

A few of the particles have reduced, combining forms, which are obligatorily used instead of the full forms in the presence of certain other, immediately following particles. The reduced form of the first particle and the following particle fuse together into one phonological word. For example, the assertive/intensifier particle has the full form of *boqo*, but when it is followed by the perfect marker *naqa*, the two obligatorily fuse together as *bo=naqa*; cf. (2) in §2 and (14) in §3.1. On the other hand, the completive particle *sui* does not have a reduced, combining form.

Grammaticalisation is often portrayed as a unilinear process, a certain meaning or function X being replaced by another function Y, which may be accompanied by a difference in form (for example due to phonological erosion), with the old and the new functions co-existing (X, Y) for some time, as shown schematically in (29) (see, for example Hopper and Traugott 1993:36):

(29) X ⟶ X, Y ⟶ Y

However, frequently the process is not unilinear. Rather, there is a split, divergence in functions, as diagrammed in (30):

(30) X ⟶ X
X ⟶ Y

The original form X continues as such, with the same meaning/function, but another form Y develops from it, with a different function. It is, of course, possible for X and/or for Y to undergo later changes in form and/or in function.

The existence of such splits is recognised (for example, Heine and Reh 1984, Hopper 1991), but perhaps not given sufficient recognition. (But see, for example, Craig [1991] for discussion of 'polygrammaticalisation' in Rama, spoken in Nicaragua.) It is precisely such splits that give rise to networks of polysemy and heterosemy, so common in languages. This is what happened in the case of *sui* in Toqabaqaita (and some of its close relatives). And as we will see in §4 and §5, splits are also found in the other two cases of grammatical change discussed here.

4 *Thafali* (intransitive, transitive) 'start, begin'; inceptive

The case I will discuss now also involves a phasal verb, but the change the verb figures in is very different from those in which the verb *sui* 'end, finish; be finished' is involved. This development is discussed in greater detail in Lichtenberk (2008), together with developments of two other verbs, 'like, want' and 'not like, not want', which undergo parallel changes that result in the same type of outcome. Here only the salient points are given that have to do with the verb 'start, begin'.

Toqabaqita has an intransitive and a transitive verb *thafali* 'start, begin'. The intransitive verb is present in (31) (from a letter) and the transitive one in (32).

(31) *Nau kwai sifo fanu kia qana Christmas holiday, kai*
1SG 1SG.FUT descend place PL(INCL) GENP Christmas holiday 3SG.FUT

thafali qana 28th Nov.
start GENP 28th November
'I will go down to our place for Christmas holidays; they will start on the 28th of November.'

In (32) there are two instances of the transitive verb. The second instance has a lexical, non-clausal noun phrase as its direct object, while the first instance has a clausal object, which is in square brackets. Both types of direct object are indexed on the verb 'start, begin' by the suffix *-a* (§2).

(32) *Manga na keki thafali-a [keki qani-a oqola kera*
time REL 3PL.FUT begin-3.OBJ 3PL.FUT eat-3.OBJ garden 3PL

qeri], keka lae keka thafali-a laqu boqo ta si
that 3PL.SEQ go 3PL.SEQ start-3.OBJ FADD ASRT some PRTT

kula faalu.
place be.new
'When (lit. the time that) they begin to eat (food from) that garden of theirs, they go and start a new place [to make it into a garden for the next crop].'

It is the transitive verb 'start, begin' that will be of relevance here, in particular when it takes a clausal object.

As shown in (32), the clausal complement of the verb 'start, begin' is finite. The clause has its own subject, which must be coreferential with the subject in the higher, 'begin' clause, and its subject marker must be future. In (32) the subject of the complement clause is indexed by the third person plural future-tense subject marker *keki*. There are then two verbs in this kind of structure, one in the main clause and one in the complement clause.

Toqabaqita has a complementiser of the form *na*, which, however, is not used very commonly and is never obligatory. It is used in (33), with a complement clause under the verb 'begin':

(33) *Nau ku thafali-a [na kwai uqunu suli-a*
1SG 1SG.NFUT begin-3.OBJ COMP 1SG.FUT narrate PROL-3.OBJ

tha Bariqi].
PERSMKR Bariqi
'I am beginning to tell the story about Bariqi [a man].'

Sentences like those in (32) and (33) are bi-clausal. Another example is given in (34):

(34) *Nau ku thafali-a [kwai uqunu naqa].*
1SG 1SG.NFUT begin-3.OBJ 1SG.FUT narrate PRF
'I am about to begin to tell a/the story.'

However, there is an alternative construction where the form *thafali* does not carry an object suffix. Instead, it carries a detransitivising suffix *-qi*, as in (35), which is synonymous with (34).

(35) *Nau ku thafali-qi uqunu naqa.*
1SG 1SG.NFUT INCEP-DETR narrate PRF
'I am about to begin to tell a/the story.'

There is evidence that constructions with *thafali-qi* are monoclausal, rather than bi-clausal, and that *thafali-qi* is not a verb but rather a preverbal particle, an inceptive marker. First, there is prosodic evidence: sentences like (35) are said under one intonation contour, whereas in the biclausal structure there is often a rise in intonation at the end of the first, matrix clause and there may be a slight pause between the matrix and the embedded clauses. Second, the inceptive element does not carry the object suffix *-a*, which is evidence that there is no direct object. Third, there can be only one subject-tense marker, which comes before the inceptive marker. And finally, the complementiser *na* is ungrammatical in sentences with the inceptive marker.

The same historically detransitivising suffix is also found on some other preverbal particles, although it so happens that there it has the form *-qe* rather then *-qi*, because the vowel in the preceding syllable is non-high. None of those particles has a corresponding verbal form in the present-day language. One such particle is present in (36), and as shown there, the particle may occur with or without the suffix *-qe*. (The *a*-to-*e* change before the suffix is regular.) The particle in question has an 'attenuative' function: it signals that the event is (to be) performed in an attenuated way, calmly, slowly, quietly or carefully. Because here the suffix does not have a detransitivising function, it is glossed, for convenience, QI.

(36) *Qoko {thafa/ thafe-qe} ngata.*
2SG.SEQ ATTN/ ATTN-QI speak
'Speak calmly/slowly/quietly.'

There are several such particles in the language, and they are all preverbal. It is possible that the immediate past/immediate future particle *biqi*, shown in (2) in §2, historically also contains the detransitivising suffix.

With the preverbal particles the suffix is optional. This is also true of the inceptive marker. In (37) the suffix is absent:

(37) *Diapana ka thafali lae mai uri fanu kia …*
Japan 3sg.SEQ INCEP go VENT ALL country PL(INCL)
'Japan/The Japanese began to come to our country …'

The detransitivising suffix *-qi*/*-qe* is the same etymon as the associative suffix *-qi*/*-qe* used on head nouns in associative noun phrases. (For discussion of the latter see Lichtenberk 2006.) The conditions on the use of the two allomorphs of the associative suffix are the same as those on the use of the variants of the detransitivising suffix: *-qi* after high vowels, *-qe* after non-high vowels. Although the associative suffix is more often than not present, it may be omitted:

(38) *fungi(-qi) baqu*
bunch-ASSOC banana
'bunch of bananas'

Some linguists have argued that there is a cognitive asymmetry between main and subordinate clauses. For example, Langacker (1991:436) says that in constructions with complement clauses the main clause 'lends its profile' to the construction as a whole; it is the profile determinant. It is the main clause that designates the overall state of affairs. So the sentence in (39) designates the state of affairs of the speaker's knowing, not that of the woman's leaving:

(39) I know she left.

Cristofaro (2003) frames the notion of asymmetry in terms of a clause expressing or not expressing an assertion. A main clause encodes an asserted state of affairs, while a dependent clause encodes a non-asserted state of affairs. So, in (39) what is asserted is the speaker's knowing [that she left].

The question I want to ask now is what happens to this asymmetry if through a grammaticalisation process a former bi-clausal structure becomes a monoclausal one. In a construction with *thafali* as a matrix verb, as in (34), what is presumably asserted is the person's *beginning* to tell a/the story. Is that also the case in the corresponding monoclausal structure in (35)? Unfortunately, I do not have evidence to answer this question one way or the other, but I do have some evidence which might be relevant in trying to answer it.

During one of my field trips I recorded a text where the speaker produced the construction in (40), with the verb 'begin' with the object suffix *-a* and a finite complement clause:

(40) *Manga na keki thafali-a [keki qani-a*
= (32) time REL 3PL.FUT begin-3.OBJ 3PL.FUT eat-3.OBJ
oqola kera qeri], ...
garden 3PL that
'When (lit. the time that) they begin to eat (food from)
that garden of theirs, …

However, later, as I was transcribing the recording of the text, the speaker said, several times, as he was listening to the recording and dictating to me, what is in (41), with a reduced, monoclausal structure:

(41) ... *keki* *thafali-q* *qani-a* …
3PL.FUT INCEP-DETR eat-3.OBJ
'... they begin to eat ...'

(In [41] the vowel of the suffix *-qi* has been deleted before a word that starts with *q* [glottal stop]. The deletion is optional.)

What was on tape and what the speaker dictated to me were two different constructions: a biclausal one with a finite complement clause, and a reduced, monoclausal one, respectively. But, evidently, to him they were semantically identical, or sufficiently close in their meanings. Does one sentence exhibit asymmetry and the other one not?

There is a danger in relying exclusively on linguistic evidence when dealing with asymmetry between states of affairs, a cognitive phenomenon. There should be other kinds of evidence that such asymmetry exists and which way it goes, rather than just linguistic structures (such as the presence or absence of subordination).

5 *Bia, bii* 'and'

In this section I want briefly to consider a different kind of grammatical change that has to do not with verbs but with a conjunction. This is where we come to the 'and' part of the title of the paper.

Toqabaqita has a conjunctive coordinator *ma* 'and' used to conjoin noun phrases and clauses, as in (42) and (43), respectively:

(42) *Nau* *ma* *wela* *nau* *ki* *mili* *too*
1SG and child 1SG PL 1PL(EXCL).NFUT be.in.certain.condition

siafaqa.
not.be.well.off
'I and my children are not well off.'

(43) *Ta* *fai* *noniqi* *wane* *keka* *lae* *uri-a* *raraqe-laa,*
some four CLF man 3PL.SEQ go PURP-3.OBJ hunt.possum-NMLZ

ma *ta* *fai* *noniqi* *wane* *keka* *lae* *uri-a*
and some four CLF man 3PL.SEQ go PURP-3.OBJ

nao-laa *qi* *laa* *kafo.*
bail-NMLZ LOC in stream
'Four of the [eight] men were going to go possum-hunting, and (the other) four men were going to go and bail water out of (lit.: in) the stream.'

There is another conjunctive coordinator 'and', one which has two variant forms, *bia* and *bii*. The form *bia* is used when the conjunct to the right is a lexical noun phrase, as in (44), and the form *bii* is used when the conjunct to the right is a pronominal noun phrase, as in (45):

(44) *Thaina-na* *bia* *maka* *nia* *kero* *lae* *qana* *uusi-a* …
mother-3SG.PERS and father 3SG 3DU.NFUT go GENP buy-DVN
'His mother and father went to the market …'

(45) … *ai* *qeri* *bii* *ni* *nau* *qe* *aqi*
woman that and PROFORE 1SG 3SG.NEG NEGV

mesi *kwai-naqo-fi.*
1DU(EXCL).NEG LIP-face-TR
'[Traditionally, when the wife of another man and I were speaking to each other,] the woman and I would not face each other.'

The forms *bia* and *bii* also function as a comitative preposition 'with'. *Bia* is used if the complement is lexical, *bii* is used if the complement is pronominal.

Apart from the restriction on the use of *bia* with lexical conjuncts to its right and *bii* with pronominal conjuncts to its right, the two forms are interchangeable with the other coordinator *ma*, which is used regardless of the type of conjunct. In (46) both *ma* and *bia* are used in neighbouring clauses, where reference is to the same participants:

(46) *Tootoo* *maka* *nia* *ma* *thaina-na* *keko* *lae* *naqa.*
later father 3SG and mother-3SG.PERS 3DU.SEQ go PRF

Maka *nia* *bia* *thaina-na* *keko* *sore-a,* …
father 3SG and mother-3SG.PERS 3DU.SEQ say-3SG.OBJ
'Later, his father and mother were about to go. His father and mother said, …'

In Lichtenberk (1991b) I reported on, among other things, a study of the use of the coordinators in texts. There were no occurrences of the form *bii*, and so only the forms *ma* and *bia* were considered. The results are shown in Table 4.

Table 4: The use of the coordinators *ma* and *bia* 'and' in a corpus of text (Lichtenberk 1991b:62)

	Noun phrase coordination		Clausal coordination	
	tokens	%	tokens	%
ma	53	63.9	352	100
bia	30	36.1	0	0

In the corpus used for that study there were no instances of *bia* (or *bii*) as clausal coordinators, and attempts to elicit such constructions were unsuccessful. However, subsequently, some years later, I encountered two spontaneous instances of such constructions, one with *bia* and one with *bii*; see (47) (from a letter) and (48), respectively:

(47) *Kada* *ku* *lae* *kwa* *too* *naqa* *i* *Yokohama*
time 3SG.NFUT go 1SG.SEQ stay PRF LOC Yokohama

Rehabilitation *Centre,* *kwai* *biqi* *thaitoqoma-na* *qani-la-na*
Rehabilitation Centre 1SG.FUT IMM know-3.OBJ eat-NMLZ-3.PERS

Japanese food, bia kwa biqi biinga leqa naqa.
Japanese food and 1SG.SEQ IMM sleep be.good PRF
'When I went and stayed at the Yokohama Rehabilitation Centre, I quickly learned (lit. knew) to eat Japanese food and quickly was able to sleep well (lit. quicky slept well).' (This after an initial period in Japan when the person could not eat Japanese food and could not sleep well.)

(48) *Laalae ku baqita naqa, kwa riki thaitoqoma-na*
when 1SG.NFUT be.big PRF 1SG.SEQ see know-3.OBJ

maka nau, bii kwa riki thaitoqoma-na thaina-ku.
father 1SG and 1SG.SEQ see know-3.OBJ mother-1SG.PERS
'When I had grown bigger, I recognised by sight my father and I recognised by sight my mother.'

According to a language consultant, *bia* in (47) could be replaced with *bii*, and *bii* in (48) could be replaced with *bia*, and both could be replaced with the more common coordinator *ma*, without any difference in meaning.

Comparative evidence tells us that the coordinator *ma* is of some antiquity: it can be reconstructed to pre-Proto Oceanic times. On the other hand, the use of *bia* and *bii* as noun phrase coordinators and clause coordinators is relatively recent. *Bia*/*bii* are reflexes of a transitive verb. Verbal cognates are found in Arosi: *bei* 'be one with, be an ally, partner', also 'assist, help' (Capell 1971:77), 'be partner, ally to', also 'assist, help' (Fox 1978:93), also *be'i* 'be in partnership with' (Fox 1978:93); and in Sa'a: *pe'i* 'be in the company of, along with', 'assist, help', also 'and' (Ivens (1929:254). Toqabaqita *bia* and *bii* function as a comitative preposition, and as a coordinator, not as a verb. Besides *bia* and *bii* there is also *laebiibii* 'walk close to s.o., walk next to s.o., walk closely behind s.o.' (*lae* 'go', 'walk'). Judging by comparative evidence, the comitative function historically precedes the coordinating function.

Structural ambiguity is frequently a factor that leads to reanalysis. One can see how structural ambiguity could lead to the development of *bia* and *bii* into noun phrase coordinators from a comitative preposition. The construction in (49) can be interpreted as involving a comitative construction or noun phrase coordination, without any great semantic difference:

(49) *Ku riki-a kukeqe bia wela nia ki.*
1SG.NFUT see-3.OBJ mature.woman COM/and child 3SG PL
a. 'I saw the woman with her children.'
b. 'I saw the woman and her children.'

On the other hand, it is difficult to see how structural ambiguity could lead to the development of the clause-coordinating function from the comitative function. It is much more likely that the at present rare clause-coordinating function is a later development from the noun-phrase coordinating function.

Communicative need or problem solving is sometimes given as a motivating factor in grammaticalisation (see, for example, Heine et al. 1991 and Cristofaro 1998). However, it is difficult to see how communicative need or problem solving might be responsible for the development of *bia* and *bii* into clause coordinators. The well-established coordinator *ma*

fulfilled this function in the past, and there was no communicative need for new clause coordinators, and there was no structural ambiguity.[6]

It was analogy-based extension that was at play here: *ma* was used both in noun-phrase coordination and in clause coordination, while *bia* and *bii* were earlier used only in noun-phrase coordination. The two types of coordinator were equivalent in noun-phrase coordination, and this equivalence was extended to clause coordination. One could still argue that the extension was functional because it increased the parallelism between the two sets of forms. However, I doubt it can be claimed that communicative need or problem-solving was a crucial factor here. Rather, it was extension due to analogy.

6 Conclusion and some remarks on grammaticalisation

Of the three sets of grammatical changes in Toqabaqita discussed here, at least two involve grammaticalisation: the development of the inceptive preverbal particle from the transitive verb 'start, begin', and the development of the conjunctive coordinator 'and' from the comitative preposition (assuming that the innovative function is somehow more grammatical than the source function). In the case of the third element, *sui* 'end, finish; be finished', sequencer, completive marker, exhaustive marker, contrast/unexpectedness coordinator, we cannot, in the absence of evidence about the initial morphosyntactic status of the etymon, be certain that grammaticalisation was involved, from a verb into the various grammatical elements, directly or indirectly. If the development went the other way, from one of the grammatical elements into the verb, this would be an instance of degrammaticalisation, a much rarer process than grammaticalisation.

All three cases involve divergence, splits, rather than replacement of the source form-function pairing by (a) new form-function pairing(s). The source elements continue alongside the new elements. Moreno Cabrera (1998:214) says that one of the properties of grammaticalisation is that '[i]t feeds the syntax and bleeds the lexicon'. However, there is no bleeding of the lexicon in the frequent cases where the source lexical form co-exists with the related grammatical form(s), such as the ones discussed here.

Grammaticalisation frequently involves reanalysis, semantic change and/or phonological erosion, although none of these characteristics is sufficient or necessary. (More on this below.) There is no phonological erosion in two of the three cases discussed here, *sui* verb 'end, finish; be finished', sequencer, completive marker, exhaustive marker, contrast/unexpectedness coordinator, and *bia*/*bii* comitative preposition and noun phrase and clause conjunctive coordinator. In the case of the inceptive marker *thafali-qi*, what was historically a detransitivising suffix is optionally omitted, but the same omission is found with the historically related associative suffix used on head nouns in associative noun phrases (example [38] in §4), where there is no grammaticalisation of the head noun.

There is a semantic/functional split, divergence in the case of *sui* verb 'end, finish; be finished', vs sequencer, vs completive marker, vs exhaustive marker vs contrast/ unexpectedness coordinator, and in the case of *bia*/*bii* comitative preposition vs noun phrase and clause conjunctive coordinator. On the other hand, there is no semantic change between the transitive verb 'start, begin' when it takes a clausal complement and the

---

[6] Ulrich Detges has suggested to me (pers. comm., 9 November 2007) that as a (noun phrase) coordinator *bia*/*bii* may have had an emphatic flavour that the older form *ma* lacked. However, in the absence of historical or comparative evidence it is impossible to evaluate this possibility.

inceptive marker, as also evidenced by the language consultant's reaction to the sentence in (40) in §4; cf. (41).

It is likely that the development of the noun-phrase coordinating function of *bia*/*bii* from the comitative function did involve reanalysis due to structural ambiguity; see (49) in §5. The later use of *bia*/*bii* in clausal coordination involved an extension from the noun-phrase coordinating function. In the case of the verb 'start, begin' there was no structural ambiguity that would give rise to the development of the inceptive function. There has, nevertheless, been significant restructuring: a new monoclausal structure (with the inceptive particle) vs the older biclausal structure (with the verb 'start, begin' in the higher clause). In the case of *sui* too there has been restructuring, for example, *sui* 'end, finish; be finished' as the head of a predicate vs *sui* as a postverbal particle. However, in the absence of historical evidence it is not possible to tell whether structural ambiguity was a factor.

For many students of grammaticalisation, this is a distinct type of language change that is characterised by unidirectionality: the development is always or almost always from a lexical element to a grammatical element or from a less grammatical element to a more grammatical element (for example, Heine et al. 1991; Hopper and Traugott 1993; Haspelmath 1999). There are no exceptions (perhaps by definition) or the exceptions are exceedingly rare. However, there have been dissenting voices, arguing that the processes claimed to be characteristic of grammaticalisation are not unique to this kind of change and that grammaticalisation is an epiphenomenon of general types of change: reanalysis, semantic change, phonological change; see, for example, Campbell (2001), Newmeyer (2001).

If the latter view is correct, and I believe it is, does that mean that the concept of grammaticalisation is superfluous? Not so. It designates grammatical changes that have a certain kind of outcome: development of a grammatical element from a lexical element or from another (less?) grammatical element. (Similarly, assimilation is a useful concept that designates a certain type of sound change.) A grammaticalisation change may, but need not, involve reanalysis, and/or semantic/functional change, and/or phonological change (such as erosion). On this view, grammaticalisation is unidirectional by definition. Unidirectionality is, then, not an empirical matter, but a definitional property. (Likewise, assimilation is defined as a change through which sounds become more alike, but that does not make the concept of assimilation useless.) What is interesting is that grammaticalisation is considerably more common than the opposite kind of process, degrammaticalisation (from grammatical to lexical, from more grammatical to less grammatical), even though this may be the consequence of the way reanalysis, semantic/functional change and phonological change operate in general. (On the whole, assimilation is considerably more common than dissimilation.)

The fact that grammaticalisation is more common (more natural?) than degrammaticalisation can also serve as a heuristic device in reconstruction, although not as a proof. And the notion of grammaticalisation also helps to highlight the fact that many (most?), though not necessarily all, grammatical elements arise ultimately from lexical sources, which may help us better to understand at least some of their properties.

References

Campbell, Lyle. 2001. What's wrong with grammaticalization? *Language Sciences* 23:113–161.

Capell, A. 1930. The language of Inakona, Guadalcanar, Solomon Islands. *Journal of the Polynesian Society* 39:113–136.

—— 1971. *Arosi grammar*. Canberra: Pacific Linguistics.

Craig, Colette G. 1991. Ways to go in Rama: a case study in polygrammaticalization. In Elizabeth Closs Traugott and Bernd Heine, eds *Approaches to grammaticalization*, vol. II: *Focus on types of grammatical markers*, 455–492. Amsterdam/Philadelphia: John Benjamins.

Cristofaro, Sonia. 1998. Grammaticalization and clause linking strategies: a typological approach with particular reference to Ancient Greek. In Anna Giacalone Ramat and Paul J. Hopper, eds *The limits of grammaticalization*, 59–88. Amsterdam/Philadelphia: John Benjamins.

—— 2003. *Subordination*. Oxford: Oxford University Press.

Crowley, Susan Smith. 1986. *Tolo dictionary*. Canberra: Pacific Linguistics.

Deck, Norman C. 1934. *Grammar of the language spoken by the Kwara'ae people of Mala, British Solomon Islands*. Reprint No.5. Wellington: Polynesian Society.

Fox, Charles E. 1974. *Lau dictionary*. Canberra: Pacific Linguistics.

—— 1978. *Arosi dictionary* (revised edition with English–Arosi index prepared by Mary Craft). Canberra: Pacific Linguistics.

Geertz, P. 1970. *'Āre'āre dictionary*. Canberra: Pacific Linguistics.

Haspelmath, Martin. 1999. Why is grammaticalization irreversible? *Linguistics* 37–6:1043–1068.

Heine, Bernd and Mechthild Reh. 1984. *Grammaticalization and reanalysis in African languages*. Hamburg: Helmut Buske.

Heine, Bernd, Ulrike Claudi and Friederike Hünnemeyer. 1991. *Grammaticalization: a conceptual framework*. Chicago and London: Chicago University Press.

Hopper, Paul J. 1991. On some principles of grammaticization. In Elizabeth Closs Traugott and Bernd Heine, eds *Approaches to grammaticalization*, vol. I: *Focus on theoretical and methodological issues*, 17–35. Amsterdam/Philadelphia: John Benjamins.

Hopper, Paul J. and Elizabeth Closs Traugott. 1993. *Grammaticalization*. Cambridge: Cambridge University Press.

Ivens, Walter. G. 1929. *A dictionary of the language of Sa'a (Mala) and Ulawa, South-east Solomon Islands*. London: Oxford University Press; and Melbourne: Melbourne University Press.

—— 1940. *A dictionary of the language of Bugotu, Santa Isabel Island, Solomon Islands*. London: Royal Asiatic Society.

Keesing, Roger M. 1975. *Kwaio dictionary*. Canberra: Pacific Linguistics.

—— 1985. *Kwaio grammar*. Canberra: Pacific Linguistics.

Langacker, Ronald W. 1991. *Foundations of cognitive grammar*, vol. II, *Descriptive application.* Stanford: Stanford University Press.

Lichtenberk, Frantisek. 1991a. Semantic change and heterosemy. *Language* 67:475–509.

—— 1991b. On the gradualness of grammaticalization. In Elizabeth Closs Traugott and Bernd Heine, eds *Approaches to grammaticalization*, vol. I: *Focus on theoretical and methodological issues*, 37–80. Amsterdam/Philadelphia: John Benjamins.

—— 2006. Associative and possessive constructions in Oceanic: the links and the differences. In Henry Y. Chang, Lillian M. Huang and Dah-an Ho, eds *Streams converging into an ocean: Festschrift in honor of Professor Paul Jen-Kuei Li on his 70th birthday*, 19–47. Taipei: Academia Sinica.

—— 2008. Asymmetry reversal. In Barbara Lewandowska-Tomaszczyk, ed. *Asymmetric events*, 173–193. Amsterdam/Philadelphia: John Benjamins.

Moreno Cabrera, Juan C. 1998. On the relationships between grammaticalization and lexicalization. In Anna Giacalone Ramat and Paul J. Hopper, eds *The limits of grammaticalization*, 211–227. Amsterdam/Philadelphia: John Benjamins.

Newmeyer, Frederick J. 2001. *Deconstructing grammaticalization*. Language Sciences 23:187–229.

17 *On the Eastern Polynesian 'direct object' marker* i

YUKO OTSUKA

1 Introduction[1]

One of the long-standing debates in the Polynesian linguistic literature is whether the case system in Proto Polynesian (PPN) was ergative or accusative. Synchronically, we find two types of case marking within Polynesian. Eastern Polynesian (EP) languages show an accusative pattern, as exemplified in (1). (Rapanui is an exception to this generalisation. See §4 below.) The subject of a transitive verb is unmarked as is the subject of an intransitive verb. Instead, the patient argument of a transitive verb is marked with a particle *i*. This particle has been called the direct object (or accusative) marker in the literature. EP languages are also said to have passive constructions, in which the verb has a suffix *-Cia* (C representing a variable thematic consonant) and the agent is marked with a particle *e*.

(1) **Hawaiian**[2]

a. *Ua hele au i nehinei.* <intransitive>
PFV go 1SG at yesterday
'I went yesterday.' (Hopkins 1992:56)

[1] It is my great pleasure to dedicate this paper to Bob Blust, an outstanding and inspiring colleague. Working with Bob in the same department motivated me to study Polynesian syntax from a historical comparative perspective. I would also like to thank Andrew Pawley and Albert Schütz for their valuable comments on early drafts of this paper.

[2] Key to abbreviations: ABS – absolutive, ACC – accusative, AGT – agent, ART – article, ASP – aspect marker, COMP – complementiser, DIST – distal, DU – dual, DEF – definite, EMPH – emphatic, ERG – ergative, EXP – experiencer, FOC – focus marker, IPFV – imperfective aspect, INDF – indefinite, ML – mainline aspect, NFUT – non-future tense, NPST – non-past tense, NS – nonspecific aspect, OBJ – object, OBL – oblique, PASS – passive, PAT – patient, PERS – personal, PFV – perfective aspect, PL – plural, POSS – possessive, PP – preposition, PRED – predicative, PRF – perfect aspect, PROG – progressive aspect, PRON – pronominal, PRS – present tense, PST – past tense, REF – referential, SG – singular, TAM – tense aspect mood marker, TOP – topic, UNS – unspecified tense. Interlinear glosses are mine for the examples cited from Biggs (1998), Coppenrath and Prevost (1975), Harlow (2001), Hawkins (1982), Hopkins (1992), and Hovdhaugen et al. (1989); otherwise, as given in the original sources.

Alexander Adelaar and Andrew Pawley, eds
Austronesian historical linguistics and culture history: a festschrift for Robert Blust, 247–265.
Pacific Linguistics, 2009.
© This edition vested with Pacific Linguistics.

b. *Ua ʻai ke kanaka* ***i*** *ka poi.* <transitive>
PFV eat DEF man OBJ DEF poi
'The man ate the poi.' (Elbert and Pukui 1979:39)

c. *Ua hoʻihoʻi ʻia ke kaʻa* ***e*** *ia.* <passive>
PFV return *Cia* DEF car AGT 3SG
'The car was returned by him.' (Hopkins 1992:157)

In contrast, we find an ergative pattern in the rest of Polynesian, in other words Tongic and Nuclear Polynesian other than EP. The patient argument of a transitive verb is unmarked, as is the subject of an intransitive verb. On the other hand, the agent argument of a transitive verb is marked with the particle *e*. In addition to the ergative construction, there is a third pattern for certain verbs, in which the object is marked with the preposition *i* or *ki*. Whether the object is unmarked or marked with a preposition is lexically determined. Verbs taking *i*/*ki*-marked objects are called middle verbs, which are typically verbs of perception and emotion (Chung 1978).

(2) **Tongan**

a. *Kuo ʻalu ʻa e tamasiʻí.* <intransitive>
PRF go ABS REF boy.DEF
'The boy has gone.'

b. *Kuo inu* ***ʻe*** *he tamasiʻí ʻa e vai.* <transitive>
PRF drink ERG REF boy.DEF ABS REF water
'The boy has drunk the water.'

c. *Naʻe sio ʻa e tamasiʻí* ***ki*** *he ʻakau.* <middle>
PST see ABS REF boy.DEF to REF tree
'The boy saw a tree.'

Altogether, four types of dyadic verbal constructions are attested in Polynesian, as schematised in (3). The order in which noun phrases (NPs) occur is relatively flexible in ergative as well as passive constructions.

(3) **Four verbal constructions in Polynesian**

ERGATIVE	V	*e*	AGT		PAT
PASSIVE	V-*Cia*	*e*	AGT		PAT
MIDDLE	V		EXP/AGT	*i/ki*	PAT
ACCUSATIVE	V		AGT	*i*	PAT

One immediately notices the striking similarity between the ergative and passive patterns. These two constructions are almost identical except for the affixation on the verb. Equally remarkable is the resemblance between the accusative and middle patterns. From a historical point of view, one would expect this curious state of affairs to have arisen due to some syntactic change, be it a drift or an innovation. The question is what the relevant change was and when it occurred. Based on the formal resemblance between the ergative and passive constructions and their complementary distribution, linguists have reached two conflicting hypotheses. Some have argued that PPN was accusative and that the ergative pattern is a consequence of a syntactic drift from, or reanalysis of, the PPN passive constructions (Hale 1968; Hohepa 1969; Chung 1978). Others have proposed that PPN was ergative and that it was the expansion of the middle pattern that gave rise to the accusative

pattern (Clark 1973, 1976; Gibson and Starosta 1990; Pawley 2001). The issue has been debated fervently, but without a conclusive consensus.

This paper examines the objects of middle verbs in Polynesian and so-called direct objects in EP to provide evidence in support of the position that PPN was ergative. I propose that the relevant innovation in PEP was the extension of the middle pattern to all dyadic verbs, with *i becoming an indicator of less affected patients. I argue that this PEP innovation should be understood not as the loss of ergative constructions, but as the development of a productive two-way system, reminiscent of the voice/focus system of western Austronesian.

2 The debate: Was PPn ergative or accusative?

The history of the debate in question goes back to the 1920s when two opposing views were proposed regarding the original function of PPN *-Cia (Pawley 2001). Herbert Williams (1928) argued that the main role of *-Cia was to mark passive voice, as in Māori. On the contrary, Spencer Churchward (1928) claimed that the original function of PPN *-Cia was to mark transitivity, as demonstrated in Samoan. This was just a precursor to the more extensive debate that was ignited in the late 1960s.

The passive-to-ergative hypothesis was first proposed by Ken Hale (1968). In this approach, case marking in PPN is taken to have been accusative. Hohepa (1969) followed suit and proposed that the ergative pattern is a result of a syntactic drift, in other words preference for passive over active constructions. One argument for the passive-to-ergative hypothesis is the high frequency of passive in languages like Māori. It is argued that in ergative languages such as Tongan and Samoan, this preference became an obligatory rule. This led to a situation in which all transitive constructions occur only in their passive forms. Having lost its original function to distinguish passive from active forms, *-Cia became optional, yielding a new, ergative pattern (Hohepa 1969). Chung (1978) also postulated the obligatory passive rule and argued that the passive pattern was then reanalysed as the canonical transitive in the absence of the contrasting active pattern.

The opponents of this position base their argument on the distribution of the two types of case marking within Polynesian. EP languages constitute a lower-order subgroup, branching off from Proto Ellicean (Marck 2000). If the ergative pattern arose due to a drift, as suggested by Hohepa (1969), one must assume parallel development in Proto Tongic as well as all other languages belonging in the Nuclear Polynesian subgroup except for EP (Clark 1973, 1976). Given Marck's (2000) subgrouping, this means that we must assume parallel development in Proto Tongic, multiple Outlier languages that directly branch off from Proto Nuclear Polynesian, Samoan, and Ellicean outlier languages. Thus, Clark argues that 'the principal change in the system of transitive case-marking within PN is the expansion of [the middle pattern] in EP' and that the rest of Polynesian retained the PPN ergative system (1976:81). Pawley (1980) supports Clark's proposal by emphasising how extremely unlikely it is that many different languages would independently develop systems that matched in such detail.

In support of Clark's position, Pawley (2001) has shown that PPN *-Cia was not a passive suffix, but that the source of *-i* in *-Cia* is the Proto Oceanic short transitive suffix *-i (Blust 1986; Pawley 1973) and the source of *-a* is a Proto Eastern Oceanic suffix *-a, which derived stative verbs from transitives. Pawley's (2001) analysis of *-Cia* explains why this suffix has been treated as the passive suffix in EP. Statives derived from transitives focus on the resulting state of previous events. Such derived stative verbs imply

the presence of an agent, but the relevant NP is demoted as a result of detransitivisation. The grammatical and semantic effects of the stativising suffix *-a, therefore, are similar to those of passive morphemes in languages like English.

Thus, from a historical perspective, it is more plausible to hypothesise that case marking in PPN was ergative. Note that the passive-to-ergative approach does not have much to say about the position of middle constructions, namely, how a subset of the proposed canonical transitive remained as such in ergative languages such as Tongan and Samoan. In the subsequent sections, I provide some evidence to support Clark's (1973, 1976) proposal that the development of the accusative pattern in (3) is an innovation in PEP.

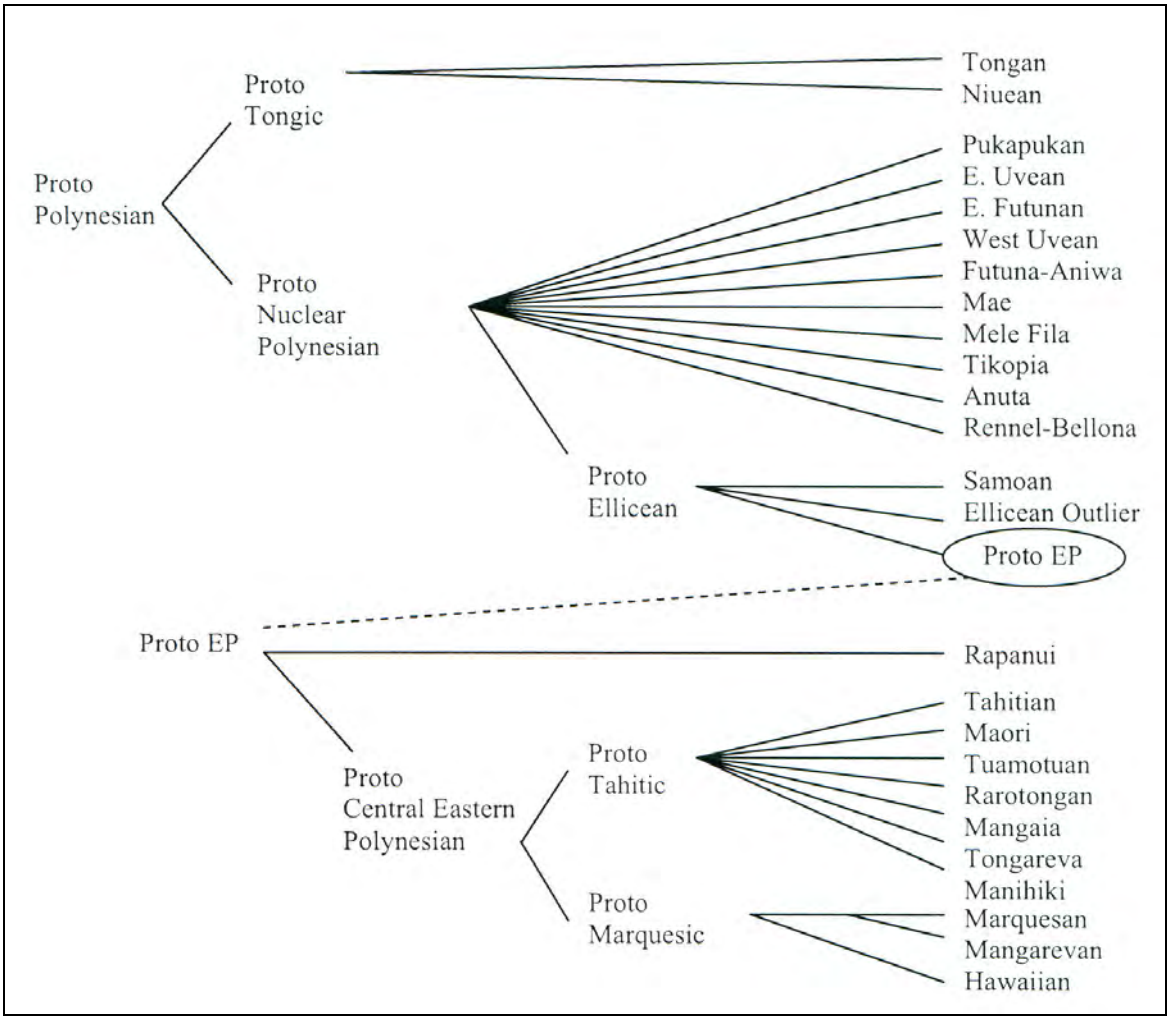


Figure 1: Marck's (2000) subgrouping of Polynesian languages

3 Middle verbs and *-Cia*

Leaving aside the 'direct' objects in EP, objects marked with *i/ki* are commonly found across Polynesian.

(4) **Tongan middle**

a. *'oku sai'ia 'a e tamasi'í* ***'i*** *he ika.*
PRS like ABS REF boy.DEF in REF fish
'The boy likes fish.'

b. *Kuo fanongo 'a e tamasi'í* ***ki*** *he ongoongó.*
PRF listen ABS REF boy.DEF to REF news.DEF
'The boy has heard the news.'

(5) **Samoan middle**

a. *'ua tago le teine* ***'i*** *lona ulu.*
PRF touch the girl to POSS.3SG head
'The girl touched her head.' (Chung 1978:56)

b. *'ua 'alo tagata* ***i-āte*** *a'u*
PRF dodge person.PL at-PRON 1SG
'People ignored me.' (Chung 1978:56)

Middle constructions also exist in EP languages with accusative case marking. In some languages, such as Hawaiian and Tahitian, it is hard to distinguish middle objects from objects of canonical transitive constructions, for the distinction between *ki and *i has been lost due to a regular sound change (i.e., PPN *k > ' possibly followed by the loss of the initial glottal stop in the reflex of *ki). However, *ki*-marked middle objects can be found in Māori and Rapanui.

(6) **Māori**

a. *E mōhio ana au* ***kia*** *ia.*
NPST know IPFV 1SG to.PERS 3SG
'I know him.' (Biggs 1998:113)

b. *Ka pīrangi ia* ***ki*** *ngā mea katoa.*
TAM want 3SG to the.PL thing all
'He wants all the things.' (Bauer 1997:197)

(7) **Rapanui** (Weber 1988:36)

a. *e haŋa era a ia* ***ki*** *te uka ...*
IPFV desire then PERS 3SG to the girl
'When he desired a girl …' (Weber 1988:36)

b. *he piri a Taparahi* ***ki*** *te tahi ŋa poki*
ML join PERS Taparahi to the other PL child
'Taparahi joined other children' (Weber 1988:36)

The choice of *i* or *ki* in middle constructions is generally lexically determined, although some verbs may take either *i* or *ki* with slight semantic differences (Bauer 1997; Biggs 1998; Clark 1973, 1976; Chung 1978). For example, in Tongan, most middle objects are marked with *ki*, with a few verbs such as *sai'ia* 'to like' requiring *i*-marked objects. At the same time, some verbs like *fanongo* 'to listen', *manako* 'to want', and *muimui* 'to follow' permit both *ki*- and *i*-marked objects. In any case, whether a verb can occur in a middle construction is lexically determined. Across Polynesian, middle constructions typically involve verbs of perception and emotion, but some other dyadic verbs also belong to this class, e.g., 'to wait', 'to follow', 'to visit', 'to speak', 'to arrive', and 'to call'.

PPN had a transitive suffix *-Ci (Clark 1973, 1976; Pawley 2001), which was used to derive transitive verbs from middle verbs. The derived transitive verbs imply that the patient is directly (or more) affected. This is attested in languages like Tongan and East Futunan, in

which (a) some middle verbs have corresponding ergative verbs ending in -*ʻi* and (b) ergative verbs imply that the patient is directly (or more) affected (Biggs 1974, Moyse-Faurie 1992).

(8) **East Futunan**

a. *e tusi a Kalala ki le toe.*
NS point.to ABS Kalala OBL DEF.SG child
'Kalala points out the child.' (Moyse-Faurie 1992:220)

b. *e tusi-**ʻi** le toe e Kalala.*
NS point.to-CI DEF.SG child ERG Kalala
'Kalala designates the child.' (Moyse-Faurie 1992:220)

(9) **Tongan**

a. *ʻoku manatu ʻa e kakaí ki he kuohilí.*
PRS remember ABS REF people.DEF to REF past.DEF
'The people remember the past.'

b. *ʻoku manatu-**ʻi** ʻa e lesoní ʻe he tamaikí.*
PRS remember-CI ABS REF lesson.DEF ERG REF children.DEF
'The children remember the lesson.'

Thus, we may assume that PPN had two classes of dyadic verbs—middle and ergative. According to the passive-to-ergative approach, the unmarked objects of ergative constructions were originally the subjects of PPN passive constructions. This hypothesis makes two incorrect predictions. First, given that some NPs such as locatives and middle objects are synchronically *i*-marked in ergative languages such as Tongan and Samoan, the theory must assume that the obligatory passive rule did not apply to these NPs. Otherwise, they would have been reanalysed as unmarked objects and therefore, should occur in the ergative pattern. Second, the theory must interpret the fact that *ki*-marked objects coexist with ergative constructions in many languages to indicate that the passive rule did not apply to middle constructions with *ki-marked objects.

Contrary to the first prediction, passivisation is available for not only *i*-marked middle objects, but also *i*-marked locative NPs in accusative language such as Māori, Hawaiian, and Tahitian.[3]

(10) **Māori**

a. *Kua kite-a e rātou he tikanga.* <middle>
PRF see-PASS AGT 3PL INDF plan
'A plan has been found by them' (Harlow 2001:187)

b. *i te ākau i kau-ria e te taniwha nei* <locative>
at the coast TAM swim-PASS by the taniwha near.I
'in the coastal waters which are swum by the taniwha' (Bauer 1997:481)

[3] As for Hawaiian and Tahitian, I regard these *i*-marked objects as middle objects rather than direct objects based on the meaning of the verb.

(11) **Hawaiian**

a. *Ua 'ike paha 'ia ka 'aihue e nā hoa noho?* <middle >
PFV see maybe PASS the thief AGT PL neighbor
'Perhaps the thief was seen by the neighbors?' (Hopkins 1992:154)

b. *Ua hele 'ia ke alanui (e lākou)* <locative>
PFV go PASS the street AGT 3PL
'The street was traveled (by them).' (Hawkins 1982:30)

(12) **Tahitian**[4, 5]

a. *'ua 'ite-a te tāvini iā na.* <middle>
PFV see-PASS the servant in.PRON 3SG
'The servant was found by him.' (Coppenrath and Prevost 1975:232)

b. *terā te 'ē'a i haere-hia e rātou.* <locative>
that ART path ASP go-PASS PP 3PL
'That is the path they took (lit. was gone by them).'
(Lazard and Peltzer 2000:67)

Likewise, we find no evidence to support the second prediction, i.e., that passivisation was unavailable for middle constructions with *ki-marked objects in PPN. Rather, we find counterexamples in ergative as well as accusative languages. For example, forms like *a'usia* 'to reach, achieve' and *siofia* 'to look deep into' in Tongan provide evidence that *-Cia* could be affixed to middle verbs taking *ki*-marked objects such as *a'u* (*ki*) 'to arrive' and *sio* (*ki*) 'to see'. Similarly, Māori permits *ki*-marked objects to undergo passivisation, as shown below. This also argues against the passive-to-ergative hypothesis, for it cannot explain why such sentences have not been reanalysed as ergative.

(13) **Māori**

a. *Ka pīrangi-tia e ia ngā mea katoa.*
TAM want-PASS by 3SG the.PL thing all
'All the things are wanted by him.' (Bauer 1997:197)

b. *Ka tāri-a te pahi e Mere.*
UNS wait-PASS the bus AGT Mary
'The bus was waited for by Mary.' (Biggs 1998:28)

Incidentally, these examples suggest that *-Cia in PPN was not a 'passive' suffix in the syntactic sense (cf. Pawley 2001). Rather, it was added to any verb—transitive, middle, or intransitive—to indicate the affectedness of the patient or the locative. In Māori, intransitive verbs can be affixed with *-Cia* only if the locative NP is taken to be directly affected (Chung 1978). Lazard and Peltzer (2000:67) have proposed that *-Cia* suffixation in Tahitian should be understood as 'pseudo passive', whose function is to indicate that the event is accomplished (or will be accomplished), that the agent is volitional, and that the patient (if there is one) is affected.

---

[4] The interlinear glosses and free translations of the Tahitian examples are translated by the author from the original French.

[5] In the examples provided by Coppenrath and Prevost (1975), the underlying subject is marked by *i* rather than agentive *e* in the 'passive' form of middle and intransitive verbs. Lazard and Peltzer (2000) give examples in which the agent is marked with *e*.

Similarly, in Tongan, although *-Cia* occurs only in some fossilised forms, the root of such *-Cia* verbs can be transitive, middle, or intransitive, including stative: *tanumia* 'to be buried' and *inumia* 'to consume by evaporation or absorption (as of the sun and the ground)' from transitive *tanu* 'to bury' and *inu* 'to drink'; *'ofeina* 'to love' and *a'usia* 'to reach' from middle *'ofa* (*'i*) 'to love' and *a'u* (*ki*) 'to arrive'; *kaiha'asia* 'to rob' and *nofo'ia* 'to be constantly occupied' from intransitive *kaiha'a* 'to steal' and *nofo* 'to stay'; and *ifo'ia* 'to find (something) tasty' and *sai'ia* 'to like' from stative *ifo* 'tasty' and *sai* 'good'. Note also that the derived verbs are not necessarily semantically passive. Furthermore, while some of the *-Cia* verbs in Tongan may co-occur with an agent marked with *'e*, the majority of them do not allow human agents (Chung 1978; Pawley 2001). These facts argue against the passive-to-ergative hypothesis. The ergative pattern with human agents could not have arisen from the 'passive' pattern with verbs affixed with *-Cia*.

Finally, if we hypothesise that 'passivised' middle constructions had indeed been reanalysed as ergative, we would then have to assume an additional innovation that followed the passive-to-ergative reanalysis in ergative languages—namely, the use of *ki to mark less affected patients. Such a scenario is undesirable for two reasons. First, it forces us to postulate parallel development in multiple languages. This is problematic, as pointed out by Clark (1976) with regard to the passive-to-ergative hypothesis. Second, such a hypothesis must take middle constructions in ergative languages to have been derived from their ergative counterparts. However, such an analysis is incompatible with the fact that some ergative forms are clearly derived from middle verbs, as seen above in the examples from Tongan and East Futunan.

4 PEP *i as a marker of less affected patients

Clark (1973, 1976) reconstructed two 'transitive' constructions for PPN, as given in (14). Here, I use the terms such as 'transitive,' 'subject,' and 'object' as used by Clark, although these terms may not be the best suited to describe Polynesian languages (cf. Biggs 1974). Pattern 2 has two variants, one with the suffix *-Cia* and the other without.

(14) Clark's (1973:569) reconstruction of PPN transitive constructions

Pattern 1	V			SUB	*i/*ki	OBJ
Pattern 2	V	(*-Cia)	*e*	SUB		OBJ

Clark also postulated two classes of transitive verbs in Polynesian. There is a group of verbs that are confined to Pattern 2 ('A-verbs'). The rest form the other group ('B-verbs'), which mainly occur in Pattern 1, but may also occur in Pattern 2. In ergative languages, A-verbs are canonical transitive, and B-verbs, middle. B-verbs are suffixed with *-Ci when occurring in Pattern 2. In EP, we find only the suffixed variant of Pattern 2, and there is no clear distinction between A-verbs and B-verbs, that is, both types of verbs may freely occur in either pattern. Clark proposed that PEP extended Pattern 1 to A-verbs, which had been confined to Pattern 2 earlier, thereby giving rise to the accusative pattern. Clark remained agnostic about the exact nature of the relation between Patterns 1 and 2 in B-verbs, although he stated that it is 'clearly not a simple matter of active versus passive' (1973:574). Note also that Clark treated *-Ci as a variant of *-Cia. He was uncertain about the (semantic) factors determining the use of these suffixes.

My analysis is basically the same as Clark's, but also incorporates Biggs's (1974) insight on the verb classes in Polynesian. Biggs argued that transitive verbs in Polynesian

divide into two classes, active and agentive. Active verbs take actor subjects, whereas agentive verbs take 'goal' (patient) subjects, where 'subject' is defined as the unmarked and indispensable NP in the sentence. These verbs may also take an additional NP: *i*-marked 'direct comment,' *ki*-marked 'indirect comment,' or *e*-marked agent.[6] In this approach, *-Cia* is understood as a suffix that changes an actor verb into an agentive verb.

(15) Biggs' (1974) classification of Polynesian (dyadic) verbs

			'SUBJECT'		
Active verbs	V		ACTOR	*i*	DIRECT COMMENT
Active verbs	V		ACTOR	*ki*	INDIRECT COMMENT
Agentive verbs	V	(*-Cia*)	GOAL	*e*	AGT

Although Biggs spoke of active versus agentive verbs, it would be more appropriate to talk of active and agentive constructions, for some verbs are ambiguous between active and agentive, as shown in (16).[7]

(16) a. **Tongan** (Tchekhoff 1973:283)
'oku ui 'a Mele.
PRS call ABS Mary
'Mary calls.' <active >
'Mary was called.' <agentive>

b. **Tokelauan** (Sharples 1976:315)
na kai te magoo.
PST eat the shark
'The shark ate.' <active>
'The shark was eaten.' <agentive>

c. **Rapanui** (Alexander 1981:142)
He kai te ika.
PRS eat the fish
'The fish eats.' <active >
'The fish was eaten.' <agentive>

Based on the preceding discussion, I propose the following scenario as to how PPN *i came to function as the direct object marker in EP.

First, PPN had three types of constructions: stative, active, and agentive (see (17) below). Each of these constructions was used either monadically, occurring only with a single, unmarked NP, or dyadically, taking an extra NP. This additional NP is a cause NP marked with *i for stative constructions, a goal/patient NP marked with *ki for active constructions, and an agent NP marked with *e for agentive constructions. I hypothesise

---

6 Tchekhoff (1973) proposed a similar approach for Tongan. The ABS marker *'a* is taken to be the marker for the 'first modifier' of the sentence, to which the 'second modifier' can be introduced either by *ki*/*'i* (goal/patient) or by *e* (agent). In Tchekhoff's analysis, the interpretation of the first modifier is determined by the nature of the second modifier: if it is *ki*/*'i*-marked, the first modifier is interpreted as the agent; if *e*-marked, it is interpreted as the patient.

7 Hovdhaugen et al. (1989) use 'agentive' and 'non-agentive' constructions to refer to this difference in Tokelauan.

that the marking of goal NPs in active constructions was *ki, based on the facts that (a) *ki*-marked objects exist across the Polynesian family, including EP languages such as Māori and Rapanui; and (b) that middle objects are overwhelmingly, if not always, marked with *ki rather than *i in languages like Tongan and Tokelauan.

Second, there were three types of verbs in PPN: stative, middle, and ergative. Stative verbs can only occur in the stative pattern. Middle verbs are confined to the active pattern and involve goals/patients that are not directly affected. Ergative verbs mainly occur in the agentive pattern, but may also occur in the active pattern when used monadically (as in the examples in (16) above).

Third, there were two separate derivational suffixes *-Ci and *-Cia. I propose that the former, a reflex of the Proto Oceanic 'short' transitive suffix *-i (Blust 1986; Pawley 1973, 2001), was used in PPN to change middle (active) verbs into ergative (agentive) verbs to indicate that the patient is directly (or more) affected. The other suffix *-Cia appears to be much more flexible in selection of bases to which it attached. It can attach to nouns to create stative verbs (e.g., Samoan *lago* 'fly' > *lago-ia* 'covered with flies'), stative verbs to create active verbs (e.g., Tongan *sai* 'tasty' > *sai-'ia* 'to like'), active verbs to create agentive verbs (e.g., Tokelauan *alofa* 'to love' > *alofa-gia* 'to love'). What these diverse instances of *-Cia* have in common is a semantic effect, namely, to indicate that the patient is directly affected. Thus, I argue that the function of *-Cia was primarily semantic, and that unlike *-Ci, it was not associated with any particular construction type.

I therefore reconstruct four sentence types for PPN, as schematised in (17).

(17) Reconstructon of PPN sentence patterns

Stative	V	PAT	(*i	CAUSE)	
Active	V	AGT/EXP	(*ki	GOAL/PAT)	[–AFFECTED]
Agentive	V	PAT/GOAL	(*e	AGT)	[+AFFECTED]
Derived Agentive	V-*Ci/(*Cia)	PAT/GOAL	(*e	AGT/EXP)	[+AFFECTED]

In this paradigm, superficially, directly affected patients were unmarked, whereas less affected patients were marked with *ki. The contrast was actually intended in the case of derived agentive versus active. I propose that this rule was extended in PEP to mark all patients that were less (or not directly) affected. In effect, the rule applied to agentive constructions that did not have corresponding active constructions, when a need arose to indicate that the patient was not directly affected.[8] This rule gave rise to the 'accusative' pattern. It appears that *-Ci as a suffix deriving agentive verbs from active verbs was replaced by the other suffix, *-Cia, also at this stage.[9]

[8] Gibson and Starosta (1990) have proposed a similar analysis for Māori. They claimed that Māori should be regarded as an ergative language, showing the so called 'passive' pattern to be the canonical transitive and the so called 'accusative' pattern to be an instance of middle constructions. They suggested that 'the major syntactic innovation in Māori is that the canonical transitive verbs developed a "middle" pattern' and that this can be viewed as 'the addition of a detransitivizing rule to mark less-affected objects' (Gibson and Starosta 1990:206)

[9] Clark (1973:589) proposes that the distinction between *-C*i* and *-C*ia* was lost in Proto Nuclear Polynesian.

(18) Reconstruction of early PEP sentence patterns

Stative	V	PAT	(*i	CAUSE)	
Active	V	AGT/EXP	(*ki	GOAL/PAT)	[–AFFECTED]
Derived Active	V	AGT/EXP	(*i	GOAL/PAT)	[–AFFECTED]
Agentive	V	PAT/GOAL	(*e	AGT)	[+AFFECTED]
Derived Agentive	V-***Cia**	PAT/GOAL	(*e	AGT/EXP)	[+AFFECTED]

As to why the preposition used for this purpose was *i rather than *ki, Clark (1973:600) suggests that it was modeled on the 'partitive' construction, reconstructable in PPN, in which a canonical transitive verb takes an oblique case nominal introduced by a reflex of *i, denoting an action which partially or incompletely affects the object. Drawing from external witnesses, namely the Fijian languages, Pawley (1980) suggests that the range of uses of *i was expanded in the following steps. Standard Fijian *e* (normally spoken as *i*) and Western Fijian *i* mark a range of oblique case roles: (a) locative 'at', (b) source 'from', (c) cause/instrument with inanimates. PPN *i additionally marked partitives and direct objects of middle verbs. Finally, in PEP, the preposition *i acquired another function to mark indirectly affected patients.

Once this rule had become productive, another generalisation was drawn from the paradigm (18): suffixed verbs co-occur with directly affected patients, whereas less affected patients co-occur with verbs without a suffix. Due to this generalisation, at some later stage, the use of *-Cia became obligatory for all verbs occurring in the agentive pattern. In other words, the suffix *-Cia acquired a syntactic function in addition to its original semantic function. The resulting paradigm (19) is what we find in the majority of EP languages today.

(19) Reconstruction of late PEP sentence patterns

Stative	V	PAT	(*i	CAUSE)	
Active	V	AGT/EXP	(*ki	GOAL/PAT)	[–AFFECTED]
Derived Active	V	AGT/EXP	(*i	GOAL/PAT)	[–AFFECTED]
Agentive	V-***Cia**	PAT/GOAL	(*e	AGT)	[+AFFECTED]

Note that *-Cia* is not productive in Rapanui, where we find the agentive and derived active patterns instead of the paradigm in (19) (Alexander 1981, Weber 1988, Du Feu 1996). This situation may have resulted if only the first rule (i.e., marking less affected patients with *i) came into effect without the second one (i.e., marking affected patients with *-Cia). Thus, the development of the paradigm in (19) may have occurred in Proto Central Eastern Polynesian rather than PEP.

(20) **Rapanui** (Alexander 1981:132)

a. *he tiŋa'i te vi'e **i** te taŋata.* <derived active>
PRS hit the woman the OBJ man
'The woman beats the man.'

b. *he tiŋa'i te taŋata **e** te vi'e* <agentive>
PRS hit the man AGT the woman
'The man is beaten by the woman.'

A similar pattern is found in Pukapukan. Unlike Rapahui, however, ergative verbs may or may not be suffixed by *-Cia* in the agentive pattern, as shown in (21).

(21) **Pukapukan** (Chung 1978:323–324)

a. *Na patu mātou **i** te tamaiti.* <derived active>
PST hit 1PL ACC the child
'We hit the child.'

b. *Na patu te tamaiti **e** mātou.* <agentive without -*Cia*>
PST hit the child ERG 1PL
'We hit the child.'

c. *Na patu-**a** te tamaiti **e** mātou.* <agentive with -*Cia*>
PST hit-PASS the child AGT 1PL
'The child was hit by us.'

Middle verbs must be suffixed with -*Cia* when occurring in the agentive pattern, as illustrated in (22). This is expected in the present approach. In any of the paradigms proposed above, middle verbs may occur only in the active or derived agentive patterns, but crucially, not in the (inherent) agentive pattern.[10]

(22) **Pukapukan** (Chung 1978:323–324)

a. *Ko mina i a-na **i** te ika* <derived active>
PROG want NOM PRON-3SG ACC the fish
'He wants the fish.'

b. *Ko mina-**ngia** te yua **e**-ku.* <derived agentive>
PROG want-*Cia* the water AGT-1SG
'I want the water.'

c. **Ko mina **e**-ku te yua.* <agentive without -*Cia*>
PROG want ERG-1SG the water
'I want the water.'

These data support the present proposal that the rise of 'accusative' constructions was a two-stage process involving two innovations, one leading to the other.

5 *I*-marked objects as a subset of middle objects

I have proposed above that what have been called direct objects in EP should be understood as a subset of the objects of middle verbs. This section provides further evidence to support this proposal. The evidence is twofold. First, *i*-marked direct objects

[10] Tokelauan is similar to Pukapukan in that a number of middle verbs must be affixed with -*Cia* when occurring in the agentive pattern: e.g., *alofa-gia* 'to love' > *alofa* 'to love', *pelo-gia* 'to deceive' > *pelo* 'to deceive', and *tago-fia* 'to remove' > *tago* 'to touch' (Hovdhaugen et al. 1989; Sharples 1976). On the other hand, a number of middle verbs can be used in the agentive pattern without a suffix, but with the expected semantic effect: e.g., *kikila* 'to see' versus 'to supervise' and *teteke* 'to oppose' versus 'to reject' (Hovdhaugen et al. 1989:86).

(i) a. *Na teteke ia ki te fuafuaga*
PST oppose 3SG to the plan
'He opposed the plan.'

b. *Na teteke e ia te fuafuaga*
PST oppose ERG 3SG the plan
'He rejected the plan.'

are treated like middle objects, but differently from unmarked objects of the agentive pattern. Second, *i*-marked objects are treated on a par with middle objects, but differently from other kinds of oblique noun phrases (e.g., *ki*-marked goals).

5.1 Anaphoric *ai* in Relative Clauses

One of the relativisation strategies used in Polynesian is the *ai* strategy, in which the relative clause contains a resumptive element *ai* in lieu of a gap. This strategy is typically used for relativisation of oblique NPs (Chapin 1974; Chung 1978; Massam and Roberge 1997). In many languages, the *ai* strategy is also required for relativisation of middle objects, as shown below.

(23) a. **Tongan**

Ko e siana [*na'a ku tokoni ki* ***ai***].
PRED REF man PST 1SG help to AI
'the man I helped'

b. **Niuean** (Seiter 1980:95)

e fakamatalaaga [*ne fanogonogo a au ki* ***ai***].
ABS speech NFUT listen ABS 1SG to AI
'the speech which I listened to'

c. **Tuvaluan** (Besnier 2000:71)

Ttino [*e ttau koe o fano ki* ***ei***] *ko Tito*
the.person NPST must 2SG COMP go to AI FOC Tito
'The person you should go see is Tito.'

In Māori, the *ai* strategy can be used for *i*-marked objects as well as middle objects. Gibson and Starosta (1990) noted this point as evidence for intransitivity of so called accusative constructions in Māori. Similar examples can be found in Hawaiian, although in both languages, the *ai* strategy is not the preferred method of relativisation for these NPs.[11]

(24) **Māori**

a. *He tāriana te poaka* [*i pupuhi* ***ai*** <*i*-marked>
INDF boar DEF pig PST shot AI

taku matua].
POSS.1.SG father
'The pig my father shot was a boar.' (Biggs 1998:159)

b. *E mōhio ana au ki te tangata* [*i* <*i*-marked>
UNS know PROG 1SG to the man PST

patu ***ai*** *a Hōne]*.
hit AI PERS John
'I know the man whom John hit.' (Chung 1978:72)

---

[11] According to Bauer (1997:569–572), the use of the pronoun strategy with *ai* for relativisation of *i*-marked direct objects is accepted only in non-past contexts and the use of *ai* with middle verbs is not obligatory and is 'sometimes rejected by fluent speakers.' Harlow (2001) also states that many speakers regard the use of this relativisation strategy for both *i*-marked direct objects and middle objects as 'not correct', although such examples are easily found.

c. *Nei te take ..., he kupu* [*i rongo* ***ai*** *au*]. <middle>
nearI the reason a word TAM hear AI 1SG
'Here is the reason … [it] is some words that I heard' (Bauer 1997:569)

(25) **Hawaiian** (Hawkins 1982:109, 113)

a. *Ua kanu ka makua i ka pōpoki* [*i* <*i*-marked>
PFV bury the parent OBJ the cat PST

pepehi ***ai*** *nā keiki].*
kill AI DEF.PL child
'The parent buried the cat that the children had killed.'

b. *Ua hoʻi mai ke keiki* [***i*** *kōkua* ***ai*** <middle>
PFV return DIR.1 the child PST help AI

ka mākaʻi]
the policeman
'The child who the policeman helped returned.'

It should be noted that when the pronoun strategy applies to subjects in languages like Tongan and Samoan, the resumptive element must be a personal pronoun, not *ai*.[12] Thus, *i*-marked objects behave like middle objects, rather than core arguments of the canonical transitive construction, not only in terms of the choice of relativisation strategy, but also with respect to the types of resumptive elements.

(26) **Tongan**

a. *Ko e tangata* [*naʻa* ****(ne)*** *langa ʻa e fale*]
PRED REF man PST 3SG build ABS REF house
'the man who built the house'

b. **Samoan** (Chung and Seiter 1980:633)

le tagata [*sā* ***(ia)*** *ʻare-ina laʻu ta'avale*]
the person PST 3SG drive-TR POSS.1SG car
'the person who drove my car'

c. **Tuvaluan** (Besnier 2000:66)

Ko ttagata teelaa [*e fai nee* ***ia*** *te aamio tonu*]
FOC the.man that NPST do ERG 3SG the behaviour right
'A [leader] who is righteous' (lit. 'that man who does the right deeds')

5.2 Other common factors

Additional data demonstrate that *i*-marked objects and middle objects together constitute a separate class from other oblique NPs. Māori provides some evidence that

[12] In Tongan, the personal pronoun strategy is obligatory for ERG subjects, but impossible for ABS subjects as well as objects (Otsuka 2000). In Samoan, it is optional for ERG as well as ABS subjects, but impossible for direct objects (Chung and Seiter 1980; Mosel and Hovdhaugen 1992:639). In Tuvaluan, the personal pronoun strategy is strongly preferred for ERG subjects and the deletion strategy is strongly preferred for ABS NPs. A personal pronoun may also be used as an alternative to *ei* if the head noun refers to an animate entity and if the relativised element is marked with the preposition *i* or *ki* (Besnier 2000:67).

middle objects and *i*-marked objects are treated alike, but differently from other oblique NPs. First, whereas 'passivisation' is available for *i*-marked objects as well as middle objects, other oblique NPs such as *ki*-marked goals may not undergo passivisation (Bauer 1997:197; Chung 1978:173–174).

(27) **Māori** (Bauer 1997:197)

a. *ka pīrangi-tia e ia ngā mea katoa* (= 13a) <middle>
TAM want-PASS by 3SG the.PL thing all
'All the things are wanted by him.'

b. *Ka pū-hia e ia te manu* <*i*-marked>
TNS shoot-PASS by 3SG the bird
'The bird was shot by him.'

c. **Ka tua-ina te toki e ia i te rākau.* <instrument>
TAM chop-PASS the axe by 3S OBJ the tree
'The axe was chopped down the tree by him.'

Second, the actor-emphatic strategy may be used for relativisation of both *i*-marked objects and middle objects, but not other types of oblique NPs.

(28) **Māori** (Bauer 1997:198–199)

a. *Ko tēnei te kōtiro nā Pani* [*i whāngai*] <*i*-marked>
TOP this the girl belong Pani TAM feed
'This is the girl Pani fostered.'

b. *Ko tēnei te kōtiro nā Pani* [*i karanga*] <middle>
TOP this the girl belong Pani TAM call
'This is the girl Pani called.'

c. **Ko tēnei te kuia nā rāua* [*i harihari* <goal>
TOP this the old.woman belong 3DU TAM carry
te harakeke].
the flax
'This is the old woman they carried the flax to'.

A third point concerns genitive relative constructions (GRCs), in which the thematic subject of the relative clause is realised as a genitive seemingly modifying the head noun. This strategy is used for middle and *i*-marked objects, but its use for oblique relatives is limited. Only some dialects permit its use for oblique relatives and even in those dialects, it is a much less common method than the *ai* strategy (Bauer 1997:577).

(29) **Māori** (Bauer 1997:569, 570, 577)

a. *Ko tēnei te whare a Hata* [*i pīrangi ai*]. <middle>
TOP this the house of Hata TAM want AI
'This is the house that Hata wanted.' (Bauer 1997:569)

b. *Ka mōhio ahau ki te tangata a Hone* [*i* <*i*-marked>
TAM know 1SG to the man of John TAM
kōhuru ai].
murder AI
'I know the man that John murdered.'

c. … *ko te whare a te mōkai rā* [*i moe ai*] <locative>
TOP the house of the slave DIST TAM sleep AI
'… the house where the servant slept'

These facts demonstrate that *i*-marked objects and middle objects are syntactically different from oblique NPs. Outside EP, various syntactic phenomena treat middle objects on a par with unmarked objects of ergative verbs, but differently from obliques: e.g., GRCs in Tongan (Otsuka 2007), quantifier float in Tongan, and incorporation in Samoan (Chung 1978).

6 Conclusion: EP languages are not accusative

The present study supports Clark's (1973, 1976) position that PPN was ergative. Two dyadic constructions are reconstructed for PPN, active and agentive. The active pattern was lexically governed by middle verbs. The agentive pattern appears to have been somewhat productive. Middle verbs were permitted in the derived agentive pattern, in which they were suffixed with transitive *-Ci. For middle verbs, the choice between the two patterns was determined by a semantic factor, i.e., the degree of affectedness of the patient. The active pattern was used to indicate a low degree of affectedness and the agentive pattern, a high degree of affectedness. This limited productivity was eventually reduced to lexical differences (i.e., middle versus ergative verbs) outside EP.

PEP underwent two innovations: first, the extension of the active pattern to ergative verbs; and second, obligatory affixation of *-Cia in the agentive pattern. Consequently, the alternation between active and agentive patterns became available for all dyadic verbs. In this view, what changed in PEP is neither the loss of ergativity nor the development of passive, but rather, the development of a two-way system that is reminiscent of the so called focus/voice system of western Austronesian. (See Blust 2002 and Ross 2002 for discussion on the historical development of 'focus' in Autronesian languages.) The active pattern is similar to the actor focus in that the actor is the prominent NP. Likewise, the agentive pattern is analogous to the patient focus in that the patient is the prominent NP. Unlike the relation between active and passive voices, neither the agentive nor the active pattern is 'basic' in this paradigm. This alternative view explains why middle verbs as well as intransitive verbs involving directly affected locatives can be 'passivised' in EP. The latter could be regarded as a very limited application of locative focus.

In addition, agentive ('passive') constructions in EP show pragmatic and syntactic characteristics similar to those of the patient focus of western Austronesian. Pragmatically, linguists have observed that the patient focus is the default choice in discourse as a means to achieve topic continuity (Huang 2000; Cooreman 1982, 1983, 1987; Cooreman et al. 1984; Payne 1994). In light of this observation, the high text frequency of 'passive' in EP is nothing unusual. Bauer (1997) has suggested that 'passive' is used in Māori in order to keep the discourse topic in the subject position. Syntactically, it is known that certain syntactic operations such as relativisation are restricted to the 'focused' NP (Keenan 1976; Schachter 1976, among others). Specifically, relativisation of the patient is permissible if the relevant construction is in the patient focus, but not in the actor focus. In EP languages, relativisation of the patients of active constructions is rare and the 'passive' strategy is strongly preferred (Bauer 1997; Harlow 1996, 2001).

These facts at least suggest that there is reason to question the long standing view that EP languages are accusative (cf. Gibson and Starosta 1990). Intriguingly, some have

proposed that Formosan and Philippine languages should be understood as ergative, with the patient focus being transitive (i.e., ergative) and the actor focus, intransitive (Aldridge 2004; Gerdts 1988; De Guzman 1988; Liao 2004; Payne 1982; Reid and Liao 2004; Starosta 1988, 1999, among others). The present study has shown that EP languages and the rest of Polynesian share the same construction types, but only with different levels of productivity. Whether we should call the relevant system ergative or not may be a matter of another long lasting debate.

References

Aldridge, Edith. 2004. Ergativity and word order in Austronesian languages. PhD dissertation, Cornell University.

Alexander, James D. 1981. Case-marking and passivity in Easter Island Polynesian. *Oceanic Linguistics* 20:131–149.

Bauer, Winifred. 1997. With William Parker, Te Kareongawai Evans and Te Aroha Noti Teepa *The Reed reference grammar of Māori*. Auckland: Reed.

Besnier, Niko. 2000. Tuvaluan: a Polynesian language of the Central Pacific. London: Routledge.

Biggs, Bruce. 1998 [1969]. *Let's learn Māori*. Auckland: Auckland University Press.

—— 1974. Some problems of Polynesian grammar. *Journal of the Polynesian Society* 83:401–426.

Blust, Robert A. 1986. A rediscovered Austronesian comparative paradigm. *Oceanic Linguistics* 16:1–51.

—— 2002. Notes on the history of 'focus' in Austronesian languages. In Wouk and Ross, eds 2002:63–78.

Chapin, Paul G. 1974. Proto-Polynesian *ai. *Journal of the Polynesian Society* 83:259–307.

Chung, Sandra. 1978. *Case marking and grammatical relations in Polynesian*. Austin: University of Texas Press.

Chung, Sandra and William J. Seiter. 1980. The history of raising and relativization. *Language* 56:622–638.

Churchward, Spencer. 1928. On the origins of the Polynesian passive. *Journal of the Polynesian Society* 37:300–305.

Clark, Ross. 1973. Transitivity and case in Eastern Oceanic languages. *Oceanic Linguistics* 12:559–605.

—— 1976. *Aspects of Proto-Polynesian syntax*. Auckland: Linguistic Society of New Zealand.

Cooreman, Ann M. 1982. Topicality, ergativity, and transitivity in narrative discourse: Evidence from Chamorro. *Studies in Language* 6:343–374.

—— 1983. Topic continuity and the voicing system of an ergative language. In Talmy Givón, ed. *Topic continuity in discourse: a quantitative cross-language study*, 425–89. Amsterdam: John Benjamins.

—— 1987. *Transitivity and discourse continuity in Chamorro narratives*. Berlin: Mouton de Gruyter.

Cooreman, Ann M., Barbara Fox and Talmy Givón. 1984. The discourse definition of ergativity. *Studies in Language* 8:1–34.

Coppenrath, Hubert and Paul Prevost. 1975. *Grammaire approfondie de la langue tahitienne (ancienne et moderne)*. Papeete: Librairie Pureora.

De Guzman, Videa P. 1988. Ergative analysis for Philippine languages: An analysis. In McGinn, ed. 1988:323–345.

Du Feu, Veronica. 1996. *Rapanui*. London and New York: Routledge.

Elbert, Samuel H. and Mary Kawena Pukui. 1979. *Hawaiian grammar*. Honolulu: University of Hawai‘i Press.

Gerdts, Donna B. 1988. Antipassives and causatives in Ilokano: evidence for an ergative analysis. In McGinn, ed. 1988:295–321.

Gibson, Jeanne and Stanley Starosta. 1990. Ergativity east and west. In Philip Baldi, ed. *Linguistic change and reconstruction methodology*, 195–210. Berlin: Mouton de Gruyter.

Hale, Kenneth. 1968. Review of Hohepa 1967. *Journal of the Polynesian Society* 77:83–99.

Harlow, Ray. 1996. *Māori*. München and Newcastle: LINCOM Europa.

—— 2001. *A Māori reference grammar*. Auckland: Longman.

Hawkins, Emily. 1982. *Pedagogical grammar of Hawaiian: recurrent problems*. Honolulu: Hawaiian Studies Program, University of Hawai‘i at Mānoa.

Hohepa, Patrick. 1969. The accusative-to-ergative drift in Polynesian languages. *Journal of the Polynesian Society* 78:295–329.

Hopkins, Alberta Pualani. 1992. *Ka lei ha‘aheo: beginning Hawaiian*. Honolulu: University of Hawai‘i Press.

Hovdhaugen, Even, Ingjerd Hoëm, Consulata Mahina Iosefo and Arnfinn Muruvik Vonen. 1989. *A handbook of the Tokelau language*. Oslo: Norwegian University Press, Institute for Comparative Research in Human Culture. Distributed by Oxford University Press.

Huang, Shuangang. 2000. The pragmatics of focus in Tsou and Seediq. *National Taiwan University Working Papers in Linguistics* 3:111–127.

Keenan, Edward L. 1976. Remarkable subjects in Malagasy. In Li, ed. 1976:247–301.

Lazard, Gilbert and Louise Peltzer. 2000. *Structure de la langue tahitienne*. Paris: Peeters.

Li, Charles, ed. 1976. *Subject and topic*. New York: Academic Press.

Liao, Hsiu-chuan. 2004. Transitivity and ergativity in Formosan and Philippine languages. PhD dissertation, University of Hawai‘i at Mānoa.

Marck, Jeff. 2000. *Topics in Polynesian language and culture history*. Canberra: Pacific Linguistics.

Massam, Diane and Yves Roberge. 1997. Operator-bound clitics and Niuean *ai*. In Jim Black and Virginia Motapanyane, eds *Clitics, pronouns and movement*, 273–300. Current Issues in Linguistic Theory 140. Amsterdam and Philadelphia: John Benjamins.

McGinn, Richard, ed. 1988. *Studies in Austronesian linguistics*. Athens, Ohio: Center for Southeast Asia Studies, Center for International Studies, Ohio University.

Mosel, Urlike and Even Hovdhaugen. 1992. *Samoan reference grammar*. Oslo: Scandinavian University Press.

Moyse-Faurie, Claire. 1992. Verb classes and argument structure variation in Futunan. *Oceanic Linguistics* 31:209–228.

Otsuka, Yuko. 2000. Ergativity in Tongan. PhD thesis, University of Oxford.

—— 2007. Genitive relative constructions and agentless transitives in Tongan. A paper presented at the 14th meeting of the Austronesian Formal Linguistics Association. May 2007, McGill University.

Pawley, Andrew. 1973. Some problems in Proto-Oceanic grammar. *Oceanic Linguistics* 12:103–88.

—— 1980. On the history of Polynesian transitive clauses. Unpublished manuscript.

—— 2001. Proto-Polynesian *-CIA. In Joel Bradshaw and Kenneth Rehg, eds *Issues in Austronesian morphology*, 193–216. Canberra: Pacific Linguistics.

Payne, Thomas E. 1982. Role and reference related subject properties and ergativity in Yup'ik Eskimo and Tagalog. *Studies in Language* 6:75–106.

—— 1994. The pragmatics of voice in a Philippine language: actor-focus and goal focus in Cebuano narrative. In Talmy Givón, ed. *Voice and inversion*, 317–364. Amsterdam: John Benjamins.

Reid, Lawrence A. and Hsiu-chuan Liao. 2004. A brief syntactic typology of Philippine languages. *Language and Linguistics* 5:433–490.

Ross, Malcolm. 2002. The history and transitivity of western Austronesian voice and voice-marking. In Wouk and Ross, eds 2002:17–62.

Schachter, Paul. 1976. The subject in Philippine languages: Actor, topic, actor-topic, or none of the above. In Li, ed. 1976:491–518.

Seiter, William J. 1980. *Studies in Niuean syntax*. New York: Garland.

Sharples, Peter R. 1976. Tokelauan syntax: studies in the sentence structure of a Polynesian language. PhD thesis, University of Auckland.

Starosta, Stanley. 1988. A grammatical typology of Formosan languages. *Bulletin of the Institute of History and Philology* 59:541–576.

—— 1999. Transitivity, ergativity and the best analysis of Atayal case marking. In Elizabeth Zeitoun and Paul Jen-kuei Li, eds *Selected papers from the Eighth International Conference on Austronesian Linguistics (8ICAL)*, 371–392. Taipei: Academia Sinica.

Tchekhoff, Claude. 1973. Some verbal patterns in Tongan. *Journal of the Polynesian Society* 82:281–292.

Weber, Nancy L. 1988. Case marking in Rapa Nui, the Polynesian language of Easter Island. MA thesis, University of Texas at Arlington.

Williams, Herbert. 1928. Some observations on Polynesian verbs. *Journal of the Polynesian Society* 37:306–317.

Wouk, Fay and Malcolm Ross, eds. 2002. *The history and typology of western Austronesian voice systems*. Canberra: Pacific Linguistics.

18 *Inclusory constructions and their development in Philippine languages*

LAWRENCE A. REID

1 Introduction[1]

In many Philippine languages it is possible to express plural participants in an activity by conjoining two or more noun phrases with the same case-marking. In Tagalog (Schachter and Otanes 1972:115–116), for example, the conjunction *at* 'and' conjoins NPs[2] that express nominative (common) nouns, as in (1a), while it conjoins NPs that express genitive and locative (personal) nouns, as in (1b-c), respectively. The second NP in such coordinate constructions may or may not be required to be preceded by a nominal specifier marking case and/or the semantic features of the following noun, as in (1)a, in which the form marking the following noun as a common noun is optional. Similar constructions occur widely in Philippine languages, as exemplified also in Masbatenyo (2), and in Khinina-ang Bontok (3).

[1] It is a great pleasure to be able to present this article to Bob Blust, whose friendship and scholarship have continued to inspire me over the four decades since we first met, he as a beginning student in linguistics and me teaching my first course as a new PhD at the University of Hawai`i. The extent of Bob's scholarship is so extensive, that it is difficult to find an area in comparative Austronesian to which he has not already made substantial contributions. It is my hope that this foray into some of the features of Philippine comparative syntax and grammaticalization will fill one of the minor gaps that Bob has not yet ventured into.

[2] Abbreviations: 1, 2, 3 – first, second, third person, PCPH – Proto Central Philippines, BAN – Bantoanon, BOH – Boholano, CEB – Cebuano, DTG – Datagnon, HIL – Hiligaynon (Ilonggo), JAU – Jaun-jaun, KUY – Kuyonon, SIB – Sibale, SOR – Sorsoganon, SUR – Surigaonon, TAG – Tagalog, WAR – Waray, ADDR – addressee, ANC – associative nominal construction, COM – common, DEF – definite, EMPH – emphasis, EXCL – exclusive, FUT – future, GEN – genitive, INCL – inclusive, IPC – inclusive pronominal construction, LG – ligature, LOC – locative, NEG – negative, NOM – nominative, NP – noun phrase, NS – nominal specifier, OBL – oblique, PERS – personal, PL – plural, PROX – proximate, RPRT – reportative, SG – singular, SPEC – specifier, SPKR – speaker, TOP – topic, TOPLK – topic linker, UNM – unmarked

Alexander Adelaar and Andrew Pawley, eds
Austronesian historical linguistics and culture history: a festschrift for Robert Blust, 267–294.
Pacific Linguistics, 2009.
© This edition vested with Pacific Linguistics.

(1) Tagalog[3]

a. *Nakita =ko ang babae **at** (ang) lalaki.*
saw =GEN.1SG NS.COM woman and (NS.COM) man
'I saw the woman and the man.'

b. *Pinanood ni Belen **at** ni Rosa ang parada.*
saw GEN.PERS.SG Belen and GEN.PERS.SG Rosa NS.COM parade
'Belen and Rosa saw the parade.'

c. *Susulat =ako kay G Reyes **at***
will.write =NOM.1SG LOC.PERS.SG Mr Reyes and

kay Gng Quizon.
LOC.PERS.SG Mrs Quizon.
'I'll write to Mr Reyes and Mrs Quizon.'

(2) Masbatenyo (Wolfenden 2001:258)

*Adi si Maria **kag** si Pedro.*
here NS.PERS.SG Maria and NS.PERS.SG Pedro
'Maria and Pedro are here.'

(3) Khinina-ang Bontok[4]

*Omey am-in nan fafarro **si** mamagkhit ay insamar.*
go all NS.COM young.men and young.women LG prepare.pondfield
'All the young men and women go to prepare the soil in the pondfields (for planting).'

Similar constructions may also be employed when the first of two potential coordinate constructions is expressed by a personal pronoun and the second by a nominal other than a personal pronoun, as in (4a), in which the NPs are nominative, and (4b), in which the NPs are formally marked with locative prepositions, and express a dative case relationship.

(4) Tagalog

a. *Nakita =ko siya **at** si Juan.*
saw =GEN.1SG 3SG and NS.PERS.SG Juan
'I saw him/her and Juan.'

b. *Ibibigay =ko ito sa kaniya **at** kay Juan.*
will.give =GEN.1SG this LOC OBL.3SG and LOC.PERS.SG Juan
'I'll give this to him/her and Juan.'

Schachter and Otanes (1972:115–116) noted, that in constructions such as these, where the first of two potential coordinates is expressed by a personal pronoun and the second by a nominal other than a personal pronoun, 'Tagalog in some cases allows, in others requires,

---

[3] Tagalog examples are either taken directly from Schachter and Otanes (1972) or have been verified as grammatical by Ricardo Nolasco and Tish Bautista. The transcriptions of published source materials throughout the paper are retained as in the originals, except for the addition of = signs to mark clitic pronouns. Translations have been modified and verified with native speakers, where appropriate, to more accurately reflect the inherent ambiguities in the data.

[4] Khinina-ang Bontok data are from my own field notes and have been confirmed as grammatical by a native speaker, Susan Catay.

the use of a SPECIAL COORDINATE CONSTRUCTION in place of the coordinate construction with *at*.' These constructions require a plural personal pronoun followed by a genitively marked NP, regardless of the case of the preceding pronoun, as in (5a-c), in which the third person pronouns are respectively unmarked for case (5a), genitive (5b) and locative (5c). In these constructions, even though the pronoun is plural, its reference is necessarily singular. The non-pronominal NP in these constructions expresses only one (or more) of the most salient participants included in the set covered by the pronoun.

(5) Tagalog

a. *Nakita =ko sila **ni** Juan.*
saw =GEN.1SG 3PL GEN.PERS.SG Juan
'I saw him/her and Juan.'

b. *Ginawa =nila **ni** Juan ang trabaho.*
made =GEN.3PL GEN.PERS.SG Juan NS.COM work
'He/She and Juan did the work.'

c. *Ibibigay =ko ito sa kanila **ng** bayaw =mo.*
will.give =GEN.1SG this LOC OBL.3PL GEN.COM brother-in-law =GEN.2SG
'I'll give this to him/her and your brother-in-law.'

Similar constructions occur also in Sorsoganon, a member of Zorc's (1977) Peripheral subgroup of the Central Bisayan dialects and, like Tagalog, a member of the Greater Central Philippine subgroup of Philippine languages (Blust 1991), as in (6).

(6) Sorsoganon[5]

a. *Nakita =ko sinda **ni** Juan.*
saw =GEN.1SG 3PL GEN.PERS.SG Juan
'I saw him/her and Juan.'

b. *Ginibo =ninda **ni** Juan an trabaho.*
made =GEN.3PL GEN.PERS.SG Juan NS.COM work
He/She and Juan did the work.'

c. *Ihahatag =ko ini sainda **san** bayaw mo.*
will.give =GEN.1SG this LOC.3PL GEN.PERS.SG brother-in-law GEN.2SG
'I'll give this to him/her and your brother-in-law.'

The first linguist to describe constructions such as these in Philippine languages was Blake (1916), who in an insightful paper referred to them as 'explicative' coordinate constructions. These kinds of constructions are found in many languages and have been referred to by Lichtenberk (2000:2) as INCLUSORY PRONOMINAL CONSTRUCTIONS (IPC).[6] In describing such constructions in Toqabaqita, Lichtenberk says, 'a pronominal form that identifies a total set of participants, a subset of which is identified by a lexical NP, will be referred to as an 'inclusory pronominal' … [and] the lexical noun phrase that identifies a subset of the set encoded by the inclusory pronominal will be referred to as the 'included noun phrase'. Lichtenberk uses the term 'inclusory pronoun' to refer specifically to independent personal pronouns that are inclusory, in that their reference includes not only

[5] All Sorsoganon examples have been provided and checked by Maria Sheila Zamar.

[6] See Lichtenberk (2000) for examples and a wide-ranging discussion of such constructions in Toqabaqita, an Oceanic language.

a single speaker, addressee or third person, but also the lexical NP that follows. Since in Philippine languages there are no 'subject-tense markers', as in Toqabaqita, that require a distinction in terminology between 'inclusory pronominals' and 'inclusory pronouns', all inclusory pronominal forms will be referred to here as 'inclusory pronouns'.

Philippine languages also typically allow the expression of one or more of the salient participants of a larger set by a construction type in which a 'plural' nominal specifier is followed by a personal noun which functions to label a total set of participants, as in (7a-c), in which the nominal specifiers are respectively unmarked for case (7a), genitive (7b) and locative (7c). This type of inclusory construction is referred to in this paper as an ASSOCIATIVE NOMINAL CONSTRUCTION (ANC).

(7) Tagalog

a. *Nakita =ko **sina** Juan.*
saw =GEN.1SG NS.PERS.PL Juan
'I saw Juan (and others).'

b. *Ginawa **nina** Juan ang trabaho.*
made GEN.PERS.PL Juan NS.COM work
'Juan (and others) did the work.'

c. *Ibibigay =ko ito **kina** Juan.*
will.give =GEN.1SG this LOC.PERS.PL Juan
'I'll give this to Juan (and others).'

Inclusory pronominal constructions will be discussed in §2, inclusory nominal, or associative nominal constructions in §3, and their historical development in §4.

2 Inclusory pronominal constructions

IPCs consist of two parts, the INCLUSORY PRONOUN and the INCLUDED NOUN PHRASE. These constructions correspond in type to the Toqabaqita constructions that Lichtenberk labels 'split-phrasal' inclusory constructions, in that while both the inclusory pronoun and the included NP frequently occur in apposition, they are separable. In §2.1 we examine the inclusory pronouns, while in §2.2 we discuss the included noun phrase, and in §2.3 give evidence for the 'split-phrasal' nature of these constructions. Section 2.4 will deal with phrasal IPCs.

2.1 Inclusory pronouns

The examples of inclusory pronouns given in (5)–(6) above are all third person plural forms. However, when they occur as part of an IPC, their reference is distinctive, in that although their form is plural, their pronominal reference is singular. The total set of participants expressed by the construction is plural, including along with the pronominal referent, the lexical referent(s) that follows, expressed by the included NP.

In addition to third person inclusory pronouns, Tagalog also has IPCs with first and second person inclusory pronouns followed by an included NP. As with third person inclusory pronouns, first person exclusive, and second person pronouns that are inclusory, unless otherwise specified (as by a numeral (8b)), identify only a single pronominal referent, as in (8a,c). A first person plural inclusive pronoun that is inclusory, expresses at least two pronominal referents, a single speaker and a single addressee and at least a single

included nominal, as in (8d), although it may be interpreted as having plural pronominal referents. For each of these, the inclusory NP represents a set of one (or more) of the salient participants included in the pronominal reference. Other Philippine languages exhibit similar constructions, as in (9)–(14). It should be noted that, unlike many other Philippine languages, most dialects of Tagalog no longer distinguish a first person dual pronoun ('we two') from a first person inclusive pronoun ('we all').

(8) Tagalog

a. *Maglalakad =kami* ***ng*** *Nanay.*
will.walk =1PL.EXCL GEN.COM.SG mother
'Mother and I will walk.'

b. *Maglalakad =tayo =ng tatlo* ***ng*** *Nanay.*
will.walk =1PL.INCL =LG three GEN.COM.SG mother
'The three of us, Mother, you and I will walk.'

c. *Ibibigay =ko ito sa inyo* ***ng*** *bayaw =mo.*
will.give =GEN.1SG this LOC OBL.2PL GEN.COM brother-in-law =GEN.2SG
'I'll give this to you (SG) and your brother-in-law.'

d. *Lilinisan =natin* ***ni*** *Maria ang bahay.*
will.clean =GEN.1PL.INCL GEN.PERS.SG Maria NS.COM house
'We, including Maria, are cleaning the house.'

(9) Sorsoganon

a. *Malakat =kami* ***ni*** *Mamay.*
will.walk =1PL.EXCL GEN.PERS.SG mother
'Mother and I will walk.'

b. *Malakat =kamo* ***ni*** *Mamay.*
will.walk =2PL GEN.PERS.SG mother
'Mother and you (SG) will walk.'

(10) Northern Subanen (Daguman 2004:169)

Miktuntultuntul =gami ***ni*** *Junjun.*
casually.talked =NOM.1PL.EXCL GEN.PERS.SG Junjun
'We, Junjun and I, casually talked.'

(11) Mansaka (Svelmoe and Svelmoe 1974:56)

Kikita =ko kamo ***si*** *Ilik.*
see =GEN.1SG NOM.2PL PERS.SG Ilik
'I see you (SG) and Ilik.'

(12) Ilianen Manobo (Wrigglesworth 1971:121)

Ne embiya egkeamin ini se egkeenen, dey
then if consumed this the food GEN.1PL.EXCL

ki *ina …*
OBL.PERS mother
'Then if our food here is used up, mine with mother's …'

(13) Ilokano[7]

*Mapan =kami **ken** Jose.*
go =NOM.1PL.EXCL OBL.PERS Jose
'Jose and I are going.'

(14) Khinina-ang Bontok

a. *Chinarosan =mi **an** Pakoran nan*
cleaned =GEN.1PL.EXCL OBL.PERS.SG Pakoran NS.COM

pantew =cha.
yard =3PL
'Pakoran and I cleaned their yard.'

b. *Inmey =kayo **an** Pakoran ay mangila =s nan*
went =2PL OBL.PERS.SG Pakoran LG see =OBL NS.COM

arang =cha.
granary =3PL
'You (SG) and Pakoran went to see their granary.'

It should be noted that although languages such as Ilokano and Khinina-ang Bontok have inclusory pronoun constructions with non-third person included pronouns, as in (13) and (14) above, they do not allow third person inclusory pronoun constructions corresponding to the Tagalog and Sorsoganon examples in (5) and (6). The sentences in (15) and (16) are grammatical only with the meanings given in parentheses; they are ungrammatical with the meanings shown with a preceding asterisk. To express meanings such as these, sentences with associative nominal constructions are used (see §3).

(15) Ilokano

a. *Mapan =da kenni Jose.*
go =3PL OBL.PERS Jose
*'He/She and Jose are going.' ('They are going to Jose's place.')

b. *Mapan =da iti gayyem =na.*
go =3PL OBL.COM friend =GEN.3SG
*'He/She and his/her friend are going.' ('They are going to his/her friend's place.')

(16) Khinina-ang Bontok

a. *Inilak cha-icha **an** Pakoran.*
saw.1SG 3PL OBL.PERS.SG Pakoran
* 'I saw him/her and Pakoran.' ('I saw them with Pakoran.')

b. *Khina-eb =cha **an** Pakoran nan afong.*
made =GEN.3PL OBL.PERS.SG Pakoran NS.COM house
*'He/She and Pakoran built the house.' ('They, with Pakoran, built the house.')

---

[7] Ilokano data have been checked and verified as grammatical by Carl Rubino and Elizabeth Calinawagan.

2.2 Included noun phrases

Just as inclusory pronouns in Philippine languages are distinctive, included NPs are also distinctive, in that although they are formally marked with either genitive or oblique case-marking forms, their interpretation is different from other NPs that are similarly marked, and depend for their interpretation on the presence of an inclusory pronoun preceding them. In Tagalog, for example, genitive NPs occur as follows:

(1) To encode a post-nominal possessor, a common noun is introduced by *ng* (/naŋ/), while a personal noun is introduced by *ni*—or *nina* when 'plural' (see §3)—or one of the enclitic genitive pronouns or demonstrative forms. A common noun functioning as a possessor in such a construction is interpretable as either definite or indefinite, as in (17a–d).

(17) Tagalog

a. *aso* ***ng*** *lalaki* 'the/a man's dog'
b. *aso* ***ni*** *Juan* 'Juan's dog'
c. *aso* ***niya*** 'his dog'
d. *aso* ***nito*** 'this one's dog'

(2) To encode the agent of a transitive clause. A genitive NP with this function is marked in precisely the same way as a post-nominal possessor, as in (18a–d).

(18) Tagalog

a. *Ginawa* ***ng*** *lalaki* *ang* *silya.*
made GEN.COM man NS.COM chair
'The/A man made the chair.'

b. *Ginawa* ***ni*** *Juan* *ang* *silya.*
made GEN.PERS.SG Juan NS.COM chair
'Juan made the chair.'

c. *Ginawa* =***niya*** *ang* *silya.*
made =GEN.3SG NS.COM chair
'He made the chair.'

d. *Ginawa* =***nito*** *ang* *silya.*
made =GEN.PROX NS.COM chair
'This one made the chair.'

(3) to encode the second NP (the patient) of a dyadic intransitive clause, i.e., a clause that is morphologically marked as intransitive, but which has two core noun phrases (Dixon and Aikhenvald 2000:3). In such a construction the genitive NP is always indefinite or partitive, and its referent can only be either a common noun introduced by *ng* (/naŋ/) or a genitive demonstrative (such as *nito*), with a partitive interpretation. A genitively-marked personal noun cannot occur, as in (19).

(19) Tagalog

a. *Gumawa* *si* *Juan* ***ng*** *silya.*
made NS.PERS Juan GEN.COM chair
'Juan made a chair/chairs.'

b. *Gumawa si Juan* ***nito.***
made NS.PERS Juan GEN.PROX
'Juan made some of these.'

In addition, in Tagalog, a genitive NP is used to express an included NP. A genitive NP with this function, however, is different from those noted above, in that it can encode either a common noun introduced by *ng* (/naŋ/), typically with a definite interpretation, or a personal noun introduced by *ni*—or *nina* when 'plural'—but not a genitive pronoun or a demonstrative. It functions moreover, not as a possessor, agent, or patient, but simply to name one (or more) of the salient members of the set specified by the previous plural pronoun, regardless of its case-marking. To receive this interpretation then, a genitive NP must be preceded by an inclusory pronoun. An included NP is distinctive, or 'special', but not in the sense suggested by Schachter and Otanes (1972:116), since it cannot be interpreted as either coordinated with, or a concomitant of, the set specified by the pronoun, but rather as one of the participants included within the set. Lichtenberk (2000) notes that 'constructions with inclusory pronominals have usually been analyzed as coordinate or comitative … In Toqabaqita … they are neither.' The same is true also of these constructions in Philippine languages.

2.3 'Split-phrasal' inclusory pronominal constructions

In his typology of IPCs in Toqabaqita, Lichtenberk distinguishes between two general types. Those in which the inclusory pronominal and the included NP form a phrase, he labels as 'phrasal' inclusory pronominals. Those constructions which do not form a single phrase he labels as 'split-phrasal type' (Lichtenberk 2000:3). All of the examples given in the preceding sections are of the latter type, in that while both the inclusory pronoun and the included noun NP frequently occur in apposition, they do not constitute a single phrase and are separable.

In Philippine languages, clitic pronouns typically occur in second position in a clause, immediately following the main lexical verb as shown in the examples in the preceding sections, but when the lexical verb is preceded by an 'auxiliary' verb, such as a negative, pronouns occur between the two verbs, resulting in the separation of the inclusory nominal from its dependent included NP, as in (20a), in which the two parts of the inclusory construction are surrounded by square brackets. The inclusory pronoun can also be separated from its included NP by temporal adverbial expressions, as in (20b-c).

(20) Tagalog

a. *Hindi* [=*nila*] *ginawa* [*ni Juan*] *ang trabaho.*
NEG =GEN.3PL made GEN.PERS.SG Juan NS.COM work
'He and Juan didn't do the work.'

b. *Binili* [=*namin*] *kahapon* [*ni Juan*] *ang kotse.*
bought =GEN.1PL.EXCL yesterday GEN.PERS.SG Juan NS.COM car
We, Juan and I, bought the car yesterday.'

c. *Nakita =ko* [*sila*] *kanina* [*ni Maria*] *sa parke.*
saw =GEN.1SG 3PL earlier GEN.PERS.SG Maria LOC park
'I saw her/him and Maria in the park earlier today.'

In Sorsoganon, similar constructions occur in which the parts of the inclusory constructions are separated. In (21a), the inclusory pronoun is fronted to follow the negative verb, while in (21b-c), temporal adverbs separate the two parts of the construction. In Sorsoganon, unlike in Tagalog, genitive common NPs may be marked as either definite (with *san*) or indefinite (with *sin*) (Zorc 1977:85). A Sorsoganon genitively-marked common NP in an inclusory construction is 'special', because it can only be marked with the definite form, as in (21b-c). The indefinite, or non-specific, form is incompatible with the function of the construction to express a salient member of the pronominal set.

(21) Sorsoganon

a. *Dili* [=*ninda*] *ginibo* [*ni* *Juan*] *an* *trabaho.*
NEG =GEN.3PL made GEN.PERS.SG Juan NS.COM work
'He and Juan didn't do the work.'

b. *Ginibo* [=*ninda*] *kahapon* [*san* *lalaki*] *an* *trabaho.*
made =GEN.3PL yesterday GEN.COM.DEF man NS.COM work
'He/she and the man did the work yesterday.'

c. *Kinuwa* [=*namon*] *kanina* [*san* *mga* *batit*] *an*
took =GEN.1PL.EXCL earlier GEN.COM.DEF PL child NS.COM
mga *burak.*
PL flower
'We, the children and I, took the flowers earlier today.'

In the Ilokano examples in (22a-b), a future adverbial enclitic (=*nto*) attaches directly to the inclusory pronoun, separating it from the included NP. In (22b), the inclusory pronoun and its enclitic future adverb are further separated by their second-position occurrence following the negative verb (*saan*), and in (22c), the inclusory pronoun is separated from its included (common) NP by a temporal adverb.

(22) Ilokano

a. *Mapan* [=*kami*] =*nto* [*ken(ni)* *Jose*].
go =1PL.EXCL =FUT OBL.PERS(.SG) Jose
'Jose and I will go.'

b. *Saan* [=*kami*] =*nto* *a* *mapan* [*ken(ni)* *Jose*].
NEG =1PL.EXCL =FUT LG go OBL.PERS(.SG) Jose
'Jose and I won't go.'

c. *Inala* [=*mi*] *itattay* [*kadagiti* *ubbing*] *ti* *sabsabong.*
took =GEN.1PL.EXCL earlier OBL.COM.PL children NS.COM flowers
'We, the children and I, took the flowers earlier today.'

The 'special' status of included NPs is also apparent in the Ilokano examples in (22a-b). In each of these examples, the bracketed NP has at least two interpretations, one of which reads the NP as included in the preceding pronominal reference, and is the interpretation given in the free translations. It should be noted that the form marking the included noun is optionally *ken* or *kenni*, in which =*ni* marks the following personal noun explicitly as singular. (The equivalent oblique personal 'plural' form is *kada*.) Without =*ni*, the NP is only interpretable as an included dependent of the inclusory pronoun (or as a conjoined

NP, see §4.2). With *=ni*, the phrase is potentially ambiguous as being either an included NP, or a locatively-marked personal location, as shown in the free translations in (23).[8] Example (22c) cannot be interpreted as containing a personal noun location, because of its position prior to the nominative NP (*ti sabsabong*), and its semantic incompatibility. (See §4.2 for further explanation of the ambiguities in these forms.)

(23) Ilokano

a. *Mapan =kami =nto kenni Jose.*
go =1PL.EXCL =FUT OBL/LOC.PERS.SG Jose
'Jose and I will go.'/'We (EXCL) will go to Jose's place.'

b. *Saan =kami =nto a mapan kenni Jose.*
NEG =1PL.EXCL =FUT LG go OBL/LOC.PERS.SG Jose
'Jose and I won't go.'/'We (EXCL) won't go to Jose's place.'

In addition to distinguishing 'phrasal' and 'split-phrasal' types of inclusory constructions in Toqabaqita, Lichtenberk (2000:3) notes that these are cross-cut by whether or not there is an overt marker between the inclusory pronoun and the included NP. If there is, he considers the construction to be 'explicit'; if not, it is 'implicit'. All of the examples discussed in the preceding sections are split-phrasal, and since all have a form, either a genitive or oblique marker introducing the included NP, they are also explicit.

2.4 Phrasal inclusory pronominal constructions

A PHRASAL inclusory pronominal construction as defined by Lichtenberk, consists of a single NP containing an inclusory pronoun and an included noun, either with or without an overt marker, making it respectively either explicit or implicit. The only Philippine language for which data is available to clearly demonstrate this kind of inclusory construction is Tboli, one of the Southern Mindanao group of languages. This language has lost much of the morphological case marking that is found in other Philippine languages, so that grammatical relations are primarily signaled by word order, and there is no overt marker before the included noun, making the inclusory nature of the construction implicit. In (24), the phrasal inclusory constructions are bracketed, with their implicit case being marked by subscripted labels.[9] These constructions can occur as both nominative and genitive, with first, second and third person plural pronouns occurring as the inclusory pronominal head. That these constructions cannot be split, is suggested by (24e), in which the whole phrase, rather than the noun alone, occurs as a postposed topic.

(24) Tboli (Forsberg 1992:11, Porter 1966:8)

a. *Lewu* [*me Kasi*]$_{GEN}$ *funen.*
two 1PL.EXCL Kasi owner.3SG
Two of us, Kasi and I, are its owners.' (lit. 'We two Kasi are its owners.')

---

[8] Carl Rubino (pers. comm.) notes that in writing it is possible to distinguish the ambiguity by retaining the combined form, *kenni*, for a single personal location, while writing the form separately as *ken ni*, when it introduces an included NP.

[9] Tboli, like other Philippine languages, is analyzable as an ergative language, with patients of transitive sentences (labeled in this paper as nominative) typically following agents (labeled in this paper as genitive).

b. *Gunun deng nù se tahu blóng* [*ye Dimas*?]$_{GEN}$
where.3SG PAST be EMPH true division 2PL Dimas
'Where is the true boundary between you (SG) and Dimas?'

c. *Omin* [*le Yê Bong*]$_{NOM}$ *gna.*
and.then 3PL Mother Big go.ahead
'And then Big Mother and her companion went ahead.'

d. *Ton Kasi* [*le Walan.*]$_{NOM}$
saw Kasi 3PL Walan
'Kasi saw Walan (and the others).'

e. *Ton le mohin* [*le Walan.*]$_{TOP}$
saw 3PL sea 3PL Walan
'They saw the sea, Walan (and the others).'

Philippine languages that appear to have explicitly marked phrasal inclusory constructions are found in Palawan. In Southwest Palawano, one of the Meso-Philippine languages and part of Blust's Greater Central Philippines, phrasal inclusory constructions typically occur as the nominative complement of a numeral predicate specifying the total number of participants in the pronominal set, as in (25a-c), the whole construction being in apposition to a preceding plural pronoun.

(25) Southwest Palawano[10]

a. *Negtabo diye, dua [diye et bayew =ko.]*$_{NOM}$
marketing 3PL two 3PL OBL brother.in.law =GEN.1SG
'They went to market, two of them, including my brother-in-law.'
(lit. 'They were marketing, they and my brother-in-law were two.')

b. *Mesubo kay banar, dua* [*kay et si Arturo.*]$_{NOM}$
early 1PL.EXCL true two 1PL.EXCL OBL PERS.SG Arturo
'We (EXCL) were/will be early, two of us, Arturo and I.'

c. *Minuli kay, telo [kay de Arlyn.]*$_{NOM}$
went.home 1PL.EXCL three 1PL.EXCL PERS.PL Arlyn
'We (excl) went home, three of us, myself, Arlyn and his companion.'

We noted above, both for Tagalog and Sorsoganon, that an included NP with a singular noun is marked by one of the genitive noun markers, *ng* 'common noun', or *ni* 'singular personal noun'. In Southwest Palawano (Davis 1995), *ni* also marks genitive personal NPs expressing agents and possessors, but in an included NP this form does not occur. Rather, the form that is used (both for common and personal nouns) is the singular 'locative/oblique/dative' *et*. Singular personal nouns are distinguished from common nouns in such phrases by being preceded by *si* (in this context functioning only as a non-case-marked personal noun specifier) or by the 'plural' personal noun specifier *de*. In the latter case, the included noun phrase is not marked with *et*, but has developed as an associative plural construction, to be discussed in the next section.

[10] Southwest Palawano data have been generously supplied by Bill Davis, New Tribes Mission, Palawan.

3 Associative nominal constructions

We noted above that a phrasal pronominal inclusory construction, as defined by Lichtenberk, consists of a single NP containing an inclusory pronominal and an included noun, either with or without an overt marker, making it respectively either implicit or explicit.

In Philippine languages, although such constructions occur (illustrated by the Tboli and Southwest Palawano examples in (24) and (25)), one commonly finds a different type of inclusive construction, one in which the forms that introduce the 'included noun' are *not* pronouns, although in many languages their forms are identical to that of (third person plural) pronouns. This type of construction corresponds to that described by Corbett and Mithun (1996:1) as an 'associative plural construction', defined by them as consisting of 'a nominal plus a marker, and denot[ing] a set comprised of the referent of the nominal (the main member) plus one or more associated members …'.

These are exemplified by the Tagalog constructions in (7) repeated here as (26). Note that the form that introduces the construction, shown in bold in the examples, are *not* third person plural pronouns in Tagalog. These constructions, although no longer pronominal, are inclusive, in that the nominal denotes only one member, the most salient, of the set represented by the whole NP. They are ASSOCIATIVE NOMINAL CONSTRUCTIONS. The form that introduces such a construction is an INCLUSORY SPECIFIER and the nominal that represents the set is an ASSOCIATED NOUN.

(26) Tagalog

a. *Nakita =ko **sina** Juan.*
saw =GEN.1SG PERS.PL Juan
'I saw Juan (and others).'

b. *Ginawa **nina** Juan ang trabaho.*
made GEN.PERS.PL Juan NS.COM work
'Juan (and others) did the work.'

c. *Ibibigay =ko ito **kina** Juan.*
will.give =GEN.1SG this LOC.PERS.PL Juan
'I'll give this to Juan (and others).'

3.1 Marking of associative nominal constructions

It is a well-known and often discussed feature of Philippine languages (Reid 2002, 2006a, 2006b; Reid and Liao 2004; Blust 2005) that NPs are typically introduced by one or more commonly monosyllabic forms that mark the case of the NP and/or the semantic features of the lexical noun that follows. The major semantic distinction that is marked is that of common vs personal, although other distinctions are commonly marked, depending on the language, such as singular vs plural, definite vs indefinite, etc. Common NPs are typically unmarked for plurality, and can be interpreted as either singular or plural, depending on the context and sometimes on the form of the lexical noun. In many languages, especially in the Central and Southern Philippine subgroups, they can be made explicitly plural by the addition of an independent morpheme /maŋa/ (often represented orthographically as *mga*), immediately before the lexical noun, as in the Sorsoganon example (21c), repeated here as (27). In (28), the Tagalog plural NP *mga Santos* is not treated as a personal noun, but as a family name, i.e., several people with the name *Santos* (Schachter and Otanes 1972:112).

(27) Sorsoganon

Kinuwa	*namon*	*kanina*	*san*	***mga***	*batit*	*an*
took	GEN.1PL.EXCL	earlier	GEN.COM.DEF	PL	child	NS.COM

mga	*burak.*
PL	flower

'We, the children and I, took the flowers earlier today.'

(28) Tagalog

Kamag-anak	*siya*	*ang*	***mga***	*Santos.*
relative	3SG	COM.DEF	PL	Santos

'He's a relative of the Santoses.'

Personal noun phrases on the other hand are typically said to be marked for plurality, with one form appearing before singular personal nouns (such as Tagalog *si*, *ni*, and *kay*, respectively 'unmarked', 'genitive' and 'locative'), and another, usually described in the literature as 'plural', appearing before personal nouns (such as Tagalog *sina*, *nina*, and *kina*, as in (26)). The lexical item immediately following such 'plural' forms, however, is not itself plural, but is an associated noun, representing a group. Thus in (29), the sequence *cha Pakoran* does not refer to more than one individual with the name 'Pakoran', but to a set of individuals of whom 'Pakoran' is the most salient member and with whom they are in some way associated, in the same way as the set specified by an inclusory pronoun can be represented by a single named individual.

(29) Khinina-ang Bontok

As	*omey*	***cha***	*Pakoran*	*si*	*wakas.*
FUT	go	PERS.PL	Pakoran	OBL.FUT	tomorrow

'Pakoran (and others) will go tomorrow.'

In what way, then, do ANCs differ from IPCs?

IPCs, such as those illustrated in (24) for Tboli, can have first, second, or third person inclusory pronouns (depending on the language), whereas ANCs are introduced by inclusory forms that are differentiated only by case, and are usually relatable to a third person plural pronoun.

(1) In IPCs, the pronoun can commute with other pronouns, but the inclusory specifier that introduces an ANC, can commute only with a singular nominal specifier.

(2) IPCs depend for their inclusory interpretation on the obligatory presence of both a plural pronoun and an included NP. Without the included NP, the pronoun can only be interpreted as plural. ANCs on the other hand require only an associated noun (with its inclusory specifier). They can optionally be expanded by an included NP in some languages (see §3.2.1).

(3) In Ilokano, ANCs which are morphologically unmarked for case are introduced by the inclusory specifier *da*, while locatively marked ANCs are introduced by *kada*. These commute with singular personal noun specifiers, respectively *ni* and *kenni*. Compare (30a) with (30b). But note that the inclusory specifier *da* is homophonous with the Ilokano third person plural clitic pronoun *=da* that commutes with other pronouns, such as the nominatively marked *=ak* 'first person singular pronoun'. Compare (30c) with (30d).

(30) Ilokano

a. *Napan* ***ni*** *Juan* ***kenni*** *Jose.*
went NS.PERS.SG Juan LOC.PERS.SG Jose
'Juan went to Jose's place.'

b. *Napan* ***da*** *Juan* ***kada*** *Jose.*
went PERS.PL Juan LOC.PERS.PL Jose
'Juan (and others) went to Jose's (family's) place.'

c. *Napan* ***=da*** *idi* *kalman.*
went =3PL LOC.PAST yesterday
'They went yesterday.'

d. *Napan* ***=ak*** *idi* *kalman.*
went =NOM.1SG LOC.PAST yesterday
'I went yesterday.'

Further evidence that the Ilokano inclusory specifier *da* is (no longer) a third person pronoun (and that the construction it introduces is not a phrasal IPC), is its behavior when the future clitic *=(n)to* occurs. When this form occurs in combination with a clitic pronoun, the future clitic obligatorily follows the pronoun, as in (31a). However, when it occurs in a sentence which contains an INC, the future clitic occurs immediately following the verb, as in (31b).[11] Similarly when the reportative adverb *kano* occurs, it follows pronominal clitics, but it precedes an NP, as in (32a-b).

(31) Ilokano

a. *Mapan* *=da* ***=nto*** *kenni* *Jose.*
go =3PL =FUT LOC.PERS.SG Jose
'They will go to Jose's place.'

b. *Mapan* ***=to*** *da* *Juan* *kada* *Jose.*
go =FUT PERS.PL Juan LOC.PERS.PL Jose
'Juan (and others) will go to Jose's (family's) place.'

(32) Ilokano

a. *Mapan* *=da* ***kano*** *kenni* *Jose.*
go =3PL RPRT LOC.PERS.SG Jose
'They will go to Jose's place, it is said.'

b. *Mapan* ***kano*** *da* *Juan* *kada* *Jose.*
go RPRT PERS.PL Juan LOC.PERS.PL Jose
'Juan (and others) will go to Jose's (family's) place, it is said.'

3.2 Expansions of associative nominal constructions

Philippine languages differ in how ANCs can be expanded. Further specification of the membership of the set represented by the associated noun can either be by an included noun phrase of the same type as discussed above (§3.2.1) or by a coordinate noun phrase (§3.2.2).

---

[11] This clitic occurs as *=nto* following vowel-final forms, but as *=to* following consonant-final forms.

3.3 Associative nominal constructions with included noun phrase expansions

Languages, such as Ilokano and Khinina-ang Bontok, which do *not* allow third person IPCs (see §2.1 above), utilize ANCs in their place. Such languages allow an oblique noun phrase to follow the associated noun, to further specify the participants in the general set named in the construction. Thus while (33a-b) and (34a-b) are ungrammatical with the senses shown, (33a´-b´) and (34a´-b´) are grammatical. These oblique NPs function just like the included NPs discussed above with reference to IPCs. They are 'explicit' in that they carry case-marking, but they cannot be preceded by a coordinating conjunction.

(33) Ilokano

a. *Mapan* [*=da*] [*ken Jose*].
go =3PL OBL.PERS Jose
*'He/She and Jose are going.'

a´. *Mapan* [*=da* [*Maria ken Jose*].
go =PERS.PL Maria OBL.PERS Jose
'Maria and Jose are going.'

b. *Mapan* [*=da*] [*iti gayyem =na*]].
go =3PL OBL.COM friend =GEN.3SG
*'He/She and his/her friend are going.'

b´. *Mapan* [*=da* [*Maria iti gayyem =na*]].
go =PERS.PL Maria OBL.COM friend GEN.3SG
'Maria and her friend are going.'

(34) Khinina-ang Bontok

a. *Inilak* [*cha-icha*] [***an*** *Pakoran*].
saw.1SG 3PL OBL.PERS.SG Pakoran
*'I saw him/her and Pakoran.'

a´. *Inilak* [*cha* [*Takcheg* ***an*** *Pakoran*]].
saw.1SG PERS.PL Takcheg OBL.PERS.SG Pakoran
'I saw Takcheg and Pakoran.'

b. *Khina-eb* [*=cha*] [***an*** *Pakoran*] *nan afong.*
made =GEN.3PL OBL.PERS.SG Pakoran NS.COM house
*'He/She, and Pakoran, built the house.'

b´. *Khina-eb* [*cha Takcheg* [***an*** *Pakoran*]] *nan afong.*
made PERS.PL OBL.PERS.SG OBL.PERS.SG Pakoran NS.COM house
'Takcheg and Pakoran built the house.'

Just as included NPs in (split-phrasal) IPCs are separable from their pronominal head (§2.3), included NPs in ANCs are also separable from their associated noun head. In (35a), for example, the construction is split with a temporal adverb. However, while an inclusory pronoun can be separated from the rest of the construction by fronting to a position between an auxiliary verb and a main lexical verb, this is not possible for an associated noun (compare (35b) with (35c)), nor for the inclusory specifier that precedes it (35d), since there is no pronoun involved.

(35) Khinina-ang Bontok

a. *Inilak* [*cha Takcheg*] *ad khanad* [***an*** *Pakoran*].
saw.1SG PERS.PL Takcheg LOC.PAST earlier OBL.PERS.SG Pakoran
'I saw Takcheg and Pakoran earlier today.'

b. *Achi khina-eb* [*cha* [*Takcheg* ***an*** *Pakoran*]]
NEG made PERS.PL Takcheg OBL.PERS.SG Pakoran
nan afong.
NS.COM house
'Takcheg and Pakoran didn't build the house.'

c. **Achi* [*cha Takcheg*] *khina-eb* [***an*** *Pakoran*] *nan afong*
NEG PERS.PL Takcheg made OBL.PERS.SG Pakoran NS.COM house
*'Takcheg and Pakoran didn't build the house.'

d. **Achi cha khina-eb Takcheg* [***an*** *Pakoran*] *nan afong.*
NEG PERS.PL made Takcheg OBL.PERS.SG Pakoran NS.COM house
*'Takcheg and Pakoran didn't build the house.'

3.4 Associative nominal constructions with coordinate noun phrase expansions

Languages, such as Tagalog, and Sorsoganon, which *do* allow third person IPCs (see §2.1 above), can further specify the participants in the general set named in the ANC only by using a coordinate construction, as in (36a-c). In Tagalog, the coordinate construction is within the scope of the inclusory specifier, in that coordinated personal nouns are not preceded by any case-marking form, unlike coordinated NPs following a third person pronoun, as in (36d). In Southwest Palawano, a coordinated personal noun in an NP that is functioning as a coordinated expander of an inclusive non-pronominal construction, is preceded by *si*, regardless of the case of the construction, as in (37).

(36) Tagalog

a. *Nakita ko* [***sina*** [*Juan at Ben*]].
saw GEN.1SG PERS.PL Juan and Ben
'I saw Juan and Ben.'[12]

b. *Ginawa* [***nina*** [*Juan at Maria*]] *ang trabaho.*
made GEN.PERS.PL Juan and Maria NS.COM work
'Juan and Maria did the work.'

c. *Ibibigay =ko ito* [***kina*** [*Juan at Rosa*]].
will.give =GEN.1SG this LOC.PERS.PL Juan and Rosa
'I'll give this to Juan and Rosa.'

d. *Nakita =ko* [***siya***] *at* [*si Juan*].
saw =GEN.1SG 3SG and PERS.SG Juan
'I saw him/her and Juan.'

[12] There appears to be no semantic difference between a construction such as this and one using simple coordination, such as Tagalog *Nakita ko si Juan at si Ben*.

(37) Southwest Palawano

Minuli	*kay,*	*telo*	*kay*	[***de***	*Arlyn*]	*bo*	[*si*	*Abil.*]
went.home	1PL.EXCL	three	1PL.EXCL	PERS.PL	Arlyn	and	PERS.SG	Abil

'We went home, three of us, Arlyn, Abil and me.'

3.5 Summary

The following general statements have been made about inclusory constructions in Philippine languages.

IPCs consist of two parts, an inclusory plural pronominal head and a dependent included NP. Some languages allow first, second and third person pronoun inclusory pronouns, others allow only first and second person inclusory pronouns.

Included NPs in most languages are separable from the inclusory pronoun, while in a few languages they are not.

Included NPs are typically case-marked as either genitive or oblique.

In addition to IPCs, languages typically allow ANCs in which a personal noun (the associated noun) is preceded by a 'plural'-marking form, the inclusory specifier, and represents a set of individuals.

In languages in which ANCs are the only way to specify a third person inclusive set, the associated noun can optionally be expanded by a (separable) included NP, case-marked as either genitive or oblique.

In languages that allow third person IPCs, further specification of the members of the set (beyond the obligatory included NP), can only be accomplished by adding a coordinated NP.

In the next section, the historical relationship between the two types of inclusory constructions will be examined, and a proposal outlined which accounts for their development.

4 The historical development of inclusory constructions

The fact that the inclusory specifiers of ANCs and third person plural pronouns in Philippine languages apparently have a common etymological source has already been noted in the literature (Reid and Liao 2004; Blust 2005; Reid 2007). Blust, for example, in his discussion of the reconstruction of PAN genitive personal noun marking (2005:219), notes that Ivatan **da* 'genitive of plural personal nouns' probably reflects *-da '3PL genitive pronoun' and that the data in Yamada and Tsuchida (1975) 'show a similar innovation in a number of the languages of the northern and central Philippines. In some of these, the derivation from a 3PL personal pronoun is transparent, as with Itbayat *sira Pedro* 'Pedro and others'' (Blust 2005:219). In fact, however, plural personal noun marking in languages from all areas of the Philippines, and in all subgroups can be shown to be relatable to reconstructed forms of third person plural pronouns, including a few of the Central Philippine languages in which the relationship is not transparent (Reid 2007).

The nature of the innovation which resulted in the inclusory specifiers was not made explicit by Blust, but he implies that third person plural pronouns replaced the reflexes of his PAN/PMP *na 'genitive plural personal noun marker' wherever the latter form did not appear. Reid (2007) argues against this position, and presents a scenario by which the inclusory specifiers developed, as shown in Figure 1, plus a suggested development of the Tagalog third person pronouns whereby *sila* and *nila* became respectively the inclusory specifiers *sina* and *nina* (see §4.1.4 below).

Innovation (1): Deletion of *si/*ni before NP

Innovation (2): 3PL PRON becomes plural NP marker

Innovation (3): Loss of unstressed initial syllable

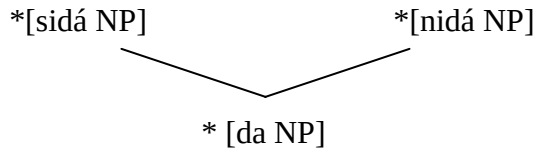


Figure 1: Innovations (from Reid (2007))

The set of changes shown in Figure 1 merely describes what appears to have taken place, that is, a third person pronoun was originally followed by an appositive NP, making explicit one of the salient individuals included in the pronominal reference. This was replaced by a single NP with loss of the redundant case-markers *si and *ni. However it does not explain the factors that brought about the shift shown as 'Innovation 2: 3PL PRON becomes plural NP marker'. In this section I claim that languages throughout the Philippines have developed inclusory specifiers (and ANCs) as a result of the grammaticalization of the IPCs (Hopper and Traugott 2003 [1993]).

4.1 Grammaticalization of pronouns

Grammaticalization, as characterized by Wanner (2006), is the holistic process by which 'originally independent elements of syntax lose autonomy and become progressively "morphologized" in their diachrony, i.e. progressively less autonomous in their syntactic freedom of occurrence, semantic referentiality, and dynamic charge' (2006:54).

The diachronic changes by which deictic forms in Latin gradually become simple pronouns, then prosodically, phonologically and syntactically reduced pronouns, and eventually (morpho)syntactically regulated clitic pronouns in Romance languages is given by Wanner (2006:54) as an example of one of the commonly observed grammaticalization clines in language. The shift from clitic pronouns in some Northern Philippine languages to person agreement markers on verbs has been described in (Reid 2001). The shift of third person plural pronouns to nominal plural markers has been described for several languages (Heine and Kuteva 2002:237–238), and is not uncommon in Philippine languages, as in Ibaloy (38) in which any noun, common or personal, can be pluralized by *ira*, originally a third person plural pronoun.

(38) Ibaloy (Ruffolo 2005:191)

a. *Sama* ***ira*** *diyang ket si'kato =y bekaan =cha nontan.*
TOP PL cave TOPLK 3SG =NS bury =3PL time.past
'As for the caves, that is where they buried (the dead) back then.'

But the shift of third person plural pronouns to inclusory specifiers has not been noted before as an instance of the grammaticalization of pronouns.

In her discussion of grammaticalization theory, Fischer (2007:115–124) discusses three of the stages in the diachronic process of grammaticalization first given in Lehmann (1985:306). These are 'weight', 'cohesion' and 'variability', each of which has paradigmatic and syntagmatic parameters, varying according to the degree of grammaticalization that has taken place. Fischer also discusses three of Hopper's principles that are also relevant to the discussion at hand. These are 'layering', 'divergence' and 'persistence' (Hopper 1991). In the discussion in the following sections each of these factors will be referred to. Thus the factor of 'variability' is relevant in the discussion of the shift from pronoun to inclusory specifier in that this results in forms that are less variable than pronouns in their syntagmatic privileges of occurrence. Inclusory specifiers are restricted to a fixed position immediately before the main lexical noun of a noun phrase, the associated noun, and commute only with singular personal noun markers, a syntagmatic position that was already present prior to grammaticalization and could have provided the 'syntactic priming' for the change (Fischer 2007:134).

There are a number of stages that can be seen in the development of ANCs, each stage sometimes co-occurring with, and not necessarily replacing, the previous stage, so that multiple ways of expressing the semantic content of inclusory constructions exist, the process referred to as 'layering' by Hopper (1991).

Languages differ according to which of the inclusive pronouns are affected, but in each situation, the end point is the same, the two parts of the construction, the inclusive pronoun and the included noun phrase become fused to form the ANC by reducing the complexity of the construction and increasing its degree of bondedness. This 'parameter' of grammaticalization will be discussed in §4.1.1. Several grammaticalization changes occurred reducing the 'weight' of the new inclusory nominal specifiers. One of these was the restriction of pronominal features, or 'semantic erosion', discussed in §4.1.2. Subsequently, a series of reductions of complex inclusory forms occurred, further reducing their 'weight' by reducing their phonemic substance, see §4.1.3. Although generally showing 'persistence' (Hopper 1991), or retention of traces of the original lexical meaning of the pronouns, the resultant forms were so underspecified that they were reformed in several languages to recover some of the lost features, §4.1.4.

4.2 Reduction of structural complexity

Constructions consisting of an independent singular pronoun followed by a coordinate construction are found in languages throughout the Philippines and represent a stage which must have been present in their (immediate) parent language.[13] The noun in such constructions probably copied the case-marking of the pronoun that preceded it, so that an unmarked pronoun would be followed by an unmarked noun, as in (39a), a genitive pronoun by a genitively-marked noun, as in (39b), and a locative pronoun by a locatively marked noun.

(39) Tagalog

a. *Nakita ko* [***siya***] *at* [***si*** *Juan*].
saw GEN.1SG 3SG and NS.PERS.SG Juan
'I saw him/her and Juan.'

[13] Proto Philippines in Blust's terms (Blust 2006), Proto Extra-Formosan in mine.

b. *Ginawa* [***niya***] *at* [***ni*** *Maria*] *ang trabaho.*
made GEN.3SG and GEN.PERS.SG Maria NS.COM work
'He and Maria did the work.'

c. *Ibibigay ko ito* [***sa kaniya***] *at* [***kay*** ***Juan***].
will.give GEN.1SG this LOC OBL.3SG and LOC.PERS.SG Juan
'I'll give this to him/her and Juan.'

Phrases with plural pronouns in which one of the salient members of the group specified by the pronoun was expressed by a concatenated appositive noun phrase (without a coordinating conjunction) must also have occurred, functioning as an included noun phrase. The noun in such constructions also probably copied the case-marking of the pronoun that preceded it. This is displayed in Table 1 for Proto Central Philippines (PCPH).[14]

Table 1: Concatenated included noun phrases in Proto Central Philippines

	Nominative	Genitive
PCPH	*[sidá] [si N]	*[nidá] [ni N]

Evidence for constructions of the type shown in Table 1 is found in several of the Central Philippine languages, such as the Bisayan languages Cebuano and Jaun-jaun, and Bikol as spoken in Iriga, as shown in Table 2 (McFarland 1974; Zorc 1977). In these languages, however, the (singular) personal noun specifier has become encliticized to the preceding pronoun and there is a reduction of structural complexity. The sequence is no longer pronominal but marks the following noun as the associated noun of an ANC.[15]

Table 2: Associative nominal constructions in some Central Philippine languages

	Nominative	Genitive
Stage 1	*[sída-si N]	*[nída-ni N]
Ceb, Jau	[síla-si N]	[níla-ni N]
Iriga	[sira-si N]	[nira-ni N]

In some languages, however, before the structural complexity was reduced, genitively marked concatenated noun phrases (*ni N) were interpreted as oblique comitative NPs (also marked with *ni), i.e., what would earlier have been understood, for example, as 'they, John' was reanalyzed as 'they with John'. This reanalysis subsequently spread to replace concatenated noun phrases in other positions in the sentence as well, as shown in Table 3 (with *ni*-marked phrases in both nominative and genitive positions), eventually giving rise to the Tagalog and Sorsoganon third person inclusory pronominal structures,

[14] In order to keep this paper within reasonable bounds, the discussion in the rest of this section will be restricted to the developments which have affected NPs only in nominative and genitive positions of a sentence. Changes affecting locatively marked NPs are not included.

[15] Zorc (1977:82) hyphenates the two parts of the form (as in Table 2) and refers to the combined form as a 'plural personal-name marker'. McFarland (1974:156) does not hyphenate the forms, but refers to the sequence as a 'plural PNE [personal noun expression] marker'.

illustrated in (5)–(6) above. Hiligaynon, like the Central Philippine languages in Table 2, then reduced the complexity of the construction producing ANCs in which the form that marks the following associated noun as inclusory shows an encliticized reflex of the oblique singular specifier *ni, see Table 4 (Zorc 1977:82).

Table 3: Third person pronouns with concatenated comitative constructions

	Nominative	Genitive
PCPH	*[sidá] [**ni N**]	*[nidá] [**ni N**]

Table 4: Associative nominal constructions in Hiligaynon

	Nominative	Genitive
PRE-HIL	*[sidá-ni N]	*[nidá-ni N]
HIL	[silá-ni N]	[nilá-ni N]

4.3 Restriction of pronominal features

The comitative included constructions, illustrated in Table 3, were not restricted to those in which a third person plural pronoun was the head. Any plural pronoun could commonly occur in that position. Prior to the development of inclusive non-pronominal constructions, I assume that the semantic features of the pronouns were not restricted, and that even with a following comitative phrase specifying one of the members of the group, the pronoun expressed all the semantic features (±SPKR, ±ADDR, ±PL) appropriate to its form. Constructions of this type are commonly found, not only in Central Philippines languages, but in Northern Philippine languages as well, as in (40)–(42), although in some languages, they are ambiguous and can also be interpreted as included pronominal constructions.

(40) Ilokano

a. *Napan* [=*kami*] [***ken*** *Marta*].
went =NOM.1PL.EXCL OBL.PERS.SG Martha
'We (EXCL) went with Martha.'

b. *Dinalosan* [=*tayo*] [***ken*** *Maria*] *ti* *balay* =*da*.
cleaned =1PL.INCL OBL.PERS.SG Maria NS.COM house =3PL
'We (INCL) cleaned their house with Maria.'

(41) Tagalog

Maglalakad [*kayo*] [***ng*** *Nanay*].
will.walk 2PL GEN.COM.SG mother
'You (PL) will walk with Mother.'

(42) Khinina-ang Bontok

Inilak [*cha-icha*] [***an*** *Pakoran*].
saw.1SG 3PL OBL.PERS.SG Pakoran
'I saw them with Pakoran.'

Restriction of pronominal features probably started in first person exclusive pronouns, since most Philippine languages have included pronominal constructions headed by this pronoun. The features of a first person exclusive pronoun without a following included noun phrase are as follows: +SPKR, –ADDR, +PL (i.e., 'we EXCL/us EXCL/our EXCL'). When the pronoun occurs as head of an IPC, however, the plurality feature is missing, and the pronoun is interpreted only as +SPKR, –ADDR ('I/me/my'). The included noun phrase supplies the additional features to enable the pronoun to be understood as plural.

That this is an on-going change, spreading from first person exclusive pronouns, to first person inclusive, second and third person plural forms (although not necessarily in that order) is clear from the various ways such pronouns are interpreted when followed by a genitive or oblique noun phrase.

A second person plural pronoun without a following included noun phrase has the features: +SPKR, +ADDR, +PL (i.e., 'you PL/your PL'). When the pronoun occurs as head of an IPC, however, the plurality feature is missing, and the pronoun is interpreted only as +SPKR, +ADDR ('you/your'). The included noun phrase supplies the additional features to enable the pronoun to be understood as plural, as in (43) (compare (41)).

(43) Tagalog

Maglalakad [*kayo*] [***ng*** *Nanay*].
will.walk 2PL GEN.COM.SG mother
'You (SG) will walk with Mother.'

Similarly, a third person plural pronoun without a following included noun phrase has the features: –SPKR, –ADDR, +PL (i.e., 'they, them, their'). When the pronoun occurs as head of an IPC, however, the plurality feature is missing, and the pronoun is interpreted only as –SPKR, –ADDR ('he/him, she/her'). The included noun phrase supplies the additional features to enable the pronoun to be understood as plural, as in (44).

(44) Tagalog

Nakita ko sila ***ni*** *Juan.*
saw GEN.1SG 3PL GEN.PERS.SG Juan
'I saw him/her and Juan.'/*'I saw them with/and Juan./*'I saw Juan (and others).'

While Tagalog allows both first person inclusive and second person plural pronouns to occur either with an optional comitative noun phrase (in which case the pronoun carries all its features) or as the heads of IPCs, resulting in the ambiguities in interpretation of the pronoun, Tagalog only allows third person plural pronouns with a following genitive noun phrase to be interpreted as ANCs.

4.4 Reduction of complex plural-marking forms

To express the unallowed meanings of (44), Tagalog uses an ANC, exemplified in (7)a, repeated here as (45).

(45) Tagalog

Nakita =ko ***sina*** *Juan.*
saw =GEN.1SG NS.PERS.PL Juan
'I saw Juan (and others).'

The nominative and genitive inclusory specifiers in Tagalog are respectively *sina* and *nina*. In Reid (2007), I proposed that these forms developed directly from the Tagalog third person plural pronouns, *sila* and *nila*, by an irregular, but not uncommon sound change, *l > *n*. Further examination of the evidence suggests, however, that Tagalog underwent a reduction in structural complexity, like Hiligaynon (Table 4), but then assimilated *l in the genitive form to the nasals occurring on either side of it. The nominative form was then changed by an analogical process to match the genitive form. The complexity of the form was then reduced by loss of the final redundant syllable, as shown in Table 5.

Table 5: The development of Tagalog non-pronominal included constructions

	Nominative	Genitive	
PCPH	*[sidá] [ni N]	*[nidá] [ni N]	
PRE-TAG 1	*[sidá-ni N]	*[nidá-ni N]	Reduction in structural complexity
PRE-TAG 2	*[silá-ni N]	*[ni**l**á-ni N]	*-d- > *-l-
PRE-TAG 3	*[silá-ni N]	*[ni**n**á-ni N]	Assimilation *-l- > *-n- in Genitive
PRE-TAG 4	*[si**n**á-ni N]	*[ni**n**á-ni N]	Analogical spread, Gen to Nom
TAG	[siná N]	[niná N]	Loss of redundant syllable

This sequence of developments was not unique to Tagalog. There are a number of other Central Philippine languages that have identical forms (see Zorc (1977:82)), or forms that probably developed from them, such as Mamanwa (*sin/nin*), and the Bisayan languages Bantoanon and Sibale (*sa/na*). In Reid (2007), I suggested one possible way in which the latter forms developed. I now consider that *nina was reduced to *na* by loss of the first syllable (the initial *n* of the resulting monosyllable marking the form as genitive), and that the nominative form developed by analogy to the genitive form, with *s* marking the form as nominative, as in Table 6.

Table 6: The development of monosyllabic specifiers in some Bisayan languages

	Nominative	Genitive	
PRE-BAN/SIB 1	*[siná N]	*[niná N]	
PRE-BAN/SIB 2	*[siná N]	*[na N]	Loss of redundant syllable in Genitive
BAN/SIB	[sa N]	[na N]	Analogical spread, Gen to Nom

Other Bisayan languages reduced the complexity of their plural marking forms by deleting the final, redundant unstressed syllable that originally marked the included noun phrase, reducing the trisyllabic specifier to disyllabic, as in Table 7.

Table 7: Reduction of trisyllabic to disyllabic specifiers in some Bisayan languages

	Nominative	Genitive	
PCPH	*[sidá] [ni N]	*[nidá] [ni N]	
PRE-B,B,S,W 1[16]	*[sidá-ni N]	*[nidá-ni N]	Reduction in structural complexity
PRE-B,B,S,W 2	*[silá-ni N]	*[nilá-ni N]	*-d- > *-l-
BUT, BOH, SUR	[silá N]	[nilá N]	Loss of redundant syllable
WAR	[híra N]	[níra N]	

4.5 Reformation of plural-marking forms

The changes shown in Table 6 resulted in inclusory specifiers that were no longer phonologically similar to third person plural pronouns, and were homophonous either with the locative preposition *sa* marking future time phrases in all the Bisayan languages, or the genitive preposition *na*, marking genitive definite common nouns in some of the Bisayan languages, neither of which was specific for personal noun or plurality. These were the conditions that probably motivated the reformation of the inclusory specifiers in a considerable number of dialects, such as Kuyonon and Datagnon, by encliticizing a (genitive) third person plural pronoun to each of the ambiguous monosyllables, and reduction of the trisyllabic specifiers to disyllabic by medial vowel syncope, as in Table 8.

Table 8: The development of non-pronominal included constructions with reformed disyllabic specifiers

	Nominative	Genitive	
PRE-KUY/DTG 1	*[sa N]	*[na N]	
PRE-KUY/DTG 2	*[sa=nidá N]	*[na=nidá N]	Encliticization of 3PL pronoun
KUY/DTG	[sánda N]	[nánda N]	Medial vowel syncope

4.6 The development of Ilokano inclusory constructions

In Ilokano, 'layering' of constructions is prevalent where grammaticalization has introduced new construction types but their source construction types remain available for use, resulting in multiple ambiguous constructions. Ilokano does not allow (or has not developed) IPCs with third person pronouns. Where plural pronouns occur followed by oblique noun phrases, these are potentially interpretable (especially when the pronoun is third person) as full pronouns followed by a location, as in (46).[17]

---

[16] PRE-BUT, BOH, SUR, WAR.

[17] In these and the following sets of examples, only first person exclusive and third person plural pronouns are illustrated. The obliquely marked phrases which follow the pronouns are in all cases shown only as singular (*ken/kenni* 'personal noun' or *ti/iti* 'common noun'), although they may also be marked as plural with case-marked specifiers grammaticalized from plural pronouns (*kada* 'plural personal noun' or *kadagiti* 'plural common noun'). Ilokano *kenni* has its ultimate source in Proto Extra-Formosan *ka 'oblique preposition' followed by a genitive personal singular specifier *ni. Change of the first vowel to schwa resulted in regular gemination of the following consonant, thus *ka=ni > *kenni*.

(46) Ilokano

a. *Mapan=kami **kenni** Jose.* ‘We (EXCL) are going to Jose’s place.’

b. *Mapan=kami **iti** gayyem=na.* ‘We (EXCL) are going to his friend’s place’

c. *Mapan=da **kenni** Jose* ‘They are going to Jose’s place.’

d. *Mapan=da **iti** gayyem=na.* ‘They are going to his friend’s place.’

Oblique noun phrases in Ilokano are also interpretable as comitative, but the marking of personal oblique noun phrases (*ken*) has extended its function to that of conjunction, not only of personal but also of common noun phrases. It has also become the coordinate conjunction for all other word and construction types (as in (47a-b)).

(47) Ilokano (Rubino 1997)

a. *Nalinis **ken** nalawa ti kuarto =na.*
clean CONJ spacious COM.SG room =GEN.3SG
‘His room is clean and spacious.’

b. *Silulukat ti barukong **ken** takiag =da nga umawat*
open COM.SG chest CONJ arm =GEN.3PL LG receive
***ken** sumarabo kenka.*
CONJ welcome OBL.2SG
‘Their chest and arms are open to receive and welcome you.’

The result is that full pronouns followed by an oblique noun phrase have either conjoined or comitative readings, as in (48). In constructions of this type the occurrence of the (singular) person marker *ni* is optional. Oblique common noun phrases are marked with *ti* in casual speech, and with *iti* in more formal styles.

(48) Ilokano

a. *Mapan=kami **ken (ni)** Jose.* ‘We (EXCL) and Jose are going.’/
‘We (EXCL) are going with Jose.’

b. *Mapan=kami **ken (i)ti** gayyem=na.* ‘We (EXCL) and his friend are going.’/
‘We (EXCL) are going with his friend.’

c. *Mapan=da **ken (ni)** Jose* ‘They and Jose are going.’/
‘They are going with Jose.’

d. *Mapan=da **ken (i)ti** gayyem=na.* ‘They and his friend are going.’/
‘They are going with his friend.’

It is from constructions such as those in (48a-b) that Ilokano (non-third person) inclusive pronominal constructions have developed.

(49) Ilokano

a. *Mapan=kami **ken (ni)** Jose.* ‘Jose and I are going.’

b. *Mapan=kami **(i)ti** gayyem=na* ‘His friend and I are going.’

5 Conclusion

There is far more that could be said both about the types of inclusory constructions that occur in Philippine languages and their historical development, but data on these constructions are not typically given in detail in the available grammatical materials, and space limitations have dictated a limited discussion of the data that are available. While studies of grammaticalization phenomena in a number of European languages can be informed by a thousand years or more of written tradition, Austronesianists are primarily limited to synchronic descriptions going back at most a few centuries, and these for relatively few languages. Inferences of direction of change then can at best be drawn only tentatively, based mainly on limited textual evidence. This paper is a first attempt to do this for these constructions in Philippine languages, and will surely be modified as more information becomes available.

In summary then, two major types of inclusory construction have been described. The first are IPCs headed by pronouns whose reference includes the lexical form that follows the pronoun, typically introduced either by a genitive or an oblique case-marking form. The second are ANCs headed by an inclusory specifier, typically described in the literature as a plural personal noun marker, since it commutes with forms that typically specify singular personal nouns. Inclusory specifiers can in all cases be shown to have developed from third person plural pronouns. They are not, however, pronouns. They are inclusive forms, depending for their interpretation on the associative feature which is carried by the personal noun that follows them.

The grammaticalization changes that have resulted in the development of pronominal inclusory constructions include an increase in bondedness, whereby conjoined constructions become appositive, and comitative constructions become included constructions syntactically dependent on their pronominal head. Third person IPCs developed into ANCs by structural simplification, creating a further increase in bondedness, as the two parts of the original construction, while typically separable in the source construction, are no longer so in the new construction. Reduction of the semantic features of inclusory pronouns and of the phonemic weight of inclusory specifiers provide further evidence of the path of grammaticalization which resulted in each of these construction types.

References

Blake, Frank R. 1916. Construction of coordinated words in the Philippine languages. *The American Journal of Philology* 37(4):466–474.

Blust, Robert A. 1991. The Greater Central Philippine hypothesis. *Oceanic Linguistics* 30(2):73–129.

—— 2005. A note on the history of genitive marking in Austronesian languages. *Oceanic Linguistics* 44:215–222.

—— 2006. The linguistic macrohistory of the Philippines: some speculations. In Hsiu-chuan Liao and Carl R. Galvez Rubino, eds *Current issues in Philippine linguistics and anthropology: parangal kay Lawrence A. Reid*, 31–68. Manila: The Linguistics Society of the Philippines and SIL Philippines.

Corbett, Greville G. and Marianne Mithun. 1996. Associative forms in a typology of number systems: evidence from Yup'ik. *Journal of Linguistics* 32:1–17.

Daguman, Josephine S. 2004. A grammar of Northern Subanen. PhD dissertation, La Trobe University.

Davis, Bill. 1995. S.W. Palawano grammar. MS. New Tribes Mission, Palawan.

Dixon, R.M.W. and Alexandra Y. Aikhenvald. 2000. Introduction. In R.M.W. Dixon and Alexandra Y. Aikhenvald, eds *Changing valency: case studies in transitivity*, 1–29. Cambridge and New York: Cambridge University Press.

Fischer, Olga. 2007. *Morphosyntactic change: functional and formal perspectives*. Oxford: Oxford University Press.

Forsberg, Vivian M. 1992. A pedagogical grammar of Tboli. *Studies in Philippine Linguistics* 9(1):1–110.

Heine, Bernd and Tania Kuteva. 2002. *World lexicon of grammaticalization*. Cambridge: Cambridge University Press.

Hopper, Paul J. 1991. On some principles of grammaticalization. In Elizabeth C. Traugott and B. Heine, eds *Approaches to grammaticalization* vol. 1, 17–35. Amsterdam: Benjamins.

Hopper, Paul J. and Elizabeth. C. Traugott. 2003 [1993]. *Grammaticalization*. Cambridge: Cambridge University Press.

Lehmann, C. 1985. Grammaticalization: synchronic variation and diachronic change. *Lingua e Stile* 20:303–318.

Lichtenberk, Frantisek. 2000. Inclusory pronominals. *Oceanic Linguistics* 39(1):1–32.

McFarland, Curtis D. 1974. The dialects of the Bikol area. PhD dissertation, Yale University.

Porter, Doris. 1966. Tagabili pronouns. MS. Library of the Summer Institute of Linguistics, Manila.

Reid, Lawrence A. 2001. On the development of agreement markers in some Northern Philippine languages. In Joel Bradshaw and Kenneth L. Rehg, eds *Issues in Austronesian morphology: a focusschrift for Byron W. Bender*, 235–257. Canberra: Pacific Linguistics.

—— 2002. Determiners, nouns or what? Problems in the analysis of some commonly occurring forms in Philippine languages. *Oceanic Linguistics* 41(2):295–309.

—— 2006a. On the origin of Philippine vowel grades. *Oceanic Linguistics* 45(2):457–472.

—— 2006b. On reconstructing the morphosyntax of Proto-Northern Luzon. *Philippine Journal of Linguistics* 37(2):1–64.

—— 2007. Another look at the marking of plural personal noun constructions in Austronesian languages. *Oceanic Linguistics* 46(1):232–252.

Reid, Lawrence A. and Hsiu-Chuan Liao. 2004. A brief syntactic typology of Philippine languages. *Language and Linguistics* 5(2):433–490.

Ruffolo, Roberta. 2005. Topics in the morpho-syntax of Ibaloy, Northern Philippines. PhD dissertation, The Australian National University.

Schachter, Paul and Fe T. Otanes. 1972. *Tagalog reference grammar*. Berkeley, Los Angeles, London: University of California Press.

Svelmoe, Gordon and Thelma Svelmoe. 1974. *Notes on Mansaka grammar,* Language Data: Asian-Pacific Series No.6. Huntington Beach: Summer Institute of Linguistics.

Wanner, Dieter. 2006. *The power of analogy*. Trends in Linguistics, Studies and Monographs 170. Berlin, New York: Mouton de Gruyter.

Wolfenden, Elmer P. 2001. *A Masbatenyo–English dictionary,* Linguistic Society of the Philippines Special Monograph Issue Number 38. Manila: Linguistic Society of the Philippines.

Wrigglesworth, Hazel. 1971. Discourse and paragraph structure of Ilianen Manobo. In Robert E. Longacre, ed. *Philippine discourse and paragraph studies in honor of Betty Mclachlin,* 85–194. Canberra: The Australian National University.

Yamada, Yukihiro and Shigeru Tsuchida. 1975. *Philippine languages,* Asian and African Grammatical Manual No.15b. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies.

Zorc, R. David Paul. 1977. *The Bisayan dialects of the Philippines: subgrouping and reconstruction*. Canberra: Pacific Linguistics.

19 *Proto Austronesian verbal morphology: a reappraisal*

MALCOLM ROSS

1 Introduction

In this paper I suggest that the system of verbal morphology hitherto reconstructed for Proto Austronesian (PAn) did not yet exist in PAn. Instead, the PAn system more closely resembled the pre-PAn system reconstructed by Ross (1995:749, 2002:40). Evidence in support of this suggestion is drawn mainly from the Formosan language Puyuma (Teng 2008a), which reflects the alleged pre-PAn system rather than the system previously reconstructed for PAn. Additional support is found in Tsou and Rukai, two other Formosan languages whose verbalsystems are more readily derived from the pre-PAn system than the PAn system.

A corollary of demoting the reconstructed PAn system to a lower node in the Austronesian tree is that the languages that reflect it belong to a subgroup which excludes Puyuma, Tsou and Rukai. This subgroup, which I dub 'Nuclear Austronesian', includes all other Austronesian languages. That is, I claim (somewhat tentatively) that PAn underwent a primary four-way split into Puyuma, Tsou, Rukai and Proto Nuclear Austronesian (PNAn). This claim entails only a minor conflict with the subgrouping proposals made by Robert Blust.[1] Blust (1999) classifies the Formosan languages into nine subgroups. The proposal here calls into question one of these subgroups, Tsouic, as it treats one of its member languages, Tsou, as a single-member off-shoot of PAn but assigns the other two members, Kanakanavu and Saaroa, to Nuclear Austronesian.[2]

[1] Bob Blust played a major role in introducing me to Austronesian historical linguistics when I first visited Canberra in 1976. He has remained a source of inspiration and has become a good friend, and it is a real pleasure to write this paper in his honour.

[2] I am grateful to Stacy Fang-ching Teng for Puyuma data, to Daniel Kaufman for discussion which stimulated the writing of this paper, and to Andrew Pawley, Lawrence Reid, Stacy Teng, John Wolff and Elizabeth Zeitoun for comments on earlier drafts.

Alexander Adelaar and Andrew Pawley, eds
Austronesian historical linguistics and culture history: a festschrift for Robert Blust, 295–326.
Pacific Linguistics, 2009.
© This edition vested with Pacific Linguistics.

2 Proto Nuclear Austronesian verbal morphology

Table 1 gives an overview of PNAn verbal morphology together with the reconstructed forms of the verb **kiRim* 'seek, look for'. It resembles the PAn morphology table presented in Ross (1995:739) and reproduced with a few changes in Ross (2002b:33).

Table 1: Proto Nuclear Austronesian verbal morphology
(= Proto Austronesian verbal morphology as previously reconstructed)

	ACTOR VOICE	UNDERGOER VOICE		
		Patient subject	Location subject	Circumstance subject
Realis (V/N)	**M*-STEM **k‹um›iRim*	*STEM-*en* **kiRim-en*	*STEM-*an* **kiRim-an*	**Sa-/Si*-STEM **Sa-/Si-kiRim*
Realis perfective (V/N)	**M-‹in›*STEM **k‹um-in›iRim*	**‹in›*STEM **k‹in›iRim*	**‹in›*STEM-*an* **k‹in›iRim-an*	**‹in›Si*-STEM **Sa-/Si-kiRim*
Realis imperfective	**M-Ca*-STEM **k‹um›a-kiRim*	**Ca*-STEM-*en* **ka-kiRim-en*	**Ca*-STEM-*an* **ka-kiRim-an*	**Sa-/Si-Ca*-STEM **Sa-/Si-ka-kiRim*
Irrealis (V/N)	**Ca*-STEM **ka-kiRim*			**Ca*-STEM **ka-kiRim*
Optative/hortative	**M*-STEM-*a* **k‹um›iRim-a*	*STEM-*aw* **kiRim-aw*	*STEM-*ay* **kiRim-ay*	**an-ay* + STEM **an-ay kiRim*
Imperative	*STEM **kiRim*	*STEM-*u* **kiRim-u*	*STEM-*i* **kiRim-i*	**an-i* + STEM **an-i kiRim*
Dependent		*STEM-*a* **kiRim-a*		

Some of the differences between Table 1 and the earlier tables concern labeling and presentation. I try here to stick to terms and frameworks used by typologists. Instead of positing four voices, I follow Himmelmann (2005) in analysing what he calls 'Philippine-type' languages as having two voices, actor voice (AV) and undergoer voice (UV).[3] Actor voice is intransitive in a number of these languages, whilst UV is transitive in all of them and is usually the default choice in discourse. The grammatical roles of a Philippine-type language are thus ergatively aligned (Starosta 1999; Reid and Liao 2004).[4] As well as default patient-subject UV (henceforth UVP) verb forms, a Philippine-type language has one or two sets of applicative-like forms which promote a location (UVL) or a circumstance role (UVC: instrument, theme or beneficiary) to transitive subject (Starosta 1986, Ross and Teng 2005).[5]

[3] Philippine-type languages include the majority of languages found in Taiwan, the Philippines, northern Borneo and northern Sulawesi. See Himmelmann (2005) for a definition.

[4] Their verbal morphology is not ergatively aligned, since most intransitive verbs are marked by the same morphemes as AV (Ross and Teng 2005).

[5] Philippine-type languages have long been regarded as typologically odd, but Peterson (2007:191–193, 217–219) comments that ergatively aligned languages with applicatives which place a referent in the highly topical subject position are relatively common, at least among languages with applicatives.

In order to compare form–function pairings across languages I have adopted the set of function terms and definitions listed in Appendix A, applying these in the analysis of Formosan verbal morphologies presented in summary form in Appendix B. Sources of language materials are listed in Appendix C.

The reconstruction in Table 1 represents an abstraction at two levels. First, it is inferred from the morphologies in Appendix B. Second, these morphologies are themselves abstractions. In Formosan languages—and in PAn and PNAn—a verb has two forms ('principal parts' in the language of Latin teachers until the mid-twentieth century), neither of which is predictable from the other but from which all other forms of the verb are usually predictable. Verbs fall into five classes on the basis of these two forms, as shown in Table 2 and illustrated from Puyuma.[6] The verb **kiRim* in Table 1 belongs to Class 1.

One of the two forms is the stem, which in PAn and PNAn and in a majority of Formosan languages is the AV imperative or dependent form.[7] A PAn/PNAn simple stem consisted of either a plain root or the root prefixed by **ka*-. There was a strong tendency for verbs with stems in **ka*- to be stative (cf. L.M. Huang 2000; Zeitoun and Huang 2000). There were also stems consisting of a root with a prefix other than **ka*- or of two roots, but these are not shown in Table 2.

The second of the two forms is the AV realis (in Tsou the AV dependent). In this form the morpheme *M*-is applied to the stem. In PAn and PNAn **M*- took three forms: the infix ‹*um*›, the prefix **ma*-, or zero. Table 2 provides Puyuma examples from the five classes and illustrates how predictability works.[8] Thus the AV irrealis form **Ca*-STEM is predictable from the AV imperative (STEM) and the AV imperfective form **M*-*Ca*-STEM is predictable from the AV realis form (**M*-STEM). All other forms of the Puyuma verb can be predicted once the two basic forms are known, and the same was evidently true for all the forms of a PAn or PNAn verb.

Table 2: Proto Austronesian, Proto Nuclear Austronesian and Puyuma verb classes

Class		1	2	3	4	5
PAn, PNAn:						
AV imperative =	STEM	*ROOT	*ROOT	*ROOT	**ka*-ROOT	**ka*-ROOT
AV realis =	**M*-STEM	*‹*um*›ROOT	**ma*-ROOT	*ROOT	**ma*-ROOT	*ROOT
Puyuma:		'bathe'	'tell'	'give'	'scold'	'be hot'
AV imperative =	STEM	*dirus*	*reŋay*	*beray*	*ka-ḍeki*	*ka-bias*
AV irrealis =	*Ca*-STEM	*da-dirus*	*ra-reŋay*	*ba-beray*	*ka-ḍa-ḍeki*	*bias*
AV realis =	*M*-STEM	*d‹em›irus*	*ma-reŋay*	*beray*	*ma-ḍeki*	*ka-ba-bias*
AV realis imperf. =	*M*-*Ca*-STEM	*d‹em›a-dirus*	*ma-ra-reŋay*	*ba-beray*	*ma-ḍa-ḍeki*	*ba-bias*

6 Tsukida (2005:315) gives a similar table for Seediq.

7 In Kanakanavu it is the UVP dependent form. In Tsou, Ishbukun Bunun and Siraya it does not occur in isolation, but several affixed forms transparently reveal the stem.

8 Puyuma verb classification is more complex than is shown here, but all forms of a verb are predictable from the two basic forms.

There are a few differences between the 1995 analysis and the present one. Some of these arise from a difference in method. The earlier analysis was largely based on a comparison of verbal affixes found in language descriptions. The present analysis is based on tabulations of forms found in the different verb classes of each language and on cross-linguistic comparison not of affixes but of whole verb forms, i.e. forms like those listed in Table 2.

Zeitoun et al. (1996) show that there is a primary division in most Formosan languages between realis mood, encoding realised events and states—present, past and sometimes habitual—and irrealis mood, encoding future and otherwise unrealised events and states. This points back to a similar division in PNAn and PAn. Three sets of realis forms are reconstructed Table 1: a set unmarked for aspect and labelled 'realis' (formerly 'neutral'), a perfective aspect set encoding completed events, and an imperfective aspect set encoding incomplete, ongoing events or changes of state.

In PNAn, unmarked realis, perfective realis and irrealis forms served both as verbs and as gerundive nominalisations. This is annotated in Table 1 by '(V/N)'.

Two comments on the imperfective are pertinent. First, it was evidently marked by **Ca*- reduplication, i.e. by reduplication of the initial syllable and replacement of its vowel by *-a-*. **Ca*- is reflected as *Ca*- in Puyuma, Kanakanavu, Saaroa, Thao, Amis and Siraya but replaced by CV- in Saisiyat, Pazih, Bunun, Paiwan, Yami and Bisayan languages. Second, the imperfective contrasted with a durative (not shown in Table 1), marked by *CVCV-reduplication (*CV- with monosyllables), which apparently encoded iterativity with telic verbs and an enduring event with atelic verbs. The contrast is reflected in Kanakanavu, Saaroa, Pazih and Siraya. There is also a CVCV- durative, but no *Ca*-imperfective, in Amis, Yami and Manobo. The earlier analysis confused matters by labelling as 'durative' what is here labelled imperfective, and Ross (2002b) compounded this confusion by suggesting that both **Ca*- and *CV- were reconstructable as markers of the 'durative' (=imperfective).[9]

Reid (2007) argues on phonological grounds that **Ca*- reduplication must be derived from earlier *CV- reduplication. This is true in principle. However, the fact that **Ca*- and *CVCV- and/or *CV- reduplication are in contrast in some Formosan languages (Zeitoun and Wu 2006) supports the reconstruction of this contrast in PNAn. Further, Puyuma is an external witness to the reconstruction of PNAn **Ca*- and supports its reconstruction in PAn (see below, Table 5). I infer that PAn **Ca*- imperfective reduplication reflects a *CV-reduplication which occurred at a pre-PAn stage for which we have no witnesses, whereas PNAn *CVCV-/*CV- durative reduplication reflects a later innovation, one which took place after the earlier *CV- had become PAn **Ca*-. **Ca*-reduplication was replaced by CV- reduplication in Saisiyat, Pazih, Bunun, Paiwan and Proto Malayo-Polynesian because of its formal and functional similarity to CVCV-/CV- durative reduplication.

**Ca*-reduplication also marked the irrealis, and Ross (1995:751–752) suggested that the irrealis ('future') was simply a functional extension of the imperfective. I am now less sure of this. I tentatively reconstruct a contrast between PNAn realis imperfective AV **‹um› Ca*-STEM and irrealis AV **Ca*-STEM on the basis of Puyuma, an external witness. I also reconstruct a contrast between the corresponding PNAn UVC forms, imperfective **Sa-/Si-Ca*-STEM (Pazih *sa-Ca*-STEM, Paiwan *si*-CV-STEM) and irrealis **Ca*-STEM (reflected in

[9] Puyuma, Kanakanavu and Pazih each have an alternative pattern whereby with certain verbs (membership in the category is morphologically or lexically determined) *Ca*-is replaced by ‹*a*› infixation, either between the morphemes of a compound stem or in Pazih after the initial consonant of the stem.

Saisiyat, Atayal and Seediq). Crucially, UVC nominalisations—which reflect irrealis rather than imperfective forms—also reflect **Ca*-STEM, never **Sa-/Si-Ca*-STEM. There was, however, no contrast between UVP or UVL imperfective and irrealis forms. How this came to be is discussed in §6.

A number of Formosan languages have optative forms encoding volition or definite intention, and hortative forms encoding a command addressed to self and hearer ('Let us …' or 'Shall we …?'). The evidence suggests that a single set of forms, labelled optative/hortative (formerly 'projective') in Table 1, had both functions in PNAn. Their reflexes have both functions in Mayrinax Atayal, are optative in Paiwan and Amis, hortative in Seediq, and imperative in Kanakanavu, Saaroa, Ishbukun Bunun, and irrealis in Pazih and Siraya. Since it is fairly clear that imperative and irrealis are encoded by other forms in Table 1, it is a reasonable inference that these meanings reflect extensions in the functions of optative/hortative forms.

Among the imperative and dependent (formerly 'atemporal') forms in Table 1, only UVP forms differ from one another. The assignment of functions to forms in **-u* and **-a* is tenuous. Forms reflecting **STEM-u* occur in Puyuma (an external witness), Saaroa, Paiwan and outside Taiwan in Lun Dayeh. All are UVP imperatives, but the Paiwan form also serves as AV imperative. Forms reflecting **STEM-a* occur as UVP dependent in Tsou (an external witness) and in Yami. In the Bisayan dialects, Manobo, Timugon, Kimaragang and Eastern Kadazan, reflexes are used both as imperatives and in dependent clauses, whilst in Ishbukun Bunun, Bonggi and Lun Dayeh they apparently occur only as imperatives (in Bunun AV, Lun Dayeh UVL). The Kavalan reflex of **STEM-a* is a general UV irrealis, whilst a Javanese reflex serves as imperative and subjunctive and a Proto Malayic reflex as subjunctive (Adelaar1992:148). The hypothesis that best accounts for these data is that PNAn **STEM-u* was imperative but has been displaced in a number of languages by an extension in function of the reflex of the dependent **STEM-a*, probably because AV, UVL and UVC imperative and dependent forms were already identical.

Dependent forms were used after certain preverbs ('auxiliaries' in conventional Formosanist terminology), including negators, as they still are in a range of Formosan and Philippine languages (Ross 1995:744–747). They are also used for foreground events in narrative in Mantauran Rukai, Paiwan, Kimaragang, Eastern Kadazan and Timugon Murut, although in this function the AV form was apparently STEM (rather than **M*-STEM). This was probably an application of their use with preverbs, as certain coordinators meaning 'and then' functioned as preverbs. A sentence thus began with an independent clause with (presumably) a realis verb, followed by one or more clauses each introduced by 'and then' and having a dependent verb. This created a coordinate-dependent form of clause linkage which still occurs in Mantauran Rukai and Paiwan (A.H. Chang 2006).[10] It is not clear whether the 'and then' preverb remains obligatory in the other languages.

The UVC forms in Table 1 require comment. Relevant data are in Table 3. First, alternant forms of the realis circumstance-subject and circumstance-nominaliser prefix, **Sa-* and **Si-*, are reconstructable. No language has regular reflexes of both (the expected Kavalan reflex of **Si-* would be *si-*, not *ti-*). Notably, there are no unambiguous reflexes of **Sa-* with the perfective infix **‹in›*, but reflexes of **S‹in›i-* occur in Saisiyat and Paiwan.

[10] The term 'coordinate-dependent' was coined by Foley (1986) and is synonymous with 'cosubordinate' (Foley and Van Valin 1984). The Papuan languages described by Foley are verb-final and have strings of dependent clauses ending in an independent clause. In Paiwan the pattern is reversed.

Pazih *s‹in›u-* reflects either **S‹in›a-* or **S‹in›i-* irregularly. Perhaps (i) an original **Sa-* became **S‹in›i-* in the perfective then (ii) in some languages a form looking like a reflex of **Si-* spread by analogy to non-perfective realis slots in the paradigm.

In my earlier analysis I doubted whether PAn had circumstance-subject UV forms, as the data then available suggested that most Formosan languages reflected **Sa-* or **Si-* in nominalisations but not in finite verbs (Ross 1995:756–758). I added UVC forms tentatively to my 2002 presentation Ross (2002b:33, 42). However, the reflexes listed in Table 3 suggest quite strongly that in PNAn **Sa-* and **Si-* both formed verbs and nouns.

Table 3: The morphology of circumstance-subject

PNAn	**Sa*-STEM		**Si*-STEM		**Ca*-STEM	
Kanakanavu	—	—	*si*-STEM	N	—	—
Saaroa	*sa*-STEM[*-a*]	V	—	—	—	—
Saisiyat	—	—	*si*-STEM	V	C*a*-STEM	V, N
Pazih	*sa-/s‹in›u*-STEM	V, N	—	—	C*a*-STEM	N
Mayrinax Atayal	—	—	*si*-STEM	V	C*a*-STEM	V, N
Seediq	*se*-STEM	V, N	—	—	C*e*-STEM	V, N
Thao	—	—	—	—	C*a*-STEM[*-an*]	N
Ishbukun Bunun	—	—	*is*-STEM	V	…	…
Paiwan	—	—	*si*-STEM	V, N	—	—
Amis	*sa*-STEM	V, N	—	—	—	—
Kavalan	*sa*-STEM[*-an*][a]	N	(*ti*-STEM	V, N)	—	—
Puyuma	—	—	*i*-STEM	N	*i*-C*a*-STEM	N
Proto Rukai	**sa*-STEM[*-ane*]	N	—	—	—	—

Another fact emerges from Table 1, namely that PNAn **Sa-* and **Si-* occurred only in realis forms (unlike **-en* UVP and **-an* UVL, which also occur in irrealis forms, and unlike **M-*, also found in the optative/hortative form). The PNAn irrealis UVC form is the same as the irrealis AV form: **Ca*-STEM. It has no voice or applicative marker, as **Ca-* characterises the imperfective and the irrealis. It too formed both finite verbs and nominalisations.

The facts in the previous paragraph allow a fresh interpretation of the material presented in Blust (1998). Blust takes **Ca*-STEM to be a template just for instrument (i.e. circumstance) nominalisations, and argues that **Si*-STEM probably had the basic function of forming verbs. Almost the opposite is true: PNAn **Ca*-STEM was a member of the set of irrealis verb forms, whilst PNAn **Sa-/Si*-STEM and its perfective and imperfective variants formed realis verbs. Both could simultaneously serve as nominalisations in PNAn. Ironically, the evidence indicates that in PAn **Ca*-STEM was verbal and **Sa-/Si*-STEM was originally nominal (see §3). This undermines Blust's conclusion that Starosta et al. were wrong to derive PAn (my PNAn) voice and applicative morphology from nominalisation .

Finally, the optative/hortative and imperative/dependent UVC forms in Table 1 need comment. I have assumed on the basis of the Wulai dialect of Atayal (L.M. Huang 1994) that **an-* was a preverb which took UVL suffixes and was followed by the stem form of the

verb.[11] In all other languages that reflect these forms, however, reflexes of **anay* and **ani* are suffixed to the stem: Mayrinax Atayal UVC optative/hortative STEM-*anay*, Paiwan UVC imperative/dependent STEM-*an* (with unexpected loss of **-i*), etc. (see Appendix B).

Where I have **-ani* Wolff (1973) reconstructs **-án*, with invariable stress,[12] contrasting with the UVL suffix **-an*, which was only stressed when suffixed to a stem with final stress. In my earlier analysis I reconstructed stress, but it seems to me that Blust (1997) is correct in maintaining that we lack decisive evidence for the reconstruction of PAn stress and I err here on the side of caution.

3 Nominalisations into verbs

The PAn forms reconstructed by Ross (1995, 2002) were similar to those reconstructed by Wolff (1973). The main advance lay in filling gaps in what are here the imperative, dependent and optative/hortative paradigms: **-aw*, **-anay* and **-u* were added and UVC **-án* was amended to **an-i*. This revealed the paradigmatic pattern in the lower part of Table 1 (Ross 1995:763, 2002b:40), repeated in (1). The middle line expands the optative/hortative suffixes into the morpheme sequences from which they are historically derived: each optative/hortative suffix consists of **-a* plus zero (AV) or an imperative suffix marking UVP or UVL. For convenience I will refer to the optative/hortative set as *a*- grade suffixes and the imperative set as zero-grade suffixes.

(1)

	AV	UVP	UVL
Optative/hortative	**-a*	**-aw*	**-ay*
	**-a-Ø*	**-a-u*	**-a-i*
Imperative	**-Ø*	**-u*	**-i*

The patternedness of this paradigm stands in sharp contrast with the hotchpotch of morphemes in the upper part of Table 1, which includes two infixes, two suffixes, one prefix and **Ca*- reduplication. Of this collection, just one member, **‹um›* AV, also appears in the lower half of the table (in the optative/hortative form).

It was suggested more than thirty years ago (by Andrew Pawley in lectures at the 1977 Institute of the Linguistic Society of America) that this jumble arose because at an earlier stage these morphemes all formed argument nominalisations,[13] which were then reanalyzed as verbs. This hypothesis was elaborated in Starosta, Pawley and Reid (1981), published in abbreviated form as Starosta, Pawley and Reid (1982).[14] The jumble of forms was accounted for by the plausible supposition that a collection of nominalisers might have disparate origins. Supporting the hypothesis is the fact that all the forms in the upper part of Table 1 except the imperfectives (which were probably a PNAn innovation; see §6) are widely reflected as nominalisers in modern languages and the fact that with undergoer verb forms the actor, whether a pronoun or a full noun phrase, in Philippine-type languages is

---

11 In Wulai Atayal the form following the preverb actually reflects **Si*-STEM. I assume this to be an analogical development.

12 Malayo-Polynesian reflexes also support **-an*.

13 Argument nominalisations are nominalisations encoding an actor, patient, location etc, as opposed to action or state nominals, the early Austronesian forms of which have not been reconstructed.

14 As Starosta, Pawley and Reid (1981) point out, various scholars had suggested that undergoer verb forms in Philippine-type languages should be analysed as nominals (Lopez 1941; Capell 1964; Egerod 1966).

almost universally encoded by the genitive (i.e. possessor) case. The hypothesis continues to find support today (Kaufman 2007).

The reanalysis envisioned by Starosta, Pawley and Reid (1981) is illustrated by reconstructed PNAn sentences in (2).[15]

(2) a. **qaLup-en ka babuy*
   i. hunt-NMLZP NOM pig
   'the pigs are something to be hunted'

   ii. hunt-UVP NOM pig
   'the pigs are hunted' > 's/he hunts the pigs'

   b. **qaLup-en na aLak ka babuy*
   i. hunt-NMLZP GEN child NOM pig
   'the pigs are the child's prey'

   ii. hunt-UVP GEN child NOM pig
   'the pigs are hunted by the child'> 'the child hunts the pigs'

The sentence in (2a) is a non-verbal clause with a nominal predicate consisting of the patient nominalisation *qaLup-en* 'something to be hunted, prey'. This is reanalyzed in (ii) as a verbal predicate 'is hunted'. In (2b) the nominal predicate *qaLup-en na aLak* 'the child's prey' includes a possessor, reanalysed in (ii) as the actor.

The sentences in (3) illustrate the corresponding reanalyses of location, circumstance and actor nominalisations.

(3) a. **qaLup-an na aLak [Ca babuy] ka bukij*
   i. hunt-NMLZL GEN child [OBL pig] NOM interior
   'the interior is the child's [pig-]hunting place'

   ii. hunt-UVL GEN child [OBL pig] NOM interior
   'the interior is hunted [pigs] in by the child'
   > 'the child hunts [pigs] in the interior'

   b. **Sa-qaLup na aLak [Ca babuy] ka asu*
   i. hunt-NMLZC GEN child [OBL pig] NOM dog
   'the dog is the child's means of [pig-]hunting'

   ii. hunt-UVC GEN child [OBL pig] NOM dog
   'the dog is used-to-hunt [pigs] by the child'
   > 'the child hunts [pigs] with the dog'

   c. **q<um>aLup [Ca babuy] ka aLak*
   i. hunt-NMLZA [OBL pig] NOM child
   'the child is the one who hunts [pigs]'

   ii. <AV>hunt [OBL pig] NOM child

[15] There are numerous pitfalls in reconstructing phrasal units in a protolanguage, but it is useful to illustrate morphosyntactic structures in this way, and no better alternative comes to mind. The lexical items are drawn from Blust (1995), paying quite careful attention to reconstructed meanings. The case-markers are from Ross (2006).

The nominalisations in (2) and (3) are gerundive (clausal) nominalisations which took arguments of their own,[16] otherwise reanalysis as transitive verbs could not have occurred. An indefinite patient was permitted as an oblique argument of the nominalisation/verb.

Unfortunately gerundive nominalisations in most Formosan languages are not well described, but they clearly encode—or encoded in the past—not only voice but also aspect and mood.[17] Mayrinax Atayal and Puyuma nominalizations have realis/irrealis and perfective/unmarked aspect distinctions (L.M. Huang 2002; Teng 2008a), although it is not entirely clear how many combinations of voice, mood and aspect actually occur in these nominalisations. These distinctions probably also occur in other Formosan languages and evidently occurred in PAn. One mood distinction that has left its mark in lexical nominalisations is the UVC distinction between realis **Sa-/Si-*STEM, reflecting an earlier nominalisation, and irrealis **Ca-*STEM, reflecting the PAn irrealis verb (*pace* Blust 1998). Table 3 indicates that only Pazih and Seediq retain the formal contrast. Whether this reflects a semantic contrast remains undetermined.

It seems improbable that a language would make distinctions in the voice, mood and aspect of nominals that it did not make in its verbal system, and we can thus infer that in the period before nominalisations underwent reanalysis as verbs, i.e. the PAn period prior to PNAn, the verbal system must also have had an actor/undergoer voice contrast, patient-, location- and circumstance-subject verb forms, and the perfective/unmarked-aspect and realis/irrealis distinctions. The question is, what were the verb forms in this period? Starosta, Pawley and Reid (1981) assumed they were the forms that I label as dependent, but the reconstruction in Ross (1995) of what are here called optative/dependent forms suggests that they were partly wrong. All the evidence points to PAn dependent forms only having followed preverbs. This leaves forms ancestral to the PNAn optative/hortative forms as candidates for independent verbs.

For convenience I will use the cover terms 'first-generation verbal affixes' for the optative/hortative (*a*-grade) and imperative and dependent (zero-grade) suffixes in the lowersection of Table 1 and 'second-generation verbal affixes' for the affixes in verbal forms that reflect renalysed nominalisations, i.e. those in the upper section of the table.

4 Puyuma

Table 4 presents a summary of Puyuma verbal morphology, based except as indicated on Teng's (2008) reference grammar, which has substantially advanced knowledge of this language. The striking feature of this table is that reflexes of the second-generation affixes **‹in›*, **-en*, **-an* and **Si-* turn up only in nominalisations. Verb forms proper, below the line, reflect only the suffix array in (1), i.e. first-generation affixes.

There are two alternative explanations of this state of affairs. Either (a) Puyuma has innovated by undoing the reanalysis of predicate nominalisations as verbs which had allegedly occurred by PAn times, or (b) Puyuma continues unchanged the state of affairs reconstructed for pre-PAn. If (b) is true, then the reanalysis of predicate nominalisations as verbs had not occurred in PAn, nor had it occurred in any interstage ancestral to Puyuma.

---

[16] Reflexes in modern languages are often lexical nominals. Gerundive nominals must have undergone lexicalisation throughout the history of Philippine-type languages.

[17] Languages that make multiple distinctions in nominalisations are probably not very common, but they do exist: Turkish nominalisations, for example, distinguish tense and voice (Comrie and Thompson 1985).

Table 4: Puyuma verbal morphology

	ACTOR VOICE (intransitive)	UNDERGOER VOICE Patient subject	Location subject	Circumstance subject
Realis (N only)[a]	*M*-STEM	‹*in*›STEM	‹*in*›STEM-*an*	*i*-STEM
Irrealis (N only)[a]	*Ca*-STEM	*Ca*-STEM-*en*	*Ca*-STEM-*an*	*i*-*Ca*-STEM
Realis	*M*-STEM	STEM-*aw*	STEM-*ay*	STEM-*anay*
Optative/hortative	*M*-STEM-*a*			
Realis imperfective	*M*-*Ca*-STEM	*Ca*-STEM-*aw*	*Ca*-STEM-*ay*	*Ca*-STEM-*anay*
Negative hortative		(no undergoer voice forms)		
Imperative	STEM	STEM-*u*	STEM-*i*	STEM-*an*
Negative	*M*-STEM	STEM-*i*		
Irrealis	*Ca*-STEM	*Ca*-STEM-*i*		*Ca*-STEM-*an*

[a]Undergoer voice forms are from the Ulivelivek, Katipul and Tamalakaw dialects (Stacy Teng, pers. comm.). Nanwang has ‹*in*›STEM[-*an*] for all undergoer voices in the realis and *Ca*-STEM-*an* for all undergoer voices in the irrealis.

Answer (a) requires that precisely the verbal functions of second-generation forms which were gained in PAn were lost again in Puyuma. This is unlikely: we would expect Puyuma to preserve some reflex of the alleged intervening PAn stage, but it doesn't. Answer (a) also requires that PAn (first-generation) undergoer-voice optative/hortative forms have extended their function in Puyuma to include the realis, displacing the PAn second-generation forms—a step which seems quite implausible. Answer (b) on the other hand requires no innovations.

If (b) is true, then, as I anticipated in §1, the pre-PAn state of affairs now needs to be reconstructed for PAn, and the alleged PAn state of affairs is reconstructed only for PNAn (as in Table 1), the interstage ancestral to Atayalic, East Formosan, Paiwan, Bunun, Western Plains, NW Formosan and Malayo-Polynesian. A strong caveat is necessary here. If it can be shown that predicate nominalisations have still not been reanalysed as verbs in some modern languages, as some scholars have suggested (Lopez 1941; Capell 1964; Egerod 1966; Himmelmann 1999; Kaufman 2007), then the innovatory feature of PNAn was that predicate nominalisations replaced first-generation verb forms in main clauses without reanalysis.

5 Matters of method

Before I turn to the details of PAn reconstruction (Table 5), however, some matters of method need to be addressed.

There is little reason to doubt the broad outlines of the reconstruction in Table 1. First, it resembles the systems found in Philippine-type languages. The assumption underlying the table is that these languages are morphosyntactically more conservative than other Austronesian languages. This assumption appears to be justified, as the verbal systems found in other Austronesian languages can be derived from a Philippine-type system, but

not vice versa (Ross 2002a:52–56; Lynch, Ross and Crowley 2002:57–63). Secondly, Starosta, Pawley and Reid (1981) and Ross (1995) assume the upper nodes of Blust's (1977) subgrouping, which makes a primary division into several Formosan subgroups and a single Malayo-Polynesian subgroup embracing all Austronesian languages outside Taiwan. The evidence for this sub-grouping is independent of the verbal system. As the system reconstructed in Table 1 is reflected in more than one Formosan subgroup and in Malayo-Polynesian languages spoken in the Philippines and northern Borneo and northern Sulawesi, it must be attributed to a language ancestral to all the subgroups in which it occurs. Until now this language has been assumed to be PAn.

The claim made here is that the language to which a system like the one in Table 1 is attributable was not PAn but PNAn, a somewhat later interstage language. Puyuma, Rukai and Tsou do not reflect the PNAn system and also have systems that have little in common with each other. One may ask why so little attention has been paid to these languages in past reconstructions of the PAn verbal system. A minor reason is that only a sketchy account of Puyuma grammar (Cauquelin 1991) was available. But the major reason was an (unconscious?) adherence to the 'majority wins' principle. The Philippine type was so common among both Formosan and Malayo-Polynesian languages that it was easy to conclude that PAn was also a Philippine-type language. But within the comparative method of historical linguistics there is no maxim which encourages one to reconstruct on the basis of a majority of witnesses. *A priori* it is just as possible that the PAn system is more nearly reflected by that of Puyuma, Tsou or the Rukai dialects, and that the system in Table 1 should instead be reconstructed for an interstage ancestral to the six groups listed above plus Malayo-Polynesian. The hypothesis I put forward here is that Puyuma verbal morphology more nearly reflects that of PAn than does the verbal morphology of any other Austronesian language.

Only Starosta (1995, 2001) has previously presented a subgrouping based on the inference that the reanalysis of nominalisations as verbs had not occurred in PAn. According to Starosta, PAn split into Rukai and an unnamed subgroup containing all other Austronesian languages; the latter split into Tsou and another unnamed subgroup; the latter into Saaroa and yet another unnamed subgroup, and so on. Puyuma is absent from the 1995 version of Starosta's subgrouping and present in the 2001 version in a lower-level group which also includes Paiwan, Bunun, Siraya, Kavalan, Amis and Proto Malayo-Polynesian. Reasons for this placement are not given. The subgrouping is based on shared innovations in morphology, with some resemblances to those presented here, but it has attracted little attention, because of the obscurity of its presentation and faults in its execution noted by Blust (1999:63–66).

There is, of course, a risk in reconstructing PAn in the way I propose. Four primary branches are now attributed to Austronesian: Puyuma, Tsou, Rukai and PNAn. Whereas PNAn is reconstructed on the basis of a large number of languages, the PAn reconstruction in Table 5 relies heavily on a comparison of Puyuma and PNAn (Tsou and Rukai play smaller roles: see below). If Puyuma has undergone substantial unrecognised innovations since PAn times, then this may distort our reconstruction. For this reason it is important to attend to the details of the reconstruction and the changes which turned PAn into PNAn.

Table 5: A tentative reconstruction of Proto Austronesian verbal morphology

	ACTOR VOICE	UNDERGOER VOICE		
		Patient subject	Location subject	Circumstance subject
Realis (N only)	**M*-STEM	(*STEM-*en*)	(*STEM-*an*)	(**Sa*-/)**Si*-STEM
Realis perfective (N only)	**M*-‹*in*›STEM	*‹*in*›STEM	*‹*in*›STEM-*an*	(*‹*in*›*Si*-STEM)
Irrealis (N only)	*Ca*-STEM	*Ca*-STEM-*en*	*Ca*-STEM-*an*	(**Sa*-/)**Si*-*Ca*-STEM
Realis	*M-STEM	*STEM-*aw*	*STEM-*ay*	**an*-*ay* + STEM
Optative/hortative	*M-STEM-*a*			
Realis imperfective	*M-*Ca*-STEM	**Ca*-STEM-*aw*	**Ca*-STEM-*ay*	**an*-*ay* + *Ca*-STEM
Imperative	*STEM	*STEM-*u*	*STEM-i^a	**an*-*i* + STEM
Dependent	*M-STEM	*STEM-*a*		
Irrealis	**Ca*-STEM	(**Ca*-STEM-*a*)	**Ca*-STEM-*i*	**an*-*i* + *Ca*-STEM

Formatting: Parentheses indicate that a reconstruction is not reflected in Puyuma, Tsou or Rukai. Note that Puyuma reflects **M*-‹*in*›STEM only in fossilised form. Puyuma *m-inaṯay* 'die' is synchronically AV-die. Diachronically, however, it reflects **M*-‹*in*›*aCay* AV‹PF›die, i.e. the AV perfective form of **paCay* 'die', attesting to PAn **M*-‹*in*›STEM.

[a] The evidence is quite strong that *STEM-*i* was not only UVL, as expected, but also UVP. Only Paiwan has reflexes of both *STEM-*u* and *STEM-*i*, and here the distinction is between imperative and hortative, not between UVP and UVL.

6 Proto Austronesian verbal morphology

A tentative reconstruction of PAn verbal morphology is presented in Table 5—tentative because of the risk just mentioned. Forms which are reflected in PNAn but not in Puyuma, Tsou or Rukai are shown in parentheses.

The principal difference between PAn and PNAn is that in PAn second-generation affixes (above the line in Table 5) only formed nominalisations but in PNAn their function was expanded so as also to encode finite verbs. It is here in Table 5 that three out of four sets of parentheses occur (excluding those around **Sa*-), because Puyuma has undergone aspect syncretism in realis nominalisations, such that the distinction between PAn unmarked-aspect and perfective has been lost and Puyuma retains only a single set of realis nominalisations.

First-generation affixes are assumed to have undergone little change in Puyuma. Puyuman ergative[18] STEM-*i* and irrealis *Ca*-STEM-*i* serve as both UVP and UVL, and expected PAn UVP irrealis **Ca*-STEM-*a* is reflected nowhere, except perhaps in Saaroa UVP imperfective STEM-*a* ('perhaps' because it is not clear whether Saaroa -*a* reflects PAn *-*a* or is an irregular reflex of *-*en*). PAn UVP dependent *STEM-*a* is not reflected in Puyuma, but it is present in Tsou. Tsou has developed quite differently from Puyuma (and other Formosan languages). In Tsou all independent and many dependent clauses begin with a preverb, followed by a verb reflecting one of the PAn dependent forms. If the preverb is realis, it agrees with the verb in voice but distinguishes only AV and UV, as in (4).[19]

[18] Puyuma negative forms follow the negative preverb *aɖi* and reflect PAn dependent forms.

[19] Interlinear glosses in examples are modified to reflect the terms used in the text. Abbreviations are: AV – actor voice, DEF – definite, GEN – genitive, IMPF – imperfective, INDEF – indefinite, IRR – irrealis, ITR – intransitive, NMLZ – nominaliser, NMLZA – actor nominaliser, NMLZC – circumstance nominaliser, NMLZL – location nominaliser, NMLZP – patient nominaliser, NOM – nominative NPERS – non-personal, OBL – oblique,

(4) *i=si* *an-a* *'o* *tac~~u~~m~~u~~* *to* *amo*
PREVERB:UV:R=3S eat-UVP NOM banana OBL father
'Father's banana has been eaten.'

The transfer of voice and mood distinctions from the verb to the preverb has meant that Tsou has no reflexes of PAn verbs other than the dependent set, as shown in Table 6. The only exceptions are fossilised forms in lexical nominalisations.

Table 6: Tsou verbal morphology

	ACTOR VOICE	UNDERGOER VOICE		
		Patient subject	Location subject	Circumstance subject
Dependent	*M*-ROOT	ROOT-*a*	ROOT-*i*	ROOT-[*n*]*eni*

If PAn was like Puyuma, then realis and optative/hortative UV forms were identical, using **a*-grade suffixes (presumably in practice the difference was marked by particles or clitics). Only the AV forms differed: the realis form was **M*-STEM, the optative/hortative form **M*-STEM-*a*. When in PNAn the realis was replaced by forms reflecting realis nominalisations, the function of the UV forms was radically narrowed to optative/hortative.

One set of PNAn forms, the realis imperfective, formed with **Ca*- reduplication and second-generation affixes, did not occur in PAn. Their putative ancestors would be PAn realis imperfective nominalisations, but no language reflects such nominalisations. Puyuma has realis imperfective verbs formed with **Ca*- reduplication and **first**-generation *a*-grade suffixes. For example, PAn **Ca*-STEM-*aw* (realis imperfective UVP) is reflected in Puyuma, where PNAn innovated **Ca*-STEM-*en*. I surmise that the PNAn imperfective forms arose by an analogy with the unmarked-aspect realis forms whereby the new second-generation verbal affixes replaced first-generation suffixes. The basis of the analogy was that PAn AV **M*-STEM served both as a realis verbal form and a realis nominalisation, i.e. **M*- was simultaneously a first-and second-generation affix. Once PAn UV nominalisations had become PNAn realis verbs, the paradigmatic relationship between PNAn realis AV **M*-STEM and realis UVP *STEM-*en* provided a basis for the analogical creation of UVP **Ca*-STEM-*en* from AV **M*-*Ca*-STEM, replacing **Ca*-STEM-*aw*. The same process applied to each of the UV imperfectives, resulting in the conflation of UVP and UVL realis imperfective and irrealis forms shown in Table 1.

Indeed, the presence of **M*- among both first-and second-generation affixes may have provided the trigger for the reanalysis of nominalisations as finite verbs, since only the syntactic context determined whether a form in **M*- was being used in a noun phrase or a verb phrase. This situation continues in Puyuma. Thus in (5), where the nominalisation is undergoer voice, the predicate is marked as nominal both by the determiner *a* and by the suffix *-en*.

If this clause were verbal, we would have no determiner and the finite verbal suffix *-i*, as in (5b).

P – plural, PF – perfective, PSR – possessor, R – realis, S – singular, UV – undergoer voice, UVC – undergoer voice/ circumstance subject, UVL – undergoer voice/location subject, UVP – undergoer voice/patient subject.

(5) a. **Katipul Puyuma** (Stacy Teng pers. comm.):

a *ka-kezeng-en* *ini* *na* *hung*
NOM:INDEF *Ca*-pull-UVP:NMLZ this:NOM NOM:DEF ox
'This ox is (something) to be pulled away.'

b. (manufactured example):

tu=ka-kezeng-i *ini* *na* *hung*
AGT:3S=*Ca*-pull-UVP:IRR this:NOM NOM:DEF ox
'This ox will be pulled away (by someone).'

In (6), on the other hand, only the determiner *na* in (6a) tells us that *na s‹em›a-senay* is a nominal. The form is the same as the verb in (6b).

(6) a. *amau* *kuiku* *na* *s‹em›a-senay*
COPULAR free:1S NOM:DEF ‹AV›IMPF-sing
'The one who was singing is me.' (Ross and Teng 2005)

b. *s‹em›a-senay* *i* *Walegan*
‹AV›IMPF-sing NOM:SG Walegan
'Walegan is/was singing.' (Teng 2008a)

It is only a few short steps from here to the reanalysis of a form like *ka-kezeng-en* in (5) as a finite verb by analogy with the construction in (6a).

The functional load of nominalisations in Puyuma and certain other Formosan languages is high because they are used in relative-clause-like modifier constructions. Teng (2008a) analyses a noun phrase as a series of one or more 'small NPs', each beginning with a determiner. By default, determiners agree in case and definiteness. There is no syntactic marking of head or modifier, and the head small NP may occur anywhere in the noun phrase. Thus in (7) there are three small NPs, all nominative and definite:

(7) *[na* *suan]* *[na* *ma-ɬina]* *[na* *uɬeuɬem]*
[NOM:DEF dog] [NOM:DEF] ITR-big [NOM:DEF ITR:black]
'the [big] [black] [dog]' (Teng 2008a)

The Puyuma equivalent of a relative clause behaves in much the same way: it is a gerundive nominalisation:

(8) *[na* *teɬu-a]* *[na* *kipiŋ]* *[na-ntu*
[NOM:DEF three-NPERS] [NOM:DEF clothes] [NOM:DEF-PSR:3S

d‹in›away *kan* *nanali]*
‹PF:NMLZ›make OBL:SG my.mother]
'the [three] [(pieces of) clothing] [that my mother made]' (Teng 2008a)

Again, when a small NP is in actor voice, there is no distinction between the form of a realis verb and the form of a nominalisation. Teng (2008a) analyses an AV verb in a small NP as a finite verb, as in (9), but, given the fact that UV verbs in this context are all nominalisations, one could also analyse it as an AV nominalisation that is homophonous with the AV realis form. It is precisely this ambiguity which provided the template for post-PAn speakers to reanalyse PAn nominalisations as finite verbs in PNAn.

(9) *[ɖa maʔɳɖangan] [ɖa ma-sangal ɖa basak]*
OBL:INDEF old.person OBL:INDEF AV-carry OBL:INDEF sack
'… older people who carry sacks (on their shoulders) …'

The syntactic ambiguity of AV forms in small NPs does not carry over into Puyuma independent clauses because a predicate nominal is always preceded by a nominative determiner, as in (5a), but a verb isn't. Starosta, Pawley and Reid (1981) reconstruct PAn (my PNAn) predicate nominals without a case-marker, as in (2) and (3), because there is usually no case-marker with a Formosan predicate nominal. Puyuma is exceptional in having both definite and indefinite case-marked determiners and in using a determiner with a predicate nominal. If PAn resembled Puyuma in this respect, then loss of the determiner in this context must have preceded reanalysis of nominalisations as verbs. Conversely, the retention of determiners with predicate nominals has prevented reanalysis of Puyuma nominalisations.

The Puyuma small NP construction exists in other Formosan languages, with two differences. First, small NPs other than the first in a sequence are introduced by an invariable linker. Second, whilst a UV verb in a relative-clause-like small NP reflects the same PAn forms—i.e. nominalisations—as in Puyuma, unlike in Puyuma the same verb form also occurs in an independent clause. These points are illustrated by the Paiwan examples in (10), from A.H. Chang (2006).

(10) a. *[a za vatu] [a ku=k‹in›eɭem katiaw], macay=aŋa*
[NOM that dog] [LNK AGT:1S=‹UVP:PF›hit yesterday] AV:die=COMPL
'[That dog] [which I hit yesterday], it's dead.'

b. *ku=k‹in›eɭem a za vatu katiaw.*
AGT:1S=‹UVP:PF›hit NOM that dog yesterday
'I hit that dog yesterday.'

The difference between Puyuma, with agreeing case-marked determiners between small NPs, and most Philippine-type languages, both Formosan and Philippine, with an invariable linker, again suggests that Puyuma is more conservative than other Philippine-type languages. The fact that Philippine-type languages do not agree on the form of the linker (Ross 2006, *pace* Starosta, Pawley and Reid 1981) suggests that linkers have evolved independently at various interstages in their histories, probably from determiners. Indeed, Nanwang Puyuma also allows a linker *na*, apparently reflecting the definite nominative determiner *na*, to be used between small NPs. The simplest inference from these facts is that Puyuma is again uniquely conservative and retains an NP construction prevalent in PAn.

One difficulty remains in the PAn verbal morphology presented in Table 5. The infix **‹in›* is reconstructed in perfective nominalisations on Rukai and Nuclear Austronesian evidence, yet no perfective aspect finite verbs are reconstructed in PAn. This seems odd, and perhaps indicates that Puyuma does not reflect PAn well in this regard. Two facts are relevant here.

First, Puyuma nominalisations with *‹in›* are simply realis and may also encode an imperfective sense. Second, Puyuma encodes perfective aspect with finite verbs with the enclitic *=la*. This situation allows the alternative sets of inferences below.

1. Like Puyuma, PAn marked the perfective in finite clauses with a clitic. A candidate reconstruction is PAn *=*(a)ŋa* (Rukai *=ŋa* PF, Paiwan *=aŋa* COMPLETIVE).
2. Puyuma at some point lost the perfective/unmarked-aspect distinction in its verbal morphology, and later innovated the perfective enclitic *=la*. This would imply that PAn had a now lost set of perfective finite verbs, perhaps combining *‹*in*› with first-generation affixes, but there is no evidence of this.[20]

The data do not allow us to decide between the two possibilities.

Peterson (2007:161–169) has recently offered an alternative account of the origin of second-generation forms. He rejects Starosta, Pawley and Reid's account (§3) on typological grounds, namely that it is unusual for anything except an action or state nominalization to be reanalysed as a main-clause verb form. Instead, he infers that *-*an* UVL arose by the capture of an erstwhile preposition and that **Si-* UVC represents a functional extension of **Si-* 'have, possess, wear'. Thus *-*an*, at least, arose in much the same manner as most applicative affixes in the world's languages. These forms, he suggests, are likely to have arisen in relative clauses, where the object or prepositional object is gapped. He also suggests that they were reanalysed as nominalisations when they were still restricted to relative clauses. Their use was then extended to main indicative clauses.[21]

There are several difficulties with this account. First, because it is framed within the diachronic typology of applicatives, it provides no account of *-*en* UVP, ignoring the fact that the UVP form is equal in morphological complexity with the UVL and UVC forms. Second, the claim that relative-clause use was prior to nominalisation is based on a generalisation from diachronic typology, but this generalisation is contradicted by DeLancey's (1986) account of the history of Lhasa Tibetan relative clauses, in which nominalisations have come to serve as relative-clause-like modifiers in a manner almost exactly parallel to that entailed in Teng's (2008) analysis of Puyuma. Third, as noted earlier, the difference between Puyuma verbal morphology and corresponding morphology in Nuclear Austronesian Formosan languages is more readily accounted for under the supposition that nominalisations became finite verb forms than that finite-verb uses of these forms were lost in Puyuma. Finally, Peterson's account offers no explanation of the fact that the genitive serves as the agentive case.

In fact first-generation suffixes are more promising candidates for captured morphemes, as noted by Starosta (1995:703–704).

7 The non-Nuclear Austronesian languages

There are three non-Nuclear Austronesian languages: that is, languages that do not reflect the innovations of PNAn verbal morphology. They are Puyuma, Tsou and Rukai. Puyuma verbal morphology has been dealt with at length above.

---

[20] The infix *‹*um*› was both a first-and second-generation affix, and it would not be surprising if this were also true of the infix *‹*in*›.

[21] Peterson (2007:167) also proposes an alternative scenario whereby 'there simply is no true direct diachronic relationship between relativizations/nominalizations and the focus constructions: they simply share related source elements.'

Tsou reflects only the PAn dependent forms (Table 6), and this entails a problem noted by Starosta (1985). Not only does Tsou lack second-generation verbal forms: it also appears to lack nominal reflexes of second-generation affixes. Starosta rightly makes the point that if these forms ever occurred in the language, they ought to be reflected at least in fossilized form in lexicalized nominals, and yet they are apparently not found.[22] This, Starosta suggests, means that second-generation affixes had not yet been grammaticised as nominalisers when Tsou broke away from the language ancestral to the rest of Austronesian, a suggestion that merits further investigation.[23]

Table 7: Proto Rukai verbal morphology

NOMINALISATIONS	
Agentive	…
Patient	**a*-STEM-*anə*
	**‹in›*STEM-*anə*
Location	**ta*-STEM-*anə*
Instrument	**sa*-STEM
Realis	**u-a*-STEM
Subjunctive	**‹u›*STEM
Imperative	**‹u›*STEM-*a*
Dependent	*STEM
Passive	**ki-[a]*-STEM

The Rukai system has moved in a direction quite different from Puyuma or Tsou. Rukai has six recognised dialects. Because of differences among them, it is simpler to work here with the Proto Rukai forms set out in Table 7, based mostly on material from Zeitoun (2003).

Proto Rukai has lost the undergoer voice and has acquired a passive reflecting grammaticisation of the PAn lexical prefix **ki-*‘get, obtain’, also reflected with the same function but a much lower functional load in Puyuma and Paiwan (Zeitoun and Teng 2006).[24] In other Formosan languages the default choice in narratives is an undergoer voice. In Rukai it is the actor voice. The history of Rukai is complex and still something of a mystery (Ross 2003), but Table 7 allows several observations and inferences. Reflexes of the PAn second-generation affixes **‹in›*, **-an* and **Sa-* are alive and well in Rukai nominalisations but not among finite verbs—the same situation as in Puyuma. Like Tables 1 and 5 and the table in Appendix B, Table 7 is a summary of forms across the PAn verb classes illustrated in Table 2, and it is only when we examine the Rukai verb classes in Table 8 that certain probable historical facts emerge.[25] The correspondence between the five

[22] Szakos (n.d.) records numerous placenames and family names in *-ana*, apparently reflecting PAn **-an*, but the absence of other lexical items in *-ana* opens up the possibility that these are all borrowings. Szakos (1994:73–76) records nominalising affixes, the most frequent of which is *le-*, prefixed to a verb inflected forvoice to form agent, location and instrument nominals. I have found no cognates.

[23] In his 1995 subgrouping, however, Starosta places Rukai—which does reflect second-generation affixes as nominalisers—at a node above Tsou. No reason is given for the abandonment of the 1985 position.

[24] There is a second, non-agentive, passive reflecting a prefix **ku-*, but it is missing from Mantauran and Maga and may be a later development.

[25] The passive is omitted from the table, as its form is **ki-*[*a-*]ROOT with all verb classes (**-a-*is reflected in all dialects except Mantauran)

PAn/PNAn verb classes in Table 2 and the five Proto Rukai classes is imperfect. Proto Rukai class 5 clearly corresponds with PAn class 4. In Proto Rukai classes 2 and 4, **m-* ROOT means that the root either has an initial vowel or an initial **p-* or **k-*, replaced by **m-*.

Table 8: Proto Rukai verb classes

Class	PAn source	1	2	3	4	5
Proto Rukai:						
Realis	?	**u-a-*ROOT	**u-a-*ROOT	**m-*ROOT	**a-m-*ROOT	**ma-*ROOT
Subjunctive	**M-*STEM	* ROOT	**<u>*ROOT	**m-*ROOT	**m-*ROOT	**ma-*ROOT
Imperative	**M-*STEM-*a*	* ROOT-*a*	**<u>*ROOT-*a*	...	**m-*ROOT-*a*	**ma-*ROOT-*a*
Dependent	* STEM	* ROOT	* ROOT	* ROOT	* ROOT	**ka-*ROOT

Crucially, the prefix **ma-*in class 5 corresponds with **ma-* in PAn class 4, and we know from Table 2 that **ma-*ROOT is a manifestation of PAn AV **M-*. There is a complication, however: in Table 8 class 5 **ma-* corresponds both with the realis and with the subjunctive and imperative forms in the other classes. The most straightforward inferences here are as follows.

1. The subjunctive and imperative sets reflect PAn AV **M-*STEM,**‹u›* reflecting **‹um›* with nasal loss.
2. The imperative reflects PAn **M-*STEM-*a* (AV optative/hortative).
3. Realis **u-a-*in classes 1 and 2 and **a-m-*in class 4 reflect a combination of PAn **M-*STEM and a prefix or infix **(-)a-*, the origin of which remains unknown.[26]
4. Dependent STEM reflects the PAn **M-*STEM AV dependent.

The fact that Rukai retains second-generation morphemes only in nominalizations meshes with the fresh reconstruction of PAn verbal morphology proposed here. It could of course be argued that if second-generation UV morphemes had been reanalysed as finite verbs, they would then have been lost in any case. However, we would expect to find some fossil record of their verbal use somewhere in the language, and we don't.

Kanakanavu and Saaroa have previously been subgrouped as 'Tsouic' along with Tsou. However, both languages reflect second-generation affixes in finite verbs (Appendix B), indicating that they are Nuclear Austronesian and do not subgroup with Tsou. Confusingly, both languages also appear to have a first-generation affix in one or more UVP forms where, on the basis of the reconstructed PNAn system in Table 1, it should have been replaced by a reflex of second-generation **-en*. The critical forms are Kanakanavu STEM-*ai* and Saaroa UVP forms in *-a*.

Kanakanavu STEM-*ai* is labeled 'OF2' (object focus 2) by Mei (1982:212–214) and 'special focus' by Tsuchida (1976). According to Mei (1982:227–228) the UVP form STEM-*ene* (his 'OF1') only occurs in certain subordinate clauses, while STEM-*ai* is the default UVP form.

[26] Starosta (1995:701–702) and Ross (1995:746–747) both suggested that Proto Rukai realis **u-a-*reflected a PAn preverb, a grammaticisation of **ua* 'go'. Zeitoun (2003) has questioned this, and it is clear from inspection of the full range of verb classes that she was right to do so.

If Mei is correct, then STEM-*ai* retains its PAn realis status,[27] a fact which would be problematic for its Nuclear Austronesian status. But Tsuchida's analysis is quite different: STEM-*ene*, regularly reflecting *-*en*, is UVP realis, and STEM-*ai* occurs in circumstances that he does not properly understand, but always in narrative. The texts in Tsuchida (2003) bear this out, and this is the analysis adopted in Appendix B. Kanakanavu is thus straightforwardly Nuclear Austronesian.

Saaroa UVP forms in *-a* raise a different difficulty, namely that *-a* is found where a reflex of PNAn *-*en* is expected. Is Saaroa *-a* (i) a reflex of PAn UVP dependent *-*a*? Or (ii) an irregular reflex of PNAn *-*en*? Two facts favour (ii). First, *-a* co-occurs with *li-*, the Saaroa reflex of *‹*in*›*, and the latter never coocurs with first-generation affixes. Second, PAn *-*a* marked the dependent, and would have undergone a massive extension of function to occur in the realis, imperfective, perfective and irrealis as Saaroa *-a* now does. For the time being I assume that (ii) is true and that Saaroa is also straightforwardly Nuclear Austronesian.

8 The agent case question

In Philippine-type languages the subject NP is marked as nominative and in an undergoer voice clause the agent NP is marked as genitive. Both are referenced by pronominal clitics.[28] In most languages these are second-position enclitics, i.e. they follow the first constituent of the clause, which in many languages is usually the verb but in some is often a preverb. Only *one* set of clitics is reconstructable for PAn, referencing either subject or agent (Ross 2006:532), and this situation apparently still prevailed in PNAn.

Starosta, Pawley and Reid (1981) argued that agents in Philippine-type languages are in the genitive case—the case of the possessor—precisely because the verbs with which they co-occur were once gerundive nominalisations which would have treated their agents as possessors.[29] If this is so, then we would expect to find that agents in languages where nominalizations have not been reanalyzed as verbs—Tsou, Rukai and Puyuma—are not in the genitive case. This is arguably what we find in Tsou. Rukai, described in §7, has lost the undergoer voice and become an accusative language, and so the agent case question does not arise. The Puyuma situation is more complicated.

Tsou enclitics reflect the probable PAn situation: there is only one enclitic pronoun set (the enclitic is attached to the preverb) and it marks both nominatives and agents.

Puyuma pronominal clitics are shown in Table 9. The agent (AGT) and possessor in nominative-NP proclitics (P/NOM) and the nominative enclitics differ little from each other in form, reflecting the single set of PAn clitics. The procliticisation of the agentive forms is explained by Starosta, Pawley and Reid (1981) and Wolff (1996): originally an enclitic to the preverb (as in Tsou), with loss of the preverb the pronominal became proclitic to the following verb. The history of Puyuma possessor pronouns awaits explanation.

With regard to NP case-marking, Tsou makes no genitive/oblique distinction: there is only an oblique case. There is thus no genitive agent marking in Tsou. The same is true of the Nanwang dialect of Puyuma, with oblique common *kana*, personal *kan*. In Katipul and Ulivelivek Puyuma, this is true of indefinite noun phrases, but case-markers in both

---

[27] But with a change of undergoer subject from UVL to UVP—a common enough shift.

[28] There is some variation across languages as to whether a clitic is invariably present or is present only when there is no subject/agent noun phrase.

[29] Capell (1964) considered this a reason to analyse undergoer voice verbs as nominals.

dialects distinguish between definite genitive (common *na* or *nina*, personal *ni*) and definite oblique (common *kana*, personal *kani*). In Ulivelivek an agent is marked as oblique, in Katipul as genitive (Teng 2008b). It is a reasonable inference that Katipul reflects pre-Puyuma morphology, since the obliques appear to consist of a preposition *ka-* and a genitive case-marker, but we cannot tell whether the pre-Puyuma agent was case-marked as genitive (as in Katipul) or as oblique (as in Nanwang and Ulivelivek).

Table 9: Puyuma pronominals

	1S	2S	3S	1IP	1EP	2P	3P
PAN	**=ku*	**=Su*	–	**=(i)ta*	**=mi[a], *[S]ami*	**=mu*	–
NOM	*=ku*	*=yu*	∅	*=ta*	*=mi*	*=mu*	∅
AGT	*ku=, ti=*	*nu=*	*tu=*	*ta=*	*mi=*	*mu=*	*tu=*
P/NOM	*ku=*	*nu=*	*tu=*	*ta=*	*niam=*	*mu=*	*tu=*

The agent is thus not encoded by a distinct genitive case clitic in Tsou, and I take this to have been the situation in PAn. It would be convenient to claim that the same was true of agent noun phrases. It is true of Tsou and of two Puyuma dialects, but not of definite agents in the third.

We are left, in any case, with a puzzle. In Nuclear Austronesian languages of the Philippine type we would expect undergoer voice verbs with second-generation forms to be accompanied by an agent in the genitive case, but those with first-generation forms to be accompanied by an agent in the oblique case. To my knowledge this situation does not occur: the agent of an undergoer voice verb is in the genitive case regardless of that verb's form. We can only infer that, perhaps as early as PNAn, the new case alignment appearing with second-generation forms was generalised to first-generation forms.

9 Subgrouping

The set of innovations involved in the reanalysis of nominalisations as finite verbs (§3) is complex, and it is improbable that they occurred independently in different languages. Instead, they probably occurred once, in PNAn. No shared innovations have been found supporting a subgroup containing two or three of Puyuma, Tsou, Rukai and PNAn, and so each is assumed to form a primary subgroup of Austronesian in its own right. That is, Austronesian has four primary branches: Puyuma, Tsou, Rukai and Nuclear Austronesian. In this concluding section I compare this subgrouping briefly with three other current hypotheses: Tsuchida's (1976:9–15), which continues to be cited in descriptive works, Sagart's (2004), and Blust's (1999).

The Nuclear Austronesian hypothesis is compatible with neither Tsuchida's nor Sagart's subgroupings, as it cuts across a major subgroup in each case. It cuts across Tsuchida's Southern Formosan (consisting of all Formosan languages except Atayalic) and across both its member subgroups, Rukai-Tsouic and Paiwanic, the latter consisting of Puyuma, Siraya, Paiwan, Amis, Bunun, Thao, Saisiyat and Pazih.

Sagart has three primary subgroups: Saisiyat and Pazeh are single-language subgroups, whilst Pituish contains all other Austronesian languages. Pituish in turn consists of some single-language subgroups (Thao and four extinct languages) and Enemish, consisting in its turn of Siraya and Walu-Siwaish. The latter contains six subgroups: Puyuma, Rukai-

Tsouic, Amis, Bunun, Paiwan and Muish, which includes Kavalan and Malayo-Polynesian. Nuclear Austronesian cuts across Walu-Siwaish, so that the Nuclear Austronesian hypothesis and Sagart's are irreconcilable.

The incompatibilities between these three hypotheses reflect differences in method (Tsuchida's subgrouping is based on lexicostatistics, modified by certain shared innovations) and the differences in the weight given to different kinds of innovation. The Nuclear Austronesian hypothesis rests on innovations in verbal morphology, whilst Sagart's Pituish, Enemish and Walu-Siwaish, are based on innovations in terms for the numerals 5 to 9.

Blust (1999:44–53) uses phonological evidence to place Formosan languages into nine subgroups. Three are established on the basis of shared phonological innovations (East Formosan, Western Plains and Northwest Formosan), four have only a single language each (Puyuma, Rukai, Paiwan and Bunun), and two, Atayalic and Tsouic, are taken as established on the basis of research by other scholars. Of the nine subgroups, six (NW Formosan, Atayalic, Western Plains, Bunun, Paiwan, and East Formosan[30]) reflect the PNAn system in Table 1 and two (Puyuma and Rukai) do not. The ninth subgroup, Tsouic, appears to fall in both camps: two members, Kanakanavu and Saaroa, discussed in §7, reflect the system in Table 1 and the third, Tsou (Table 6) does not.

Blust's hypothesis and the Nuclear Austronesian hypothesis can thus be reconciled provided that Blust's Tsouic subgroup is broken into Kanakanavu and Saaroa on the one hand and Tsou on the other. Significantly, perhaps, Tsouic is one of the two subgroups[31] which Blust (1999:52) takes as established on the basis of the work of other scholars. Since he establishes other groups on the basis of phonological innovations, I infer that he has not found phonological innovations which uniquely associate the Tsouic languages. In his list of 'significant mergers' Tsou, Kanakanavu and Saaroa appear to share three innovations: (i) PAn **j* is lost; (ii) merger of PAn **S* and **s*; (iii) merger of PAn **k* and **g*. None of these is convincing as a uniquely shared innovation, i.e. one which occurred in a putative Proto Tsouic, nor does Blust claim that they are. Loss of **j* is categorical in Tsou, but occurs only adjacent to **i* in Kanakanavu and Saaroa. Mergers (ii) and (iii), on the other hand, are categorical in Kanakanavu and Saaroa but only partial in Tsou. There are grounds here for a subgroup comprising Kanakanavu and Saaroa, but not including Tsou.

Blust's case for Tsouic rests on Tsuchida's (1976) work, but Tsuchida does not provide a list of shared innovations. Blust perhaps assumes that the exclusively shared lexicon of Tsou, Kanakanavu and Saaroa (Tsuchida 1976:6–10), established lexicostatistically, is extensive enough for him to infer that the three languages share a significant collection of shared innovations. However, identifying lexical innovations entails distinguishing between them and shared inheritances. This is easier to do within an Austronesian subgroup like Oceanic, where one can appeal to non-Oceanic languages as external witnesses. It is difficult when one is dealing with the primary subgroups of Austronesian, as there are no external witnesses to help determine which items should be reconstructed for PAn—and are therefore inherited into daughter languages—and which items are innovations in primary subgroups. Tsuchida (1976:15), incidentally, considered Kanakanavu and Saaroa to form a subgroup within Tsouic.

---

[30] Blust's NW Formosan consists of Kulon-Pazih and Saisiyat and is only weakly supported, as he points out. The only diagnostic innovation is **C* > *s*, and this was also reflected in the extinct Central Western Plains languages Taokas, Papora and Hoanya.

[31] The other subgroup is Atayalic, whose members are so similar that their relationship is obvious by inspection.

In a recent paper, H.Y. Chang (2006) has also questioned the Tsouic subgroup, recognising that Tsou displays morphosyntactic features which do not occur in any other Formosan language. His observations are correct (and it is useful to have them in a single publication), but he takes them to be innovations relative to the earlier reconstruction of PAn morphosyntax (Wolff 1973; Ross 1995) which single Tsou out as a subgroup in its own right. However, as he says himself, these innovations do not necessarily speak against a Tsouic subgroup: they could have occurred after Tsou speakers had become separated from Kanakanavu and Saaroa. Refuting the existence of a Tsouic subgroup entails showing that alleged shared Tsouic innovations are not what they seem, as I have tried to do above. It also entails proposing an alternate subgrouping hypothesis and showing that Tsou, Kanakanavu and Saaroa cannot belong to the same subgroup: this is a spin-off of the Nuclear Austronesian hypothesis.

If the Nuclear Austronesian proposal is superimposed on Blust's subgrouping, the result is as shown below. Ten primary subgroups are reduced to four:

1. *Puyuma*
2. *Rukai*
3. *Tsou*
4. Nuclear Austronesian
    a. *Kanakanavu* and *Saaroa*
    b. Northwest Formosan: *Saisiyat, Kulon-Pazih*[32]
    c. Atayalic: the dialects of *Atayal* and *Seediq*
    d. Western Plains: *Thao, Taokas, Favorlang-Babuza, Papora, Hoanya*
    e. *Bunun*
    f. *Paiwan*
    g. East Formosan: *Basay-Trobiawan, Kavalan, Amis, Siraya*
    h. Malayo-Polynesian: all extra-Formosan Austronesian languages (including *Yami*, which lies within Taiwan's political boundary)

Appendix A: Functions terms and their definitions

In order to compare form–function pairings across languages the following function terms and definitions are used:

Realis:	non-future, used for present, past and sometimes habitual.[33]
Irrealis:	future and hypothetical non-future events.[34]
Subjunctive:	irrealis used only in subordinate clauses.[35]
Non-past:	future, habitual (in Amis only).
Perfective:	completed, usually past, realis event.

[32] The evidence for Northwest Formosan is weak, and I would prefer to treat Saisiyat and Kulon-Pazih as separate subgroups, but this issue lies beyond the scope of this paper.
[33] In Teruku also after one future auxiliary.
[34] In Saaroa also habitual and negative.
[35] In Mantauran Rukai also used as an imperative.

Imperfective: incomplete realis event regardless of tense, typically an ongoing atelic event (progressive, or, with a stative verb, a change of state), or sometimes an iterative or habitual event.

Durative: a process continuing for an appreciable time: 'keep/kept on doing something', a repeated or frequent action; contrasting in some languages with imperfective.

Imperative: command addressed to hearer.

Hortative: inclusive imperative: command addressed to self and hearer ('let us').

Optative: volition or intention.

Narrative: event in a narrative sequence (in Kanakanavu) or after a coordinator in a coordinate dependent clause (Paiwan, Mantauran Rukai, Kimaragang, Eastern Kadazan, Timugon Murut).

Dependent: after a preverb.

Negative: after a negative preverb (i.e. negative is a subset of dependent).

Timerative: in Amis only, 'I am afraid that'

Appendix B: Verb forms in Formosan languages

This appendix sets out the verb forms in Formosan languages on which the reconstructions in the paper are based. These forms are the result of an application of the analytic approach described in §2 to materials from the sources listed in Appendix C.

Verb forms in Formosan languages

	ACTOR VOICE	UNDERGOER VOICE		
		Patient subject	Location subject	Circumstance subject
Puyuma				
Realis nominal	*M*-STEM	<*in*>STEM	<*in*>STEM-*an*	*i*-STEM
Irrealis nominal	*Ca*-STEM	*Ca*-STEM-*en*	*Ca*-STEM-*an*	*Ca*-STEM-*an*
Realis	*M*-STEM	STEM-*aw*	STEM-*ay*	STEM-*anay*
Optative/hortative	*M*-STEM-*a*	STEM-*aw*	STEM-*ay*	STEM-*anay*
Imperfective	*M-Ca*-STEM	*Ca*-STEM-*aw*	*Ca*-STEM-*ay*	*Ca*-STEM-*anay*
Imperative	STEM	STEM-*u*	STEM-*i*	STEM-*an*
Negative	*M*-STEM	STEM-*i*	STEM-*i*	STEM-*an*
Irrealis	*Ca*-STEM	*Ca*-STEM-*i*	*Ca*-STEM-*i*	*Ca*-STEM-*an*
Proto Rukai				
Nominal	–	**a*-STEM-*anə*, *<*in*>STEM-*anə*	**ta*-STEM-*anə*	**sa*-STEM
Realis	**M-Ca*-STEM ?	–	–	–
Subjunctive	**M*-STEM	–	–	–
Imperative	**M*-STEM-*a*	–	–	–
Narr./Dependent	*STEM	–	–	–

	ACTOR VOICE	UNDERGOER VOICE		
		Patient subject	Location subject	Circumstance subject
Tsou				
Dependent	*M*-STEM	STEM-*a*	STEM-*i*	STEM-[*n*]*eni*
Kanakanavu				
Realis	*M*-STEM	STEM-*ene*	STEM-*ene*	–
Future	...	...	a-STEM-*ene*	–
Imperfective	*M*-*Ca*-STEM	...	...	–
Perfective	<*in*>*M*-STEM	<*in*>STEM	<*in*>STEM-*ane*	–
Nominal	–	<*in*>STEM	*ta*-STEM-*ane*	*si*-STEM[36]
Narrative	–	STEM-*ai*	–	–
Imperative	*M*-STEM-*a*	STEM-*au*/-*i*	STEM-*au*/-*i*	–
Dependent	–	STEM	–	–
Durative	*M*-CV-STEM	–	–	–
Saaroa				
Realis	*M*-STEM	STEM-*a*	STEM-*a*[*na*]	*sa*(*a*)-STEM[-*a*]
Imperfective	*M*-*Ca*-STEM	*Ca*-STEM-*a*	*Ca*-STEM-*a*[*na*]	–
Perfective	*ɬi*-*M*-STEM	*ɬi*-STEM-*a*	*ɬi*-STEM-*a*[*na*]	V
Irrealis	*a*-STEM	*a*-STEM-[*a*]	*a*-	STEM-*a*[*na*]
Imperative	*M*-STEM-*a*	STEM-*u*	STEM-*i*	STEM-*ani*[37]
Negative	STEM	–	–	–
Saisiyat				
Realis	*M*-STEM	STEM-*en*	–	*si*-STEM
Imperfective	CV-*M*-STEM	–	–	–
Perfective	<*in*>*M*-STEM	<*in*>STEM	<*in*>STEM-*an*[38]	<*in*>*si*-STEM[39]
Irrealis	(*ʔam M*-STEM)	*ka*-STEM-*en*	*ka*-STEM-*an*	*Ca*-STEM, *ka*-STEM
Nominal	*ka*-*ma*-STEM	<*in*>STEM, *ka*-STEM[-*en*]	*ka*-STEM-*an*	*Ca*-STEM
Imper./dependent	STEM	STEM-*i*	–	STEM-*ani*
Pazih				
Realis	*M*-STEM	STEM-*en*	STEM-*an*	*saa*-STEM
Future	CV-STEM-*ay*	CV-STEM-*en*	CV-STEM-*ay*	...
Imperfective	CV-STEM, <*a*>STEM	CV-STEM-*en*	CV-STEM-*an*	*sa*-CV-STEM
Durative	*M*-CVCV-STEM	...	...	...
Perfective	<*in*>M-STEM	<*in*>STEM	<*in*>STEM-*an*	*s*<*in*>*u*-STEM
Nominal	*ta*-STEM	<*in*>STEM, CV-STEM-*en*	<*in*>STEM-*an*, [*ta*-/*Ca*-]STEM-*an*	*saa*-STEM, *Ca*-STEM
Irrealis	*M*-STEM-*ay*	STEM-*aw*	...	...
Imper./hortative	STEM	STEM-*i*	…	…

[36] I am indebted to Stacy Teng for examples drawn from Tsuchida (2003).
[37] Tsuchida (1976:80) gives only one example each of STEM-*i* and STEM-*ani*, and assumes them both to be UVI.
[38] Rare.
[39] Rare.

	ACTOR VOICE	UNDERGOER VOICE		
		Patient subject	Location subject	Circumstance subject
Mayrinax Atayal				
Realis	*M*-STEM	STEM-*un*	STEM-*an*	*si*-STEM
Irrealis	*pa*-STEM	*Ca*-STEM-*un*	*Ca*-STEM-*an*	*Ca*-STEM
Perfective	*M*‹*in*›-STEM	STEM	‹*in*›STEM-*an*	*si*-STEM
Nominal	*M*-STEM	‹*in*›-STEM[-*an*], [*Ca*-]STEM-*an*, STEM	‹*in*›-STEM-*an*, *Ca*-STEM-*an*	*Ca*-STEM
Optative/hortative	*M*-STEM-*ay*	STEM-*aw*	STEM-*ay*	STEM-*anay*
Imperative	STEM	STEM	STEM-*i*	STEM-*ani*
Negative	BASE	STEM-*i*	STEM-*i*	STEM-*ani*
Seediq				
Realis	*M*-STEM	STEM-*un*[40]	STEM-*an*	*se*-STEM
Irrealis	*mpe*-STEM	–	–	[*Ce*-]STEM
Perfect	*M*-‹*in*›STEM	‹*en*›STEM-*an*	–	‹*en*›STEM
Nominal	*M*-[‹*en*›]STEM	‹*in*›STEM, STEM-*un*	[‹*in*›]STEM-*an*	*se*-STEM, *Ce*-STEM
Imper./hortative	STEM-*a*	STEM-*aw*[41]	STEM-*ay*	STEM-*anay*
Imper./dependent	STEM	STEM-*i*	–	STEM-*ani*
Thao				
Realis	*M*-STEM	STEM-*in*	STEM-*an*	–
Imperfective	*Ca*-*M*-STEM	–	–	–
Perfective	*M*-‹*in*›-STEM	‹*in*›-STEM[-*in*]	‹*in*›-STEM-*an*	–
Irrealis	*a*-*M*-STEM	*a*-STEM-*in*	–	–
Nominal	*M*-STEM	[‹*in*›]STEM-*an*, STEM-*in*, *Ca*-STEM-*an*	[‹*in*›]STEM-*an*	*Ca*-STEM[-*an*]
Irrealis	STEM	STEM-*a*	STEM-*a*	–
Imperative	STEM	STEM-*i*	STEM-*an*	–
Ishbukun Bunun				
Realis	*M*-STEM	STEM-*un*	STEM-*an*	*is*-STEM
Imperfective	*M*-CVCV-STEM	CVCV-STEM-*un*	CVCV-STEM-*an*	–
Perfective	‹*in*›*M*-STEM	‹*in*›STEM-*un*	‹*in*›STEM-*an*	*sin*-STEM
Nominal	–	–	*Ca*-STEM-*an*	–
Imperative	STEM-*a*	STEM-*av*	–	–
Paiwan				
Realis	*M*-STEM	STEM-*en*	STEM-*an*	*si*-STEM
Imperfective	*M*-CV-STEM	CV-STEM-*en*	CV-STEM-*an*	*si*-CV-STEM
Perfective	*na M*-STEM	[‹*in*›]STEM,	‹*in*›STEM-*an*,	‹*in*›*si*-STEM
Nominal	*M*-STEM	[‹*in*›]STEM, [*Ca*-]STEM-*en*	‹*in*›STEM-*an*, [*Ca*-]STEM-*an*	*s*‹*in*›*i*-STEM

[40] Used after certain preverbs. STEM-*an* is the unmarked predicate form.

[41] Pecoraro (1979:106). Tsukida (2005) does not record *-aw*. Instead she records *-ay* as UV.PAT and no UV.LOC form.

	ACTOR VOICE	UNDERGOER VOICE		
		Patient subject	Location subject	Circumstance subject
Optative	–	STEM-*aw*	STEM-*ay*	–
Imperative	STEM-*u*	STEM-*u*	–	STEM-*an*
Hortative	STEM-*i*	STEM-*i*	–	–
Narr./Dependent	STEM	STEM-*i*	STEM-*an*	STEM-*an*
Dependent imperf.	CV-STEM	…	…	…
Kavalan				
Realis	*M*-STEM		STEM-*an*	*ti*-STEM[42]
Realis perfective	–		‹*in*›STEM	–
Nominal	*pa*-STEM-*an*	[‹en›]STEM-*an*, STEM-*an*	*sa*-STEM[-*an*][43]	
Irrealis	STEM		STEM-*a*	STEM-*a*
Imperative	STEM		STEM-*i*	–
Haian Amis[44]				
Non-past	*M*-STEM	–	*M*-STEM-*an*	*ma*-s*a*-STEM
Future	–	ROOT-*en*		s*a*-STEM-*en*
Perfective	–	*ma*-STEM, [*ka*-]ROOT-*en*[45]	–	–
Imperative	STEM	ROOT-*en*	–	s*a*-STEM-*en*
Habitual	–	–	STEM-*an*	s*a*-STEM
Irrealis	*Ca*-*M*-STEM	*Ca*-ROOT-*en*	–	–
Optative 2	*sa*-STEM-*an*	–	–	s*a*-STEM-*aw*
Nominal	STEM	STEM-*an*, [*Ca*-]STEM-*en*	STEM-*an*	s*a*-STEM
Optative 1	*M*-STEM-*aw*	ROOT-*aw*	–	–
Timerative	*ma*-STEM-*aw*	*ma*-[*M*-]STEM-*aw*	–	–
Non-past negative	STEM	k*a*-STEM		
Siraya				
Realis	*M*-STEM	STEM-*en*	STEM-*an*	–
Imperfective	*M*-*Ca*-STEM	*Ca*-STEM-*en*	–	–
Realis past	‹*in*›*M*-STEM	‹*in*›STEM-*en*	…	–
Imperfective past	‹*in*›*M*-*Ca*-STEM	–	–	–
Durative	Apparently CVCV- applied to any realis form.			

[42] The expected reflex of **Si*- is *si*-.

[43] Elizabeth Zeitoun observes that *sa*-STEM occurs in Kavalan alongside the form *sa*-STEM-*an* reported by other sources.

[44] Amis displays a striking reorganisation of affixes. Discussion lies outside the subject matter of this paper. There are three stem categories, ‹*um*›STEM, *pi*-STEM and *ka*-STEM. The following rules provide for expansions of forms summarised in the table: *M*-+‹*um*›STEM > ‹*um*›ROOT, Ø+‹*um*›STEM > *ka*-‹*um*›ROOT, *ka*-+‹*um*›STEM > *ka*-ROOT, *sa*-+‹*um*›STEM > *sa*-*ka*-‹*um*›ROOT, *ma*-+*pi*-STEM > *ma*-ROOT, *ka*+*pi*-STEM > *ka*-ROOT.

[45] Agentive.

Appendix C: Data sources

Sources of data used in this work are listed below. Works in parentheses were mostly consulted only for minor points.

Formosan: Puyuma: Teng 2008a (Stacy Teng pers. comm.); Rukai: Zeitoun 2003, 2007; Tsou: Zeitoun (2005); Kanakanavu: Tsuchida 1976, 2003 (Mei 1982); Saaroa: Tsuchida 1976 (Radetzky 2003, 2006); Saisiyat: Yeh 1991, 2003; Pazih: Li and Tsuchida 2001, 2002 (Li 2002); Mayrinax Atayal: Huang 1995, 2000; Seediq: Tsukida 2005 (Pecoraro 1979); Thao: Blust 2003 (Wang 2004); Ishbukun Bunun: Yeh n.d.a,b (Qi 2000); Paiwan: A.H. Chang 2006 (Egli 1990); Kavalan: Tsuchida 1993 (H.Y. Chang 1997, pers. comm., Chang and Lee 2002, Lee 1997, Li and Tsuchida 2006); Amis: Wu 2006 (Y. Huang 1988, Liu 1999, Wu 2003); Siraya: Adelaar (1999).

Malayo-Polynesian: Yami: Rau 2004, 2006 (Shih 1997); Bisayan: Zorc 1977; Western Bukidnon Manobo: Elkins 1970; Bonggi: Boutin 2002; Lun Dayeh (Sarawak Murut): Clayre 2005; Kimaragang: Kroeger 2004; Eastern Kadazan: Hurlbut 1988; Timugon Murut: Prentice 1971; Brewis and Levinsohn 1991.

References

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

—— 1999. Siraya reduplication. *Oceanic Linguistics* 38:33–52.

Adelaar, K. Alexander and Nikolaus P. Himmelmann, eds. 2005. *The Austronesian languages of Asia and Madagascar*. London: Routledge.

Blust, Robert A. 1977. The Proto-Austronesian pronouns and Austronesian subgrouping: a preliminary report. *University of Hawaii Working Papers in Linguistics* 9/2:1–15.

—— 1995. Austronesian comparative dictionary. Computer files.

—— 1997. Rukai stress revisited. *Oceanic Linguistics* 36:398–403.

—— 1998. *Ca*-reduplication and Proto-Austronesian grammar. *Oceanic Linguistics* 37:29–64.

—— 1999. Subgrouping, circularity and extinction: some issues in Austronesian comparative linguistics. In Zeitoun and Li, 1999:31–94.

—— 2003. *Thao dictionary*. Taipei: Institute of Linguistics (Preparatory Office), Academia Sinica.

Boutin, Michael E. 2002. Nominative and genitive case alternations in Bonggi. In Wouk and Ross, 2002:209–239.

Brewis, Richard and Stephen H. Levinsohn. 1991. Topic and emphasis in Timugon Murut. In Stephen H. Levinsohn, ed. *Thematic continuity and development in languages of Sabah*, 29–43. Canberra: Pacific Linguistics.

Capell, Arthur. 1964. Verbal systems in Philippine languages. *Philippine Journal of Sciences* 93:231–249.

Cauquelin, Josiane. 1991. The Puyuma language. *Bijdragen tot de Taal-, Land- en Volkenkunde* 147:17–60.

Chang, Anna Hsiouchuan. 2006. A reference grammar of Paiwan. PhD thesis. The Australian National University.

Chang, Henry Yungli. 1997. Voice, case and agreement in Seediq and Kavalan. PhD dissertation. Hsinchu, Taiwan: National Tsing Hua University.

—— 2006. Rethinking the Tsouic subgroup hypothesis: a morphosyntactic perspective. In Chang, Huang and Ho, 2006:565–583.

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

Chang, Henry Yungli and Amy Peijung Lee. 2002. Nominalization in Kavalan. *Language and Linguistics* 3:349–368.

Clayre, Beatrice M. 2005. Kelabitic languages and the fate of 'focus': evidence from the Kerayan. In I Wayan Arka and Malcolm Ross, eds *The many faces of Austronesian voice systems: some new empirical studies*, 17–57. Canberra: Pacific Linguistics.

Comrie, Bernard and Sandra A. Thompson. 1985. Lexical nominalization. In Timothy Shopen, ed. *Language typology and syntactic description*. vol. 3, 349–398. Cambridge: Cambridge University Press.

DeLancey, Scott. 1986. Relativization as nominalization in Tibetan and Newari. Paper presented at the 19th International Conference on Sino-Tibetan Languages and Linguistics.

Egerod, Sören. 1966. Word order and word classes in Atayal. *Language* 42:346–369.

Egli, Hans, 1990. *Paiwangrammatik*. Wiesbaden: Otto Harrassowitz.

Elkins, Richard E. 1970. *Major grammatical patterns of Western Bukidnon Manobo*. Norman: Summer Institute of Linguistics of the University of Oklahoma.

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

Foley, William A. and Robert D. Van Valin. 1984. *Functional syntax and universal grammar*. Cambridge: Cambridge University Press.

Himmelmann, Nikolaus P. 1999. Lexical categories and voice in Tagalog. Unpublished MS.

—— 2005. The Austronesian languages of Asia and Madagascar: Typological characteristics. In Adelaar and Himmelmann, 110–181.

Huang, Lillian M. 1994. *A study of Atayal syntax*. Taipei: The Crane.

—— 1995. *A study of Mayrinax Atayal*. Taipei: Crane Publishing Co.

—— 2000. Verb classification in Mayrinax Atayal. *Oceanic Linguistics* 39:364–390.

—— 2002. Nominalization in Mayrinax Atayal. *Language and Linguistics* 3:197–225.

Huang, Yajiun. 1988. Amis verb classification. MA thesis. Taipei: Fu-jen Catholic University.

Hurlbut, Hope M. 1988. *Verb morphology in Eastern Kadazan*. Canberra: Pacific Linguistics.

Kaufman, Daniel. 2007. The nominalist hypothesis in Austronesian. Paper presented at the Workshop 'Beyond focus and ergativity: towards a more comprehensive view of Austronesian morphosyntax', Zentrum für allgemeine Sprachwissenschaft, Humboldt-Universität Berlin.

Kroeger, Paul. 2004. Morphosyntactic vs. morphosemantic functions of Indonesian *-kan*. Unpublished MS.

Lee, Amy Peijung. 1997. The case-marking and focus systems in Kavalan. MA thesis. National Tsing Hua University, Hsinchu, Taiwan.

Li, Paul Jen-kuei. 2002. Nominalization in Pazih. *Language and Linguistics* 3:227–239.

Li, Paul Jen-kuei and Shigeru Tsuchida. 2001. *Pazih dictionary*. Taipei: Institute of Linguistics (Preparatory Office), Academia Sinica.

—— 2002. *Pazih texts and songs*. Taipei: Institute of Linguistics (Preparatory Office), Academia Sinica.

—— 2006. *Kavalan dictionary*. Taipei: Institute of Linguistics, Academia Sinica.

Liu, Dorinda Tsai-hsiu. 1999. Cleft constructions in Amis. MA dissertation. National Taiwan University.

Lopez, Cecilio. 1941. *A manual of the Philippine national language*. 3rd ed. Manila: Bureau of Printing.

Lynch, John, Malcolm Ross and Terry Crowley. 2002. *The Oceanic languages*. Richmond: Curzon Press.

Mei, Kuang. 1982. Pronouns and verb inflection in Kanakanavu. *Tsing Hua Journal of Chinese Studies* New Series 14:207–231.

Pecoraro, Ferdinando. 1979. *Eléments de grammaire taroko, précédés de la présentation dela vie et de la culture des taroko*. Paris: Association Archipel.

Peterson, David A. 2007. *Applicative constructions*. Oxford: Oxford University Press.

Prentice, D.J. 1971. *The Murut languages of Sabah*. Canberra: Pacific Linguistics.

Qi, Lisha [Elizabeth Zeitoun]. 2000. *Bùnóngyǔ cānkǎo yǔfǎ [Bunun reference grammar]*. Taipei: Yuanliu.

Radetzky, Paula. 2003. Grammaticalization of a definiteness marker in Saaroa. Unpublished MS, University of Oregon.

—— 2006. The semantics of the verbal complex with particular reference to Saaroa (Taiwan). Paper presented to the Tenth International Conference on Austronesian Linguistics, Puerto Princesa.

Rau, D. Victoria and Maa-neu Dong. 2006. *Yami texts with reference grammar and dictionary*. Taipei: Institute of Linguistics, Academia Sinica.

Rau, Der-hwa Victoria. 2004. Iconicity and temporal-aspectual-modality in Yami. Paper presented to the Ninth International Symposium on Chinese Languages and Linguistics, Graduate Institute of Linguistics, National Taiwan University, November 2004.

Reid, Lawrence A. 2007. Abductive and deductive change in reduplicative plurals of human nouns in Philippine languages. Paper presented to the Second Conference on Reduplication, University of Graz.

Reid, Lawrence A. and Hsiu-chuan Liao. 2004. A brief syntactic typology of Philippine languages. *Language and Linguistics* 5:433–490.

Ross, Malcolm, 1995. Reconstructing Proto Austronesian verbal morphology: evidence from Taiwan. In Paul Jen-kuei Li, Dah-an Ho, Ying-kuei Huang, Cheng-hwa Tsang and Chiu-yu Tseng, eds *Austronesian studies relating to Taiwan*, 727–791. Taipei: Institute of History and Philology, Academia Sinica.

—— 2002a. Takia. In John Lynch, Malcolm Ross and Terry Crowley, eds *The Oceanic languages*, 216–248. Richmond: Curzon Press.

—— 2002b. The history and transitivity of western Austronesian voice and voice-marking. In Wouk and Ross, 2002:17–62.

—— 2003. Notes on reconstructing Proto Rukai case-marking. Unpublished MS.

—— 2006. Reconstructing the case-marking and personal pronoun systems of Proto Austronesian. In Chang, Huang and Ho, 2006:521–564.

Ross, Malcolm and Stacy Fang-ching Teng. 2005. Formosan languages and linguistic typology. *Language and Linguistics* 6:739–781.

Sagart, Laurent. 2004. The higher phylogeny of Austronesian and the position of Tai-Kadai. *Oceanic Linguistics* 43:411–444.

Shih, Louise Yu-mien, 1997. Yami morphology. MA thesis. Providence University, Taijung, Taiwan.

Starosta, Stanley. 1985. Verbal inflection versus deverbal nominalisation in PAN: the evidence from Tsou. In Andrew Pawley and Lois Carrington, eds *Austronesian linguistics at the 15th Pacific Science Congress*, 281–312. Canberra: Pacific Linguistics.

—— 1986. Focus as recentralisation. In Paul Geraghty, Lois Carrington and S.A. Wurm, eds *FOCAL I: Papers from the Fourth International Conference on Austronesian Linguistics*, 73–95. Canberra: The Australian National University.

—— 1995. A grammatical subgrouping of Formosan languages. In Paul Jen-kuei Li, Dah-an Ho, Ying-kuei Huang, Cheng-hwa Tsang and Chiu-yu Tseng, eds *Austronesian studies relating to Taiwan*, 683–726. Taipei: Institute of History and Philology, Academia Sinica.

—— 1999. Transitivity, ergativity, and the best analysis of Atayal case marking. In Zeitoun and Li, 1999:371–392.

—— 2001. Reduplication and the subgrouping of Formosan languages. Paper presented at the International Symposium on Austronesian Cultures: issues relating to Taiwan, Academia Sinica, Taipei.

Starosta, Stanley, Andrew Pawley and Lawrence A. Reid. 1981. The evolution of focus in Austronesian. Paper presented at the Third International Conference on Austronesian Linguistics, Bali.

—— 1982. The evolution of focus in Austronesian. In Amran Halim, Lois Carrington and S.A.Wurm, eds *Papers from the Third International Conference on Austronesian Linguistics, 2: Tracking the Travellers*. vol. 2, 145–170. Canberra: Pacific Linguistics.

Szakos, József. 1994. Die Sprache der Cou: Untersuchungen zur Synchronie einer austronesischen Sprache auf Taiwan. PhD thesis. Rheinische Friedrich-Wilhelms-Universität zu Bonn.

—— n.d. Cou-Wörterbuch. Unpublished computer file, Providence University, Taichung.

Teng, Stacy Fang-ching. 2008a. *A reference grammar of Puyuma, an Austronesian language of Taiwan*. Canberra: Pacific Linguistics.

—— 2008b. Case syncretism in Puyuma. Unpublished MS, Academia Sinica, Taipei.

Tsuchida, Shigeru. 1976. *Reconstruction of Proto-Tsouic phonology*. Tokyo: Institute for the Study of Languages, Cultures of Asia, and Africa.

—— 1993. Short sketch of Kavalan grammar. Unpublished MS.

—— 2003. *Kanakanavu texts (Austronesian Formosan)*. Osaka: Endangered Languages Pacific Rim, Osaka Gakuin University.

Tsukida, Naomi. 2005. Seediq. In Adelaar and Himmelmann, 2005:291–325.

Wang, Shan-shan. 2004. An ergative view of Thao syntax. PhD dissertation. University of Hawai‘i.

Wolff, John U. 1973. Verbal inflection in Proto-Austronesian. In Andrew Gonzalez, ed. *Parangal kay Cecilio Lopez*, 71–91. Quezon City: Linguistic Society of the Philippines.

—— 1996. The development of the passive verb with pronominal prefix in western Austronesian languages. In Bernd Nothofer, ed. *Reconstruction, classification, description: festschrift in honor of Isidore Dyen*, 15–40. Hamburg: Abera.

Wouk, Fay and Malcolm Ross, eds. 2002. *The history and typology of western Austronesian voice systems*. Canberra: Pacific Linguistics.

Wu, Joy J. 2003. Voice affixes and verb classification in Amis. Paper presented at the Second Workshop on Formosan Languages, Academia Sinica, Taipei, November 2003.

—— 2006. Verb classification, case marking and grammatical relations in Amis. PhD dissertation. State University of New York at Buffalo.

Yeh, Marie Meili. 1991. Saisiyat structure. MA thesis. National Tsing Hua University.

—— 2003. A syntactic and semantic study of Saisiyat verbs. PhD dissertation. Taipei: National Taiwan Normal University.

—— n.d.a. Bunun reference grammar. Unpublished MS.

—— n.d.b. Morphological causativization in Bunun. Unpublished MS, Lien Ho Institute of Technology.

Zeitoun, Elizabeth. 2003. Toward a reconstruction of Proto-Rukai morpho-syntax. Paper presented at AFLA X, Honolulu.

—— 2005. Tsou. In Adelaar and Himmelmann, 2005:259–290.

—— 2007. *A grammar of Mantauran (Rukai)*. Taipei: Institute of Linguistics, Academia Sinica.

Zeitoun, Elizabeth and Lillian M. Huang. 2000. Concerning *ka-*, an overlooked marker of verbal derivation in Formosan languages. *Oceanic Linguistics* 39:391–414.

Zeitoun, Elizabeth, Lillian M. Huang, Marie M. Yeh, Anna H. Chang and Joy J. Wu. 1996. The temporal, aspectual and modal systems of some Formosan languages: a typological perspective. *Oceanic Linguistics* 35:21–56.

Zeitoun, Elizabeth and Paul Jen-kuei Li, eds. 1999. *Selected papers from the Eighth International Conference on Austronesian Linguistics*. Taipei: Institute of Linguistics (Preparatory Office), Academia Sinica.

Zeitoun, Elizabeth and Stacy Fang-ching Teng. 2006. The passive *ki-*in Rukai, Paiwan and Puyuma: borrowing, shared in innovation parallel development. Unpublished MS, Institute of Linguistics, Academia Sinica.

Zeitoun, Elizabeth and Chen-huei Wu. 2006. An overview of reduplication in Formosan languages. In Chang, Huang and Ho, 2006.

Zorc, R. David. 1977. *The Bisayan dialects of the Philippines: subgrouping and reconstruction*. Canberra: Pacific Linguistics.

20 *Possession in South Halmahera–West New Guinea: typology and reconstruction*

RENÉ VAN DEN BERG

1 Introduction[1]

The Austronesian languages belonging to the South Halmahera–West New Guinea subgroup (henceforth SHWNG) show a considerable variety of possessive systems. Most (but not all) display the familiar Oceanic contrast between alienable and inalienable possession, but the semantic basis of this division as well as further subdivisions and the morphosyntactic encoding of possession is by no means uniform. Contrary to the attention given to possession in Oceanic, this topic has not been investigated for SHWNG. In fact, hardly any comparative studies have been carried out for this subgroup.

In this paper I want to tackle the issue of possessive marking in the SHWNG group by first providing a typology of the various languages (§1.3) and subsequently an attempt at reconstruction (§2). I will be guided by six basic research questions, listed in §1.2. Before I move into the details of this study, I start off by giving some background information about the SHWNG subgroup (§1.1).

1.1 The South Halmahera West New Guinea subgroup

The SHWNG languages form a clearly defined subgroup of some 40+ languages spoken in the eastern part of the Indonesian half of New Guinea and surrounding islands. Specifically they are spoken in five locations, as shown on Map 1.

[1] It is a pleasure and an honour for me to offer this paper to Robert Blust in friendship and admiration. I was fortunate to have Bob as a teacher during his years at the University of Leiden, when he introduced me to the fascinating world of Austronesian linguistics during my MA-studies in 1981–83. I cherish the memories of his lectures and the example of his scholarship, and I want to thank him for being such an inspiration to me. This paper has benefitted from comments by Ger Reesink and Malcolm Ross.

Alexander Adelaar and Andrew Pawley, eds
Austronesian historical linguistics and culture history: a festschrift for Robert Blust, 327–357.
Pacific Linguistics, 2009.
© This edition vested with Pacific Linguistics.

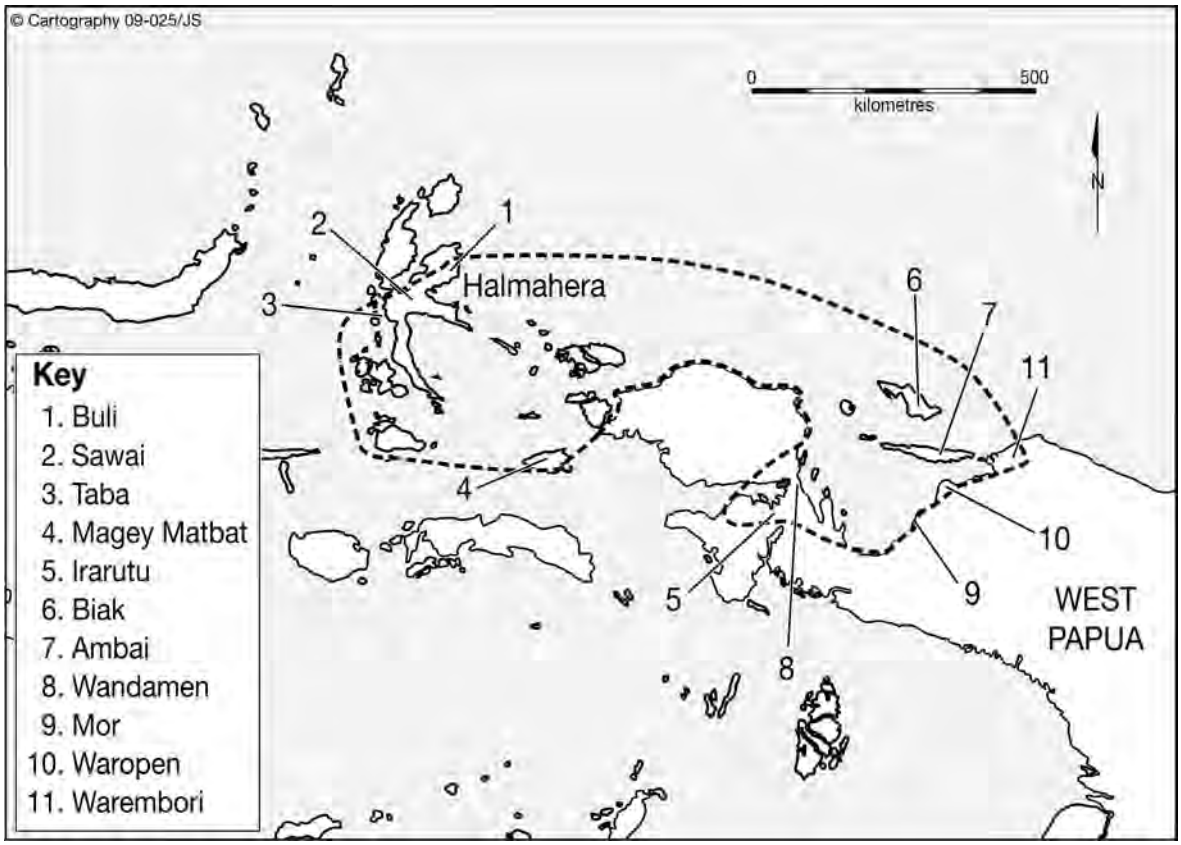
- on the southern half of the island of Halmahera in the Northern Moluccas, including the satellite islands of Makian and Gebe; (the northern half of Halmahera is occupied by languages of the West Papuan group);
- on the Raja Ampat islands (Waigeo, Misool, Salawati and Batanta) between Halmahera and the eastern tip of New Guinea;
- on the major islands of Yapen and Biak;
- along the coast of the Cenderawasih Bay (also known as the Geelvink Bay) and small islands near that coast;
- in the interior of the Bomberai peninsula.

Notice that the major landmass of the eastern tip of New Guinea, the so-called Bird's Head Peninsula (formerly known as the Vogelkop), is the exclusive home of Papuan languages, belonging to several unrelated families (Reesink 2005).

The existence of a SHWNG subgroup was first postulated by Adriani and Kruyt (1914), followed by Esser (1938) and subsequently put on a firm evidential basis by Blust (1978). The current estimate is that there are some 42–45 languages in this group, some of which (Biak, Buli, Wandamen, Waropen) have been the object of study by Dutch missionaries starting in the late 19th and early 20th century. In spite of this early attention, very few modern descriptions of SHWNG grammars are available, with the notable exception of Taba (Bowden 2001) and Biak (van den Heuvel 2006). Traditionally the group has been divided into two subgroups: South Halmahera (SH) and West New Guinea (WNG). Remijsen (2001) has argued that the languages of the Raja Ampat islands should be subgrouped with the South Halmahera languages into a RASH subgroup (Raja Ampat—South Halmahera), a proposal which I follow here. Blust (1978) tentatively divides the West New Guinea languages (his Sarera group) into four main branches: 1. Biak group; 2. Waropen; 3. Mor; 4. Yapen group. I will not be concerned with the details of subgrouping in this article.

Apart from Blust (1978) and some notes in Ross (1995), no comparative studies have been carried out within this subgroup, which is all the more surprising given its unique position between the western Austronesian languages and the eastern Oceanic subgroup. Many typical Oceanic features can already be observed among the SHWNG languages and the question can be asked at which point such innovations entered the Austronesian family. The alienable–inalienable distinction is a case in point. Another interesting domain for research is the influence of Papuan languages on the subgroup as a whole, as well as on individual languages. Klamer et al. (2007) and Reesink (2005) are important studies for the understanding of the Papuan-Austronesian interrelationships in eastern Indonesia. The typological variation in some areas of morphosyntax within this subgroup is quite large, and hence many avenues for investigation can be fruitfully explored. Possession, the topic of this article, is one such area.

In this historical-comparative study I restrict myself to the eleven SHWNG languages for which a reasonable amount of descriptive material is available, though in some cases (e.g. Mor) the information is minimal. These languages are listed in Table 1 and their location is indicated on Map 1.



Map 1: Location of the SHWNG languages

Table 1: SHWNG languages used in this study

Language	Subgroup	Location	Source
Buli	SH	South Halmahera	Maan (1951)
Sawai	SH	South Halmahera	Whisler (1996)
Taba	SH	Makian island	Bowden (2001)
Magey Matbat	RA	southeast Misool	Remijsen (2001, to appear)
Irarutu	WNG	interior of Bomberai peninsula	Voorhoeve (1989), Matsumura and Matsumura (1991), van den Berg and Matsumura (2008)
Biak	WNG	Biak and surrounding islands	van den Heuvel (2006)
Ambai	WNG	eastern Yapen	Silzer (1983)
Wandamen	WNG	west coast of Cenderawasih bay	Cowan (1955), Anceaux (1961), Ramar et al. (1983), Henning (n.d.), own field notes
Mor	WNG	island in Cenderawasih bay	Anceaux (1961), Laycock (1978),
Waropen	WNG	east coast of Cenderawasih bay	Held (1942), own field notes
Warembori[2]	WNG	eastern tip of Cenderawasih bay	Donohue (1999)

[2] Although Warembori is listed as a non-Austronesian language in the *Ethnologue* (following Donohue's (1999) tentative classification), I follow Ross (2005:30, fn.14) and Donohue (pers. comm.) and treat it as a Papuanised Austronesian language. Mainly because of its location I assign it to the WNG group, though this remains to be confirmed.

For the sake of completeness, I list all other SHWNG languages in Table 2 following the 15th edition of the *Ethnologue* (Gordon 2005), divided by geographic location.

Table 2: Other SHWNG languages

Subgroup	Location	Languages
SH	South Halmahera	Gane (Giman), Maba, Patani, Gebe
RA	Raja Ampat.	Biga, Kawe, Maden, Matbat, Ma'ya, Wauyai, Waigeo, Legenyem, As[3]
WNG	Bomberai	Kuri (Nabi)
	Cenderawasih Bay	Roon, Dusner, Meoswar, Iresim, Tandia, Yaur, Yeretuar
	Yapen	Ansus, Busami, Marau, Munggui, Papuma, Pom, Serui-Laut, Woi, Wabo, Kurudu.

Remijsen (2001) gives a slightly different list for the Raja Ampat languages. He recognises Matbat, Ma'ya, Biga and Ambel but considers Wauyai, Legenyem (his Laganyam) and Kawe to be Ma'ya dialects spoken on Waigeo. The situation on Salawati is confusing - the varieties spoken there could be one or two separate languages, or simply dialects of Ma'ya, including what has been termed Maden. In Remijsen (to appear) Matbat is tentatively split into Magey Matbat and Tomolol Matbat.

For most of these languages little or no information is available apart from a wordlist, and quite a few of them are highly endangered. The number of speakers listed in the *Ethnologue* for Tandia, for example, is only 2, for Dusner 20, for Iresim 70, for As 230, for Meoswar 250, and for Yaur and Yeretuar 350 each. Clearly time is running out fast if these languages are to be properly documented before its speakers have completely and irrevocably shifted to Papuan Malay.

1.2 The questions

Much research has been done on possession in Oceanic languages and Proto Oceanic (Lichtenberk 1985 and references there; see also Lynch, Ross and Crowley 2002). The possessive system of Proto Austronesian has also been reconstructed with a reasonable degree of certainty (Blust 2005; Ross 2002, 2006). Because the wider picture of possession in Austronesian is fairly clear, this will enable us to ascertain where the SHWNG languages have gone their own paths and to what extent the present-day languages deviate from the system as it can be reconstructed for the protolanguage.

In this paper I propose to answer the following six research questions. The answers to these questions will lead to a fairly comprehensive reconstruction of possession in Proto SHWNG.

1. Do all SHWNG languages make a morphosyntactic distinction between alienable and inalienable possession? Can this categorial distinction be reconstructed for Proto SHWNG?

[3] As is actually spoken on the northwest coast of the Bird's Head, but is clearly a Raja Ampat language (Remijsen 2001).

2. How is inalienable possession encoded in SHWNG languages?
3. What inalienable forms can be reconstructed for Proto SHWNG?
4. How is alienable possession encoded in the SHWNG languages? And what can be reconstructed for Proto SHWNG?
5. What is and what was the semantic basis of the alienable versus inalienable possession?
6. What is and what was the order of the possessor versus the possessed noun?

Before I proceed however, two words of caution are in order. In the first place one necessary step in historical-comparative research has been skipped, namely the reconstruction of Proto SHWNG phonology. Though both Blust (1978) and Ross (1995) offer lists of phonological innovations (which are partially different), no phonological reconstruction has been proposed in the literature. This means that the actual phonological shape of the proposed reconstructions stands on a rather unstable footing and must be subject to future scrutiny.[4]

Secondly, although I have made use of all available information on possession in SHWNG languages, it is possible that new data from hitherto undescribed languages will modify the reconstruction considerably. This is simply a given when doing comparative work in large language families with undocumented members, and although I believe enough is known to warrant the proposed reconstructions, this attempt is therefore no more than a first step towards Proto SHWNG morphosyntax.

1.3 The data

In this section I present some of the possessive data from the 11 languages. Space does not permit a detailed overview of each language, so I restrict myself to a table with pronominal forms, followed by a statement of the main patterns and a few illustrative examples. For each language I present the minimal necessary information for an understanding of possession; details, complications and non-relevant issues are usually left out of consideration. For the sake of comparison I also give the full pronouns. Variation is indicated by ~. Notice that I restrict myself to singular and plural forms. Though duals and trials are found in many SHWNG languages, I have simply decided to leave them out of this study. They are often (but not always) transparently related to the words for ‘two’ and ‘three’; structurally they usually add little to the analysis, and since they are mostly lacking in SH, RA and in Irarutu, it is not clear to me at this point whether duals and trials should be reconstructed for Proto SHWNG. This is clearly an area for future study.

[4] I suspect that the few reconstructions which I propose will turn out to be phonologically uncontroversial. One issue that does need to be addressed in this respect, however, is Ross’s (1995:85) implied claim that Proto SHWNG had a six-vowel system, including /ə/.

1.3.1 Buli

Table 3: Buli

	Free pronoun	Inalienable	Alienable general	Alienable edible
sg 1	*ya*	*ya-…-k*	*(ya-)ni-k*	*(ya-)na-k*
sg 2	*awe ~ au*	*a-…-m*	*(a-)ni-m*	*(a-)na-m*
sg 3	*i*	*i-*	*(i-)ni*	*(i-)na*
pl 1in	*ite*	*ite-…-r*	*(ite-)ri-r*	*(ite-)ra-r*
pl 1ex	*ame(k)*	*ame-…-mam*	*(ame-)ni-mam*	*(ame-)na-mam*
pl 2	*meu*	*meu-…-meu*	*(meu-)ni-meu*	*(meu-)na-meu*
pl 3	*si ~ sil ~ sile*	*si-…-ri ~ sile-…-ri*	*(si-)ri-ri*	*(si-)ra-ri*

Inalienable nouns are directly marked by a circumfix; alienable nouns are divided into two categories (general and edible[5]) and indirectly marked on two classifiers, for which long and short forms exist.

Examples:	inalienable	*ya-boboko-k*	‘my head’
		a-boboko-m	‘your head’
		i-boboko	‘his/her head’
		ame-boboko-mam	‘our (excl) heads’
	alienable general	*ya-ni-k ebai*	‘my house’
		meu-ni-meu ebai	‘your (pl) house’
	alienable edible	*ya-na-k pinge*	‘my rice’
		a-na-m pinge	‘your rice’

1.3.2 Sawai

Table 4: Sawai.

	Free pronoun	Inalienable	Alienable general	Alienable edible
sg 1	*ya*	*(a-)…-g ~ (ya-)…-g*	*ni-k ~ (y)a-ni-k*	*na-k ~ (y)a-na-k*
sg 2	*aw(e)*	*(a-)…-m*	*(a-)ni-m*	*(a-)na-m*
sg 3	*i*	*(i-)*	*(i-)ni*	*(i-)nó*
pl 1in	*it(e)*	*(ite-)…-r*	*(ite-)ri-r*	*(ite-)rór*
pl 1ex	*am(e)*	*(a-)...-mam*	*a-mam*	*a-mam*
pl 2	*méw(e)*	*(me-)…-mi*	*(me-)ni-mi*	*(me-)nó-mi*
pl 3	*si*	*(si-)…-ri*	*(si-)ri-ri*	*(si-)ró-ri*

The system is the same as in Buli, except for vowel alternations in inalienable roots between 1sg/2sg versus 3sg and all plurals.

---

[5] I use ‘edible’ for the category which has also been termed ‘alimentary’ or ‘intimate’. Even though in many languages some non-edible items are also included in this category, the edible items do seem to represent the prototypical meaning associated with this form. The word ‘edible’ should be understood to mean ‘meant for consumption’.

Examples:	inalienable	*mta-g*	‘my eyes’
		mta-m	‘your eyes’
		mtó-mam	‘our (excl) eyes’
	alienable general	*(si)rórі bet*	‘their garden’
	alienable edible	*((y)a-)na-k yof*	‘my sago’

1.3.3 Taba

Table 5: Taba

	Free pronoun	Possessive ligature
sg 1	*yak*	*ni-k ~ nig ~ dik*
sg 2	*au*	*ni-m*
sg 3	*i*	*ni*
pl 1ex	*am*	*amam ~ am*
pl 1in	*tit*	*ni-t*
pl 2	*meu*	*memeu ~ mmeu*
pl 3	*si*	*ni-di ~ di*

All possessed nouns are marked with the possessive ligature.

Examples:	*ni-k wwe*	‘my foot’
	ni sso	‘his name’
	ni-t kabin	‘our goat’
	memeu golo	‘your (pl) snot’
	ni-di calana	‘their trousers’

1.3.4 Magey Matbat

Table 6: Magey Matbat

	Free pronoun	Inalienable	Alienable
sg 1	*ya^{21}ka*	*-ŋ*	*ak-*
sg 2	*ya^{21}wa*	*-m*	*aw-*
sg 3	*i^{21}*	*Ø*	*i-*
pl 1ex	*ya^{21}ma*	*-m*	*am-*
pl 1in	*ya^{21}ta*	*-n*	*at-*
pl 2	*mi^{21}na*	*-m*	*min-*
pl 3	*hafo^{21}*	*Ø*	*ha-*

Inalienable nouns are marked through suffixes, or infixes on what appear to be frozen compounds. Alienable nouns are marked with prefixes. Many inalienable roots show vowel alternations between 3sg/3pl and the rest of the paradigm. Superscript numbers mark tone.

Examples:			
	inalienable	*fa-ŋ*	‘my husband’
		fa-m	‘your husband’
		ni^{3}-ŋ-su	‘my body’ (*ni^{3}su* ‘his/her body’)
		si^{3}-m-bo^{21}m	‘your back’ (*sabo^{21}m* ‘his/her back’)
	alienable	*ak-wa^{3}ŋ*	‘my canoe’
		aw-wa^{3}ŋ	‘your canoe’
		ha-wa^{3}y	‘their child/children’

1.3.5 Irarutu

Table 7: Irarutu

	Free pronoun	Possessive prefix	Possessive suffix
sg 1	*a ~ ja*	*a-*	*-(û)g*
sg 2	*o*	*o-*	*-(û)m*
sg 3	*i*	*i- ~ Ø*	*Ø*
pl 1ex	*am*	*am-*	*-(û)g*
pl 1in	*it*	*it-*	
pl 2	*e*	*e-*	*-(û)m*
pl 3	*ir*	*ir-*	*Ø*

Inalienable nouns are marked through the combination of prefixes and suffixes; alienable nouns through prefixes only. The variants with *û* (a lowered high vowel) occur following consonants.

Examples:			
	inalienable	*a-fa-g*	‘my foot/leg’
		o-rit-ûm	‘your skin’
		(i-)tgra	‘his/her ear’
	alienable	*a-fide*	‘my house’
		am-ju	‘our (ex) canoe’
		ir-damri	‘their light’

1.3.6 Biak

Table 8: Biak

	Inalienable			Alienable
	Non-paired body part	Paired body part	Kinship (sg possessum)	Sg possessum
sg 1	*-ri*	*-si*	(unique forms)	*(a)yedya*
sg 2	*-m-ri*	*-m-si*	*-m(=i)*	*bedya*
sg 3	*-ri*	*-si*	*-r(=i)*	*vyedya*
pl 1ex	*nko-…-s-na*		(not possible)	*(i)nkovedya*
pl 1in	*ko-…-s-na*			*kovedya*
pl 2	*mko-…-s-na*			*mkovedya*
pl 3 animate	*si-…-s-na*			*sedya*
pl 3 inan.	–			*nbedya*

Possession in Biak is extremely complex, as shown by the 25-page discussion in van den Heuvel (2006). This chart is an oversimplification, but it does represent the basic patterns. For inalienables I ignore alternative possibilities of analysis relating to *si* and *ri* and various idiosyncratic patterns of kinship terms. For alienables, I give only one form (for a singular possessum), ignoring the extra forms for dual, trial, plural animate and plural inanimate possessums. I also ignore prepausal variants, in which final *-ya* is replaced by *-i*, e.g. *sedya ~ sedi* 'their'.

Examples:	inalienable (non-paired body parts)	*vru-ri*	'my head; his/her/its head'
		vru-m-ri	'your head'
		mko-vru-s-na	'your (pl) heads'
	inalienable (paired body parts)	*we-si*	'my leg; his/her/its leg'
		ko-we-s-na	'our (incl) legs'
	inalienable (kinship)	*me-m(=i)*	'your cross-uncle'
		kma-r(=i)	'his/her father'
	alienable	*romawa yedya*	'my son'
		snonson vyedya	'his/her name'
		rum sedya	'their house'

1.3.7 Ambai

Table 9: Ambai

	Free	Inalienable	Alienable
sg 1	*yau*	*-ku*	*ne-ku*
sg 2	*wau*	*-mu*	*ne-mu*
sg 3	*i*	*-n ~ -na*	*ne*
pl 1ex	*amea*	*ame-…-mi*	*ame-ne(-mi)*
pl 1in	*tata*	*ta-…-mi*	*ta-ne(-mi)*
pl 2	*mea*	*me-…mi*	*me-ne(-mi)*
pl 3	*ea*	*e-…mi*	*e-ne(-mi)*

Inalienable nouns take a suffix with singular possessors (3sg *-n* on body parts, *-na* on kinship terms) and a circumfix with plural possessors. With alienable nouns an invariant possessive noun *ne* is used.[6] The variants with *-mi* in the plural are only used with body parts treated as alienable nouns. 1s *-ku* is phonetically [hu].

Examples:	inalienable	*nu-ku*	'my head'
		wara-n	'his/her hand'
		ina-na	'his/her mother'
		ta-nu-mi	'our heads'
	alienable	*ne-ku ina*	'my bone'
		ne-mu tarai	'your body'
		ne anteni	'his/her heart'
		ta-ne-mi tarai	'our bodies'
		e-ne munu	'their house'

[6] Since Silzer (1983:123–125) does not give a full paradigm for alienable possession, some of these forms have been extrapolated.

1.3.8 Wandamen

Table 10: Wandamen

	Free pronoun	Inalienable	Alienable
sg 1	*ya ~ yau*	*-ne, -nei, ine*	*ine*
sg 2	*au*	*-mu*	*nomu (~ nemu)*
sg 3	*andi ~ i*	*-ni, -pa*	*nie*
pl 1ex	*ama*	*ama(N)-…-mi*	*amane*
pl 1in	*tata*	*ta(N)-…mi*	*tane*
pl 2	*mia*	*me(N)-…mi*	*mene*
pl 3 an	*sia*	*se(N)-…mi*	*sene*
pl 3 inan	*si ~ asi*	–	*sine*

Inalienable nouns take suffixes in the singular and circumfixes in the plural. The 3sg *-ni* is probably for kinship terms only. Possession on alienable nouns is built around an invariant verbal root *ne*.

Examples:	inalienable	*vara-ne(i)*	‘my hand’
		ru-ne(i), ine ru	‘my head’
		vara-mu	‘your hand’
		vara-pa	‘his/her hand’
		tama-ni	‘his father’
		amam-bara-mi	‘our (excl) hands’
		sen-du-mi	‘their heads’
	alienable	*nomu anio*	‘your house’
		nie anio	‘his house’
		tane wa	‘our canoe’

1.3.9 Mor

Table 11: Mor

	Free pronoun 1	Free pronoun 2	Inalienable
sg 1	*i(g)wa*	*i(g)wo*	*-‘a*
sg 2	*awa*	*awo*	*-ma ~ -mo* (L),*-mu* (A)
sg 3	*i*	*ijo ~ io ~ jo ~ Ø*	*-Ø / -ro*
pl 1ex	*ma’o*	*ma’oro*	*-ma*
pl 1in	*a’o ~ i’o*	*a’oro ~ i’oro*	*-ta*
pl 2	*mo’o*	*mo’oro*	*-mu*
pl 3	*ti’o*	*ti’oro*	*-ri* (L), *-ti* (A)

Inalienable nouns take suffixes. Alienable possession is marked by a following free pronoun, taken from set 2 (these are possibly pronominal enclitics). It is likely that the final *-o* and *-ro* are deictic elements. The forms given by Laycock (L) and Anceaux (A)

differ on some points and possibly represent different dialects.[7] Laycock's article does not contain a single example of an inalienable noun.

Examples:	inalienable	*varea-'a*	'my hand'
		varea-mu	'your hand'
		varea-ti	'their hands'
	alienable	*javar iwo*	'my garden'
		aun i(g)wo	'my dog'
		ate jo	'his father'

1.3.10 Waropen

Table 12: Waropen.

	Free pronoun	Inalienable	Alienable general	Alienable edible
sg 1	*ra*	*ra-*	*rai-*	*ra-na*
sg 2	*auo*	*a-*	*ari-*	*a-na*
sg 3	*i*	*Ø*	*ri-*	*na*
pl 1ex	*ako ~ amo*	*a(N)-* + fortition	*andi-*	*a-na*
pl 1in	*iko*	*Ø* + fortition	*di-*	*na*
pl 2	*mu*	*mi(N)-* + fortition	*mindi-*	*mi-na*
pl 3	*ki*	*ki-*	*kiri-*	*ki-na*

Inalienable possession is marked through prefixes, with fortition in all non-singular forms (except 3pl). Fortition refers to the morphological opposition between the lenis-fortis pairs *f - p, s - t, r - d, w - b* and zero - *k*. Alienable possession distinguishes general possession (in which a form *ri* can be recognised) and edible possession (with *na*). Edible possession is often used in benefactive constructions.

Examples:	inalienable	*ra-worai*	'my head'
		am-borai	'our (excl) heads'
		worai	'his/her/its head'
		borai	'our (incl) heads'
		ki-worai	'their heads'
	alienable general	*rai-ruma*	'my house'
		ari-gha	'your canoe'
	alienable edible	*ra-na sabaku*	'my tobacco'
		na fogha	'her/his pig; our (incl) pig'

[7] During a survey in mid-2007, researchers from SIL-Indonesia took a 239-item wordlist and 20 clauses in four Mor communities (now officially spelled as Moor). Thanks to the SIL Survey Department for sharing the data with me, which confirm the existence of several dialects of Mor. Unfortunately, no new information on possession has become available through this survey. The information on free pronouns in the list from the village of Hariti confirms the free pronouns (set 1) presented here (with the exception of *mu'o* for 2pl); the pronouns in the other three lists were not used as they appear to contain a number of anomalies.

1.3.11 Warembori

Table 13: Warembori.

	Free pronoun	Possessive set 1	Possessive set 2
sg 1	*iwi*	*e-*	*e-*
sg 2	*awi*	*a-*	*a-*
sg 3	*yi*	*i-*	*Ø*
pl 1ex	*ami*	*(ami)*	*(ami)*
pl 1in	*ki*	*ki-, ke-*	*ki-, ke-*
pl 2	*mi*	*mi-, me-*	*mi-, me-*
pl 3	*ti*	*ti-, te-*	*ti-, te-*

The choice between the two possessive sets is lexically governed. Both alienables and inalienables are marked with proclitics/prefixes, but inalienables show a tighter phonological cohesion between the prefix and the noun. On the basis of the noun *bera* 'arm', the 1sg form is *e-vera-ro*, where *b* changes to *v* (/β/); *-ro* is an indicative suffix). With the alienable root *bo* 'coconut', no such lenition occurs. Inalienables also occur with a preposed free pronoun.

Examples:				
	inalienable	*e-vera-ro*	'my arm'	(*bera* 'arm')
		e-voro-ro	'my tongue	(*boro* 'tongue')
	alienable	*e-bo-ro*	'my coconut'	(*bo* 'coconut')
		e-waro	'my canoe'	(*wa(ro)* 'canoe')
		i-mero	'his house'	(*mero* 'house')

Now that the data for these 11 languages has been presented, we are ready to venture into the task of reconstruction. It will be helpful to first look at what has been reconstructed for Proto Austronesian (the parent language for all 1200+ Austronesian languages) and Proto Oceanic, the parent language of the 450+ Oceanic languages and a sister to Proto SHWNG (see Figure 2 in §3). This will help us decide which features are innovations and which are retentions.

1.4 Possession in Austronesian

The possessive system of most Formosan and Western Malayo-Polynesian languages is fairly straightforward and can be summarised by the following three statements, which also hold for Proto Malayo-Polynesian (PMP) and Proto Austronesian (PAn) (see Blust 2005):

- pronominal possession is encoded by possessive enclitics (or suffixes) on all nouns;
- there is no alienable-inalienable distinction;
- the order within the noun phrase is possessed—GENITIVE MARKER—possessor.

Table 14 shows the PAn pronominal system as reconstructed by Ross (2006:532). In order to avoid a plethora of protoforms, I ignore the PAn 'nominative' and 'accusative' pronouns. Third person plural forms present special comparative problems, and following Ross they are left out of consideration here.

Table 14: PAn neutral and genitive pronouns.

	Neutral	GEN1	GEN2	GEN3
1s	**i-aku*	**=[a]ku*	**(=)m-aku*	**n-aku*
2s	**iSu[qu]*	**=Su*	**(=)m-iSu*	**n-iSu*
3s	**s-ia*	*(*=ya)*	…	**n-ia*
1p ex	**i-ami*	**=mi[a]*	*(*(=)m-ami)*	**ni-am, *n-ami*
1p in	**ita*	**=(i)ta*	**(=)m-ita*	**ni-ta*
2p	**i-mu[qu], *i-amu*	**=mu*	**(=)m-amu*	**ni-mu, *n-amu*

In spite of the apparent complexity of three sets of genitive forms, the possessive system in PAn was essentially fairly basic. The set called GEN1 was the unmarked set which was retained in PMP; it also served as the syntactic pivot, an issue which does not concern us here. GEN2 is a non-third person set only found in some Formosan languages, while GEN3 consists of the genitive personal phrase marker **ni*, followed by the free set. There were two other genitive phrase markers: **na* (for common nouns present) and **nu* (for common nouns absent).

The main points (possessive suffixes, no alienable-inalienable distinction, possessor following the head noun) are illustrated in the following examples from Muna, a Western Malayo-Polynesian language spoken in Southeast Sulawesi (see van den Berg 1989; notice that in sentence (2) a nominal possessor is obligatorily indexed on the head noun by means of the 3sg possessive suffix *-no*).

(1) a. *ama-ku*
father-1s
'my father'

b. *fotu-mu*
head-2s
'your head'

c. *lambu-no*
house-3s
'his/her/its house'

(2) *lambu-no* *ama-ku*
house-3s father-1s
'my father's house.'

When we move to Oceanic languages, the system is much more complex. For Proto Oceanic Lynch, Ross and Crowley (2002) reconstruct the following system:

- the possessive system makes a distinction between inalienable and alienable nouns;
- inalienable nouns (body parts, kinship terms) take direct possession, that is, they are affixed with possessive suffixes as shown in Table 15;
- alienable nouns take indirect possession, that is, they require prenominal possessive classifiers which in turn take possessive suffixes;
- three such classifiers are reconstructed: **ka* 'food'; **m(^{w})a* 'drink' and **na* 'general'; possibly there were other classifiers;
- the possessor followed the possessed item.

Table 15: POC personal pronouns.

	Independent	Possessive suffix
1s	**[i]au*	**-gu*
2s	**[i]ko[e]*	**-mu*
3s	**ia*	**-ña*
1p ex	**ka[m]i, *kamami*	**-ma[m]i*
1p in	**kita*	**-da*
2p	**ka[m]u, *kamiu*	**-m[i]u*
3p	**[k]ira*	**-dra*

The following examples from Vitu illustrate these points. Vitu is a conservative Oceanic language belonging to the Meso-Melanesian linkage, spoken on islands northwest of New Britain, Papua New Guinea (PCF is possessive classifier for food items):

(3) a. *tama-gu*
father-1s
'my father'

b. *baka-a*
head-2s
'your head'

c. *ka-gu ruma*
PC-1s house
'my house'

d. *ha-na beti*
PCF-3s banana
'his/her bananas'

(4) *ruma ka-na kaua*
house PC-3s dog
'a dog's house'

In (3a) and (b) we see direct possession with the inalienable nouns *tama* 'father' and *baka* 'head'. In (3c) and (d) we see indirect possession with the alienable nouns *ruma* 'house' and *beti* 'banana', using the classifiers *ka* for general possession and *ha* (phonetically [ɣa]) for edible possession.[8] Example (4) illustrates a full possessor following the possessed noun. The reverse order, though less frequent, is also possible in Vitu. For further details, see van den Berg and Bachet (2006).

2 Six questions

Having outlined the reconstructed systems for both PAn and POC, we are now in a better position to look at some of the unusual systems displayed in the SHWNG group and arrive at a well-founded reconstruction. I will deal with the six questions posed in §1.2 in turn.

2.1 Do all SHWNG languages mark an alienable-inalienable distinction?

With one exception all the SHWNG for which data are available do indeed make this distinction, so that it can be safely reconstructed for Proto SHWNG. The one exception is Taba, where there are no separate forms.[9] The question can be raised whether Taba represents a relic (implying Proto SHWNG did not have the distinction) or whether it is

[8] The edible classifier *ha* in Vitu is a regular derivation from POC **ka*, but the general classifier *ka* is an innovation in Meso-Melanesian (Ross 1988:271).

[9] Bowden (2001:233–234) points out that certain part-whole relationships have an obligatory possessor. This leads Klamer et al. (2007) to say 'In Taba this means that inalienables have obligatory expression of the possessive relationship', but I doubt whether this is a correct reinterpretation. The notions of (in)alienability and obligatory possession are in principle independent parameters.

highly innovative and has lost the distinction. The major clue to the answer is that the only possessive strategy employed in Taba is the so-called 'indirect possession', i.e. the use of a possessive morpheme or ligature *ni-* to which possessive suffixes are attached (see Table 5 in §1.3.3),[10] as with the following nouns (which are usually inalienable in other languages):

(5) *yak* *ni-k* *mapin*
1s POSS-1s woman
'my wife'

(6) *ni-m* *mto*
POSS-2s eye
'your eyes'

(7) *ni-Ø* *mtu*
POSS child
'his/her child'

The point that needs to be made is that this situation can best be explained if we assume that Taba lost the distinction between alienable and inalienable possession by generalising the indirect possession strategy to all nouns. The presence of *ni* in the language unambiguously points to the existence of an earlier dichotomy between alienable and inalienable possession, since other South Halmahera languages use *ni* as well. If Taba had been a very conservative language, it would almost certainly have retained direct possession on nouns. However, contrary to the general pattern in the area, Taba has collapsed the distinction. A merger scenario is confirmed by data on the Tahane dialect of Taba (Collins 1982:101), where the alienable-inalienable distinction seems to hold for only 1st and 2nd person singular. The forms given are *mta-g* 'my eye'; *nim mta-m* 'your eye' and *nim mta* 'his/her/its eye'.[11] It is likely that Taba lost the distinction under the influence of neighbouring West Papuan languages. While it is true that most of the West Papuan languages on the New Guinea mainland have an alienable-inalienable distinction (which may be responsible for the appearance of this distinction in Austronesian in the first place), most of the North Halmahera languages appear to lack the distinction, including Tobelo (Gary Holton pers. comm.), Tabaru (Ed Kotynski pers. comm.) and Sahu (John Severn pers. comm.). It is true that a number of nouns referring to kin terms and body parts are obligatorily possessed in some of these languages, but the morphosyntax of possession on these nouns is in no way different from that of other nouns. According to Gary Holton (pers. comm.), the inalienable/alienable distinction is either not present or weakly present in North Halmahera. Taba appears to have followed this pattern.

It should also be pointed out that in Wandamen the distinction between alienable and inalienable possession has broken down for 1sg. Again, this is almost certainly a recent innovation. In conclusion, it is virtually certain that Proto SHWNG had both alienable and inalienable possession.

---

[10] First person plural exclusive and second person plural are exceptional in that they do not use *ni-* but normally reduplicate the pronoun: *amam* (variant *am*) '1pe' and *memeu* (variants *mmeu, meu*) '2pl' Given the variation for 3pl (*ni-di* ~ *di*), it seems reasonable to assume that *ni-* was present in all these forms in an earlier stage of the language

[11] The *-m* on *nim* for 3sg is puzzling.

2.2 How is inalienable possession encoded in the SHWNG languages?

The SHWNG languages show a bewildering variety of forms and positions for inalienable possession. Table 16 summarises the various possibilities for inalienable possession discussed and illustrated in §1.3. It also gives the illustrative example 'head', if available, or another body part. I use the term prefix to include proclitics.

Table 16: Inalienable possessive strategies in selected SHWNG languages.

	Singular	'my head'	Non-singular	'our (incl) heads'
Buli	prefix + suffix	*ya-boboko-k*	prefix + suffix	*it-boboko-r*
Sawai	optional prefix + suffix	*béboko-g* ~ *(y)a-béboko-g*	optional prefix + suffix	*béboko-r* ~ *it-béboko-r*
Taba	suffix on possessive ligature *ni*	*ni-k wwe* 'my foot'	suffix on possessive ligature *ni*[12]	*ni-t wwe* 'our feet'
Magey Matbat	suffix or infix	*ni^{3}ŋsu* 'my body'	suffix or infix	*ni^{3}nsu* 'our bodies'
Irarutu	prefix +suffix	*a-rgun-ûg*	prefix + suffix	*it-rgun-ûg*
Biak	suffix	*vru-ri*	prefix + number suffix + suffix *-na*	*ko-vru-s-na*
Ambai	suffix	*nu-ku*	prefix + suffix *-mi*	*ta-nu-mi*
Wandamen	suffix	*ru-ne(i)*	prefix + suffix *-mi*	*taŋ-du-mi*
Mor	suffix	*varea-'a* 'my hand'	suffix	*varea-ta* 'our hands'
Waropen	prefix	*ra-worai*	prefix (with C fortition)	*Ø-borai*
Warembori	prefix (with cohesion)	*e-vera-ro* 'my arm'	prefix (with cohesion)	*ke-vera-ro* 'our arms'

From Table 16 it is immediately clear that the typological variation encountered within SHWNG is quite large. Not only are there simple possessive prefixes and suffixes, but also circumfixes and infixes. Several languages have an invariant element for the plural, and others show rare phonological phenomena.

As Klamer et al. (2007) have noted, one can observe a general drift towards possessive prefixation in eastern Indonesia, which probably arose under the influence of neighbouring Papuan languages which are prefixing. If this scenario is correct, suffixation is older than prefixation, and Mor is thus the most conservative of the SHWNG languages—at least on this parameter—having only suffixes. Waropen and Warembori, on the other hand, are the most innovative, having only prefixes. The other languages show combinations of pre- and suffixes, and various kinds of paradigm splits. It is interesting to observe that the singular forms are more conservative than the non-singulars, with suffixing only still occurring in Ambai and Biak singulars. The second person singular has been particularly resistant.

The following rough scale can be set up

[12] See fn.10 for further clarification.

Stage 1	sg:	only suffixes	Mor, Matbat
	pl:	only suffixes	
Stage 2	sg:	only suffixes	Biak, Wandamen, Ambai
	pl:	prefixes and suffixes	
Stage 3	sg:	optional prefixes and suffixes	Sawai
	pl:	prefixes and suffixes	
Stage 4:	sg:	obligatory prefixes and suffixes	Buli, Irarutu
	pl:	obligatory prefixes and suffixes	
Stage 5:	sg:	only prefixes	Waropen, Warembori
	pl:	only prefixes	

This chart is again an oversimplification. Taba does not fit at all, since it has no separate inalienable category. Also, the third person singular is sometimes structurally different from first and second. Thirdly, the suffix component in the later stages shows considerable variety. Wandamen, Biak and Ambai (as well as the Raja Ampat language Ambel) have an invariant possessive suffix (*-ni* in Ambel, *-na* in Biak, *-mi* in Ambai and Wandamen—possibly a generalisation of the 2nd person plural), while Buli and Irarutu, both at stage 4, have retained old suffixes. Buli has retained the original plural suffixes, while Irarutu has generalised the 1st person sg suffix *-(û)g* to cover all first persons, and the 2nd person singular suffix *-(û)m* to cover all second persons.

Anceaux (1961:157–162) provides some additional information on other languages from Yapen and the Cenderawasih Bay area, although his data rarely extend beyond a single inalienable paradigm. Kurudu only uses prefixes and is therefore at stage 5 (*aywekampe* 'my hand', *tawekampe* 'our hands'; '*-pe* seems to indicate definiteness'). With a combination of prefixes and suffixes Pom is at stage 3 (*warani* 'my hand', *tambarami* 'our hands'). Ansus shows a 'curious collection of forms' for 'hand' (including *varau* 'my hand', *tambaraminekuira* 'our hands'), but appears to be at stage 2. For all these languages one should of course be very cautious about drawing conclusions based on a single paradigm.

The main point, however, is clear: the general direction of the morphological drift is unmistakenly towards prefixation. Given such conservative languages as Mor, and given the fact that Proto Austronesian used suffixation for possession and that Proto Oceanic is reconstructed with suffixes for inalienable nouns, we can safely assume that Proto SHWNG had possessive suffixes for all persons with inalienable nouns. I therefore reconstruct Proto SHWNG with suffixes on inalienable nouns.

2.3 What inalienable forms can be reconstructed for Proto SHWNG?

It is now time to look at the actual forms and propose formal reconstructions. I will in turn look at South Halmahera, Raja Ampat and West New Guinea.

The South Halmahera languages constitute a distinct subgroup with relatively little variation. All the forms in Table 17 are taken from the inalienable nouns, as well as from the possessive classifiers. When these differ, as with Maba and Sawai in 1sg, they are separated by a comma. Taba forms are simply taken from the 'possessive ligatures'. The data from Maba, identical to Buli, is taken from Adriani and Kruyt (1914:335–338). For Giman, Patani and Gebe no information on possessive marking has appeared in print, but the languages appear to be sufficiently similar to warrant the reconstruction.

Table 17: Inalienable possession in SH languages

	Proto South Halmahera	Taba	Buli	Sawai	Maba
sg 1	*-*k*	*-k ~ -g*	*-k*	*-g, -k*	*-g, -k*
sg 2	*-*m*	*-m*	*-m*	*-m*	*-m*
sg 3	*Ø*	*Ø*	*Ø*	*Ø*	*Ø*
pl 1ex	*-*mam*	*amam ~ am*	*-mam*	*-mam*	*-mam*
pl 1in	*-*t*	*-t*	*-r*	*-r*	*-r*
pl 2	*-*miu*	*memeu ~ mmeu ~ meu*	*-meu*	*-mi*	*-meu*
pl 3	*-*ri*	*-di*	*-ri*	*-ri*	*-ri*

Comments

- 1sg *-*k*. Several languages show sporadic voicing to *-g*. Maba and Sawai have *-k* on the classifiers (*a-ni-k* and *a-na-k* 'my'), but *-g* on inalienable nouns (*mta-g* 'my eye'). An alternative solution is to reconstruct *-*gu* (with sporadic loss of voicing; cf. POC *-*gu*), but since PMP also had *-*ku*, it makes more sense that the voicing started in the SHWNG group.
- 1pl inclusive *-*t*. Although most languages have *-r*, external witnesses favour *-*t* (PMP *-*ta*, POC *-*da*). The little-known language of Gebe also has *-t* as in *itne-kuto-t* 'our head' (Maan 1951:53, fn.64).
- 2pl *-*miu*. Taba, Buli and Maba show sporadic lowering of **i* to *e*. Sawai *-mi* has lost final **u*.
- 3pl *-*ri*. Taba has irregular fortition to *-di*.
- As explained in §2.1 (fn.10), Taba has innovated by using simple or reduplicated pronouns for possession in 1pl ex and 2pl.

Moving now to the Raja Ampat languages, Table 18 presents the inalienable suffixes.

Table 18: Inalienable possession in RA languages

	Proto Raja Ampat	Magey Matbat	Ambel
sg 1	*-*k*	*-ng*	*-k*
sg 2	*-*m*	*-m*	*-m*
sg 3	*Ø*	*Ø*	*Ø*
pl 1ex	?	*-m*	*am-…-ni*
pl 1in	*-*ni* ?	*-n*	*-ni*
pl 2	?	*-m*	*mem-…-ni*
pl 3	*-*ni* ?	*Ø*	*-ni*

Notice that Matbat has a reduced system; it has collapsed 3sg and 3pl (both Ø), as well as 2sg, 2pl and 1pl exclusive (all *-m*). The attempt at reconstruction is somewhat shaky, as next to Magey Matbat I only have Ambel data at my disposal, from Remijsen (2001).[13] For Ambel only one noun is given (*taji* 'eye')—which makes generalising very hazardous—

[13] I was unable to access A.C. van der Leeden's work on Ma'ya, another Raja Ampat language. Almost all of the information on Ma'ya morphology and syntax, however, is contained in unfinished and unpublished manuscripts (see references in Remijssen 2001).

while in Matbat the six inalienable nouns seem to belong to various form classes, the details of which I ignore here. The singular paradigm is fairly clear, however, with Matbat 1sg *-ng* obviously an innovation, but for the original plural forms I can only make educated guesses at this stage and leave some blanks. One possible scenario is that there was an original 3pl suffix **-ni* (possibly from **-di* < **-ri*; compare Proto SH **-ri*), which was also used for 1pl inclusive. Such an alignment may appear unlikely, but a parallel is found in Hatam, a Papuan language of the eastern Bird's Head, where the possessive pronoun *i-de* is used for both 3pl and 1pl inclusive (Reesink 1999:40–41), as well as in Muna (Sulawesi), where the verbal prefix *do-* is also used for both 3pl and 1pl inclusive. Matbat lost the final *-i* of *-ni* and generalised the 3sg form Ø to 3pl, thereby obliterating the original meaning of *-ni*. As mentioned above, it also generalised 2sg *-m* to 2pl and 1pl ex (though it is just possible that 1pl ex *-m* retains an older form, cf. Proto SH **-mam*). Ambel in turn generalised *-ni* as a plural possessive marker, using the verbal prefixes *am-* and *mem-* to distinguish the remaining forms. Such general possessive markers for the plural are also found in Ambai and Waropen, albeit in different forms.

The infixation in Matbat forms such as *ni*^{3}*-ŋ-su* 'my body' (cf. *ni*^{3}*su* 'his/her body') is best explained as suffixation on the first part of frozen compounds, a situation which is parallelled in Irarutu; see van den Berg and Matsumura (2008).

The last subgroup to be discussed are the WNG languages, laid out in Table 19.

Table 19: Inalienable possession in WNG languages

	Proto WNG	Irarutu	Biak	Ambai	Wandamen	Mor	Waropen	Warembori
sg 1	**-ku*	*a-… -(û)g*	*-ri*	*-ku*	*Ø ~ -ne(i) ~ ine*	*-'a*	*ra-*	*e-*
sg 2	**-mu*	*o-… -(û)m*	*-m-ri*	*-mu*	*-mu*	*-mu* (A) *-ma ~ -mo* (L)	*a-*	*a-*
sg 3	**-na*	*i-*	*-ri*	*-n ~ -na*	*-ni ~ -pa*	*Ø ~ -ro*	*Ø*	*i-, Ø*
pl 1ex	**-ma*	*am-… -(û)g*	*nko-… -s-na*	*ame-… -mi*	*ama(N)-… -mi*	*-ma*	*a(N)-* + fortition	*ami-*
pl 1in	**-ta*	*it-… -(û)g*	*ko-… -s-na*	*ta-… -mi*	*ta(N)-…-mi*	*-ta*	*Ø* + fortition	*ki-, ke-*
pl 2	**-mi*	*e-… -(û)m*	*mko-… -s-na*	*me-… -mi*	*me(N)-…-mi*	*-mu*	*mi(N)-* + fortition	*mi-, me-*
pl 3	**-ri* ?	*ir-*	*si-… -s-na*	*e-…-mi*	*se(N)-…-mi*	*-ri*	*ki-*	*ti-, te-*

With the exception of Mor, all the WNG languages use prefixes or a combination of prefixes and suffixes. Prefixes appear to have two sources. For some languages they are almost always identical to verbal prefixes (Biak, Waropen, Warembori), though these have not been presented here. In other languages, the prefixes are almost always identical to the free pronouns and probably arose as proclitic pronouns (Irarutu).

In the following paragraphs I will briefly defend my reconstructions and try to account for non-cognate forms. It will be clear that many of the reconstructions have been made with a knowledge of the PMP and POC forms in mind.

1. First person singular *-*ku*. The reconstruction is based on Ambai (where *-ku* is phonetically [hu]), as well as external evidence from PMP and POC. Irarutu *a-* is the 1sg proclitic on verbs as well as the short form of the pronoun, but the suffix *-(û)g* retains the original form with voicing. The vowel in the variant *-(û)g* (which follows consonant-final roots) is either an epenthetic vowel (though Irarutu allows for complex consonant clusters) or—more likely—metathesis of earlier **-gu*, with unexplained lowering of *u* to *û*. Biak *-ri* (also used for 3sg) is a possessive element which occurs in many SHWNG languages. Below I will argue this was originally probably a classifier. Wandamen *-ne* and *-nei* are deictic elements meaning 'this (near speaker)', which may apparently function as possessives. Many WNG languages have an elaborate system of deictics and demonstratives which are obligatory on nouns and noun phrases (cf. Mor 3sg *-ro*). Wandamen *ine* is originally an alienable form based on the verb *ne* (see §2.4 below). Since the distinction between alienable and inalienable possession is lost or blurred in Wandamen for 1sg, both these forms can be used. Mor *-'a* is unusual, in that Mor is highly conservative, but obviously not in 1sg. It is probable that the suffix is simply the reflex of the free pronoun *iwa ~ igwa*, itself a fortition of **ia or* **ya*. Waropen *ra-* is a fortition of *ya-* (*ya* 'I' is the free pronoun in the Napan dialect) and really the verbal prefix. The origin of Warembori *e-*, also the 1sg verbal prefix, is somewhat unclear, but probably arose as a contraction from earlier **ya-*.

2. Second person singular *-*mu*. The form **-mu* is reconstructed on the basis of this form in Mor (Anceaux's data), Wandamen, Ambai, Irarutu *-(û)m* (after consonant-final roots and parallelling 1sg) as well as the reduced *-m* in Biak *-m-ri*. Waropen and Warembori *a-* are again verbal prefixes, probably reductions of the free pronoun *auo* (Waropen) or *awi* (Wandamen), both going back to **aw*.

3. Third person singular *-*na*. The form **-na* is reconstructed on the basis of Ambai, which has *-na* for 3sg kinship terms (but *-n* for body parts). Wandamen *-ni* contains a deictic element *-i*, while *-pa* is indeed a deictic. Mor *-ro* is almost certainly also a deictic or an indicative element (cf. Warembori *-ro* 'singular indicative'). Irarutu *i* is the verbal prefix, as well as the free pronoun. Similar to the SH and the RA languages, Waropen, Mor and Warembori have a zero form for 3sg. For Biak *-ri*, see the comments under 1sg.

4. First person plural exclusive *-*ma*. The form *-ma* is reconstructed on the basis of Mor *-ma*. All other languages have innovated, usually prefixing the verbal prefix or a pronominal clitic to the inalienable noun, sometimes generalising from 1sg (Irarutu), or using a generalised plural possessive suffix. The origin of the Waropen fortition remains unclear at this point.

5. First person plural inclusive *-*ta*. Again Mor is the most conservative, while the other languages have prefixed reduced pronouns to the noun.

6. Second person plural *-*mi*. This form is reconstructed on the basis of the general possessive marker *-mi* in Wandamen and Ambai, very likely a relic form of the 2pl suffix which has been generalised to all non-singular forms. Mor *-mu* is then a generalisation from 2sg. Alternatively, the reconstruction should be **-miu*, which has gone to either *-mi* or *-mu*.

7. Third person plural **-ri.* This is again tentative, based on Mor *-ri* and the SH and RA evidence. The forms *si-* (Biak) and *se-* (Wandamen) go back to a verbal prefix *si-*, which is probably also true for Warembori *ti- ~ te-* and Ambai *e-* (possibly through **si* > **hi* > **he* > *e*; cf. the free pronoun *ea* 'they' < PMP **sida*). Waropen *ki-* is probably an irregular development from *ti-* (the change *t* to *k* is fairly common in SHWNG).

I now present the reconstructed forms for these three languages groups together in Table 20. Column 4 gives the reconstructed Proto SHWNG forms on the basis of these three subgroups, of which SH and RA are closely linked. For the sake of comparison I also give the POC forms (repeated from Table 15) and my own reconstruction of the suffixes in Proto EMP. Four of the six forms for Proto EMP are uncontroversial, but the forms for 1sg and 1pl inclusive warrant some comments. I reconstruct **-ku* and **-ta*, rather than their voiced variants **-gu* and **-da* which have been posited for POC), mainly on the basis of Proto SHWNG **-ku* and **-ta*. I assume that the POC voiced forms are best explained as due to sporadic voicing, noting that this voicing can also be observed in several SHWNG languages (including Buli and Irarutu). By the time the speakers of POC (or their ancestors, speakers of pre-POC) left their eastern Indonesian location somewhere around the Bird's Head, this sporadic sound change had firmly established itself.

Table 20: Reconstruction of inalienable suffixes.

	1	2	3	4	5	6
	Proto South Halmahera	Proto Raja Ampat	Proto West New Guinea	Proto SHWNG	Proto Oceanic	Proto EMP
sg 1	**-k*	**-k*	**-ku*	**-ku*	**-gu*	**-ku*
sg 2	**-m*	**-m*	**-mu*	**-mu*	**-mu*	**-mu*
sg 3	*Ø*	*Ø*	**-na*	**-na*	**-ña*	**-ña*
pl 1ex	**-mam*	?	**-ma*	**-mam*	**-ma[m]i*	**-mami*
pl 1in	**-t*	**-ni* ?	**-ta*	**-ta*	**-da*	**-ta*
pl 2	**-miu*	?	**-mi*	**-miu*	**-m[i]u*	**-miu*
pl 3	**-ri*	**-ni* ?	**-ri* ?	**-ri* ?	**-dra*	?

What is surprising about this chart is how conservative Proto SHWNG now appears to be. With the exception of 3pl **-ri* (which is uncertain), all the forms are clearly derived from their PMP cognates and very similar to POC.

2.4 How is alienable possession encoded? And what can be reconstructed for Proto SHWNG?

Table 21 summarises the encoding of alienable possession.

Table 21: Alienable possessive strategies in SHWNG

	How many alienable categories?	Which morphemes?
Buli	2	a. classifier *ni* b. classifier *na*
Sawai	2	a. classifier *ni* b. classifier *na/no*
Taba	(no alienable–inalienable distinction)	–
Magey Matbat	1	possessive prefix
Irarutu	1	possessive prefix
Biak	1	inflected verb *ve* with possessor suffix and specifier
Ambai	1	possessive noun *ne*
Wandamen	1	inflected verb *ne*
Mor	1 ?	separate set of pronouns
Waropen	2	a. set of prefixes b. possessive noun *na*
Warembori	1	possessive prefix

With the exception of Taba, languages are split as to whether there is a single or a double category of alienable encoding. Buli and Sawai have two clearly definable possessive classifiers. Waropen also has two possessive alienable strategies, but the other languages have just one. In each case, the morphosyntactic strategy is unique. Irarutu, Matbat and Warembori simply use prefixes for alienable possession. Mor (for which data is extremely limited) appears to employ a separate set of pronouns, e.g. *ate awo* 'your father' (cf. the free pronoun *awa* 'you' and the inalienable noun *warea-mu* 'your hand'). It is likely that *awo* is a reduction of *awa-o*, with *o* as a deictic marker, in which case alienable possession is simply marked by independent pronouns. Ambai uses an 'auxiliary possessive noun' *ne*, while Wandamen and Biak inflect a possessive verb, *ne* and *ve* respectively. As mentioned earlier, Biak is extremely complex as it also suffixes this inflected verb with a possessum agreement marker and a specifier.

Given this bewildering variety, what should be reconstructed for Proto SHWNG? Did it have one or two inalienable strategies? And how was the distinction morphologically encoded? Given the variety of systems in use, this aspect of the reconstruction is perhaps the most tentative. My hypotheses are as follows.

There were two alienable strategies in Proto SHWNG, one for general possession and one for edible possession. The fact that two strategies are found at either end of the SHWNG language area—Buli and Sawai in South Halmahera as well as Waropen on the east coast of the Cenderawasih Bay makes it highly likely that this was an inherited feature of Proto SHWNG. Loss of the distinction in individual languages is more easily accounted for than its emergence at extreme points of the group. The fact that Taba has only one possessive strategy shows that such loss can happen relatively quickly in a fairly tight subgroup. The fact that Proto Oceanic is reconstructed with at least two possessive classifiers is corroborating evidence for this. It shows that the distinction already existed by the time of Proto EMP (see also §3).

I propose the following two possessive classifiers for Proto SHWNG: **na* 'classifier for edible possession' and (tentatively) **ri* 'classifier for general possession'. These two classifiers took the set of possessive suffixes, as they still do in e.g. Buli. The reconstruction of **na* is fairly straightforward, given its occurrence in Buli, Sawai and Waropen, where it is used for food, tools and certain kinship terms (see discussion in §2.5). Positing **ri*, however, is more tentative. I mainly reconstruct it on the basis of its occurrence in South Halmahera and in Waropen, where the form has undergone some changes but where its function has been retained, as well as scattered occurrences of *ri* in other languages of the subgroup, including Irarutu and Biak. Specifically, Waropen uses verbal prefixes on the classifier *ri* (e.g. *a-ri gha* 'your canoe') instead of the proposed reconstruction **ri-mu* with suffixes. Also, some of the forms have become obscured through sound change. The form *rai* 'my', for instance, probably originates from **ya-ri* via **ra-yi* or **ra-ri*). For the South Halmahera languages we have to posit an irregular change **ri* > *ni*, resulting in the classifier pair *na* and *ni*. Buli forms such as 1pl inclusive *(ite-)ri-r* and 3pl *(si-)ri-ri* have retained the original *ri* or represent later assimilations.[14] In Taba *ni* became the sole marker of possession, even incorporating the inalienable category. In Irarutu *na* was lost and *ri* was retained for emphatic possession, again with prefixes (see van den Berg and Matsumura to appear). Wandamen and Ambai similarly lost the distinction between *na* and *ri*, merging them as *ne*. In Ambai *ne* is still a possessive noun with suffixes (e.g. *ne-mu tarai* 'your body'), but in Wandamen this *ne* has been e analyzed as a verb (comparable to the Biak possessive verb *ve*) taking verbal prefixes, and an infix *i* for the 3sg: *i-ne anio* 'my house', *n<i>e anio* 'his/her house'.[15] Mor seems to have lost *ri*, though more data could certainly reveal it still exists. Biak has a form *ri* in some forms of its paradigm, but surprisingly it marks inalienable possession on unpaired body parts (*vru-ri* 'my head' or 'his/her/its head', *vru-m-ri* 'your head'), as well as 3sg inalienable possession on kinship terms: *kma-ri* 'his/her father'.[16] Because of the phonological and semantic correspondence, it is highly likely that Biak *ri* is related to the possessive classifier **ri* which I reconstruct for Proto SHWNG, but its syntactic behaviour as a singular suffix on just a limited set of inalienable nouns is so aberrant that explaining its diachronic development is a real challenge.

These two classifiers acted as nouns which were affixed with the possessive suffixes. Some languages also allow for prefixes to occur on these classifiers (e.g. the long forms of Buli), but this was no doubt a later development, on the analogy of the inalienable nouns.

Notice that Biak has a homophonous morpheme *ri* which has an anaphoric function (van den Heuvel 2006:320). Interestingly, the nearby Papuan language Hatam also has a marker *ri*, roughly indicating 'givenness' (Reesink 1999:66). The usage of *ri* in Biak and Hatam appears to be rather different from the possessive constructions, and it is therefore unclear to me whether these two morphemes (seemingly related to each other) are connected to or can shed light on the history of the possessive *ri* in SHWNG.

---

[14] I am aware of an alternative explanation in which the classifier *ni* in Buli and Sawai is simply not cognate with the forms in Waropen. Buli forms in *ri* such as *(ite)ri-r* and *(si)-ri-ri* simply arose through assimilation preceding a suffix with initial *r*. My proposal, though tentative, is neither phonetically nor morphologically implausible and offers a reasonable (and in my opinion much more interesting) alternative.

[15] Wandamen 2sg shows quite some variation in the sources: *nomu* (Henning, Ranar et al., my own field notes), *nemu* (Anceaux), *nue* (Cowan) and *memu* (Cowan, basing himself on published folk tales). *Nemu* must be the oldest form, *nomu* arose by vowel assimilation, *memu* by consonant assimilation, while *nue* is the regular 2sg verbal inflection, with *u* infixed. It is striking that *nomu* is the current form.

[16] *Kma-ri* can also be analysed as *kma-r-i* 'father-POS-3s'.

2.5 What was the semantic basis of alienable and inalienable possession?

Having established the existence of two alienable classifiers, the next question is the semantic basis of the reconstructed system. Which nouns were inalienably possessed, which nouns took the classifier **na* and which ones took **ri*? Table 22 shows the semantic alignment for each of the eleven languages. When nouns are listed as exceptions to the inalienable category, they take general possession.

Table 22: Semantic alignment between possessive strategies.

	Inalienable possession	Exceptions to the inalienable category	Edible possession	General possession
Buli	a. most body parts, including bone, flesh b. kinship terms (father, mother, parent, brother, sister, child, grandchild) c. others: fruit, name	a. hair, eyebrows, nail b. married kin	a. food and drink items b. a few tools	all other nouns (objects, tools)
Sawai	a. most body parts b. kinship terms (siblings, descendants, parents-in-law) c. parts of whole (only 3s)	a. body liquids (e.g. blood, pus, sweat), eyebrows, breath, bone; body parts of animals b. father, mother, grandparent, aunt, uncle, brother-in-law	a. edible items b. hand implements (plate, sago pounder) c. garden	a. several body parts b. several kinship terms c. all other nouns
Taba	–	–	–	
Magey Matbat	a. most body parts b. kinship terms (including husband, mother-in-law)	a. 'head' b. 'father'	-	all other nouns
Irarutu	a. most body parts b. 8 kinship terms (cross-sibling, younger sibling, older sibling, spouse, child-in-law, parent-in-law, grandchild, friend/brother')	a. lung, brains, throat, vein, breast, blood b. most kinship terms (incl. father, mother, child, aunt, uncle)	–	all other nouns
Mor	a. kinship terms b. ?	?	–	all other nouns
Biak	a. some body parts (leg, arm, belly-area, head, penis, vagina, mouth, tooth, ear, nose, eye, neck) b. some kinship terms (e.g. father, mother, spouse, grandparent/grandchild, cross-uncle, cross-aunt, cross-sibling, cross-cousin, cross-sibling's child) c. 4 locational nouns (middle, inside, upside, downside)	a. forehead, back-head, jaw, chin, throat, tongue, head-hair, moustache, tongue, joint, vein, buttocks, breast. Also: all compounded body parts b. some kinship terms (e.g. great-grandparent, son, daughter, parellel sibling)	–	all other nouns
Ambai	a. some body parts b. some kinship terms	a. some body parts b. some kinship terms	–	all other nouns

Wandamen	a. some body parts (head, hand, leg, belly) b. some kinship terms (father, mother, parent-in-law, child, grandparent, grandchild, wife, husband)	a. body parts ? b. older sibling, parent	–	a. all other nouns b. all other body parts
Waropen	a. most body parts b. 4 kinship terms (father, mother, grandparent, sibling-in-law)	a. breast b. all other kinship terms	a. food and drink b. tools (knife) c. chief and slave d. also allows for all kin terms	all other nouns
Warembori	a. all body parts b. all kinship terms c. house, canoe, net bag			all other nouns

The distribution of the various nominal categories is far from uniform, which makes the reconstruction somewhat problematic. I tentatively propose the following division.

1 Inalienable possession (marked by direct possession on the noun):

a. Most body parts. Exceptions were probably body liquids (blood, pus, sweat), internal organs (bone, lungs, heart), and removable items such as hair, eyebrows and nails.

b. Some kinship terms. With the possible exception of Warembori, all of the languages divide the kinship terms between inalienable possession and general possession, though note that Waropen has the extra option of encoding them all with 'edible' possession, and that information on Wandamen and Ambai is limited. The organising principle behind this division in the individual languages is far from clear. It seems that there was (and is) a general drift away from inalienable possession. Are there cultural factors at play such that married kin are not marked as inalienable in Buli? Why is 'father' no longer inalienable in Matbat? In some cases I suspect that the syntactic status of a word may have been a factor. In Biak, for instance, *romawa ~ roma* 'son' and *inai* 'daughter' are alienable forms, probably because they also (and originally) mean 'boy, child' and 'young girl' respectively. In Wandamen, *baba* 'parent' is an alienable noun, presumably because it is based on the adjective/stative verb *baba* 'big'.

c. Locative nouns. Since this is mentioned for Biak and hinted at in several other languages, it is likely that at least some locational nouns were treated as inalienable nouns.

d. The word for 'name'.

2 Alienable 'edible' possession (marked by suffixation on the classifier **na*)

a. Food and drink items.

b. Tools, probably originally only tools used to procure or consume food.

All three languages that have the edible category agree on these two points. Waropen has expanded the category to include all kinship terms (clearly an innovation), and by extension the words for 'chief' and 'slave'.

3. Alienable 'general' possession (marked by suffixation on the tentative classifier **ri*). All other nouns went into this category.

To summarise, even though the semantic details are not entirely clear, nouns were clearly divided into three categories following the schema outlined in Figure 1, which illustrates 1sg.

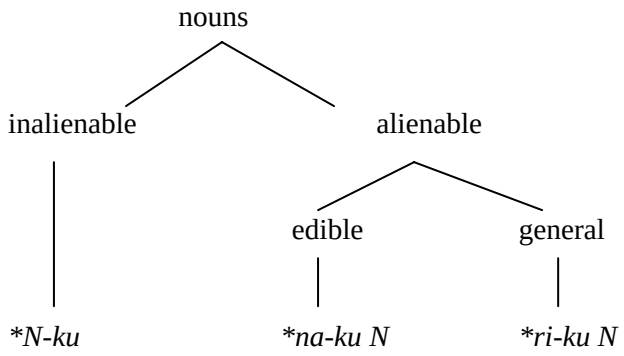


Figure 1: Noun division in Proto SHWNG

2.6 What was the order of the possessor versus the possessed?

The last question concerns the position of the possessor vis-à-vis the possessed head noun. Eastern Indonesia is well-known as the home of the so-called 'reverse genitive', a preposed possessor within the noun phrase. Table 23 summarises the situation for the SHWNG languages. Space prevents me from giving representative examples.

Table 23: Position of possessor

	Preceding possessed	Following possessed	Comment
Buli	yes	no	
Sawai	yes	no	
Taba	yes	no	
Matbat	yes	no	
Irarutu	yes	no	
Biak	yes	yes	
Ambai	yes	no	
Wandamen	yes	no ?	Not enough data.
Mor	yes	yes	The second order is only allowed if the possessor itself is not possessed.
Waropen	yes	no	
Warembori	yes	no	

In most languages the only possible order is possessor–possessed. However, given that at least two languages allow for the possessor to follow the possessed items (Biak and Mor) and the fact that both Proto Austronesian and Proto Oceanic had the 'regular' order (possessed–possessor), I propose that in Proto SHWNG both orders were possible, with the preposed possessor the dominant order, but the other one still in active use, possibly with some subtle pragmatic difference. It is quite possible that the alternative word order has been an overlooked feature in some grammatical descriptions.

3 Conclusion and further questions

The following statements sum up the conclusions reached so far concerning possession in Proto SHWNG.

- Proto SHWNG had a possessive system which contrasted inalienable and alienable possession.
- Inalienable possession was marked by direct affixation on inalienable nouns using a set of possessive suffixes. Inalienable nouns included most body parts, a limited set of kinship terms and some locational nouns.
- Alienable possession was marked by indirect possession, i.e. affixation of the possessive suffixes on two possessive classifiers: **na* and (tentatively) **ri*. The edible classifier **na* was used for food items and some tools, while the general classifier **ri* was used for all other nouns.
- The order within the noun phrase was possessor-possessed, with a minor variant possessed-possessor.

The following tendencies seem to capture most of the developments that individual languages have undergone:

- A tendency towards prefixation, carried to its extreme in Waropen and Warembori.
- A tendency to collapse the distinction between the two alienable classifiers.
- A tendency to move body parts and especially kinship terms into the class of alienable nouns.

I end this paper with some thoughts on the implications for higher subgroups. One very important question is this: at which point in the Austronesian family tree was the alienable-inalienable distinction introduced as an innovation? If Proto SHWNG had this distinction and Proto Oceanic did as well, then we can safely reconstruct it for Proto Eastern-Malayo-Polynesian (PEMP), the next higher node in the Austronesian family tree (see Figure 2).

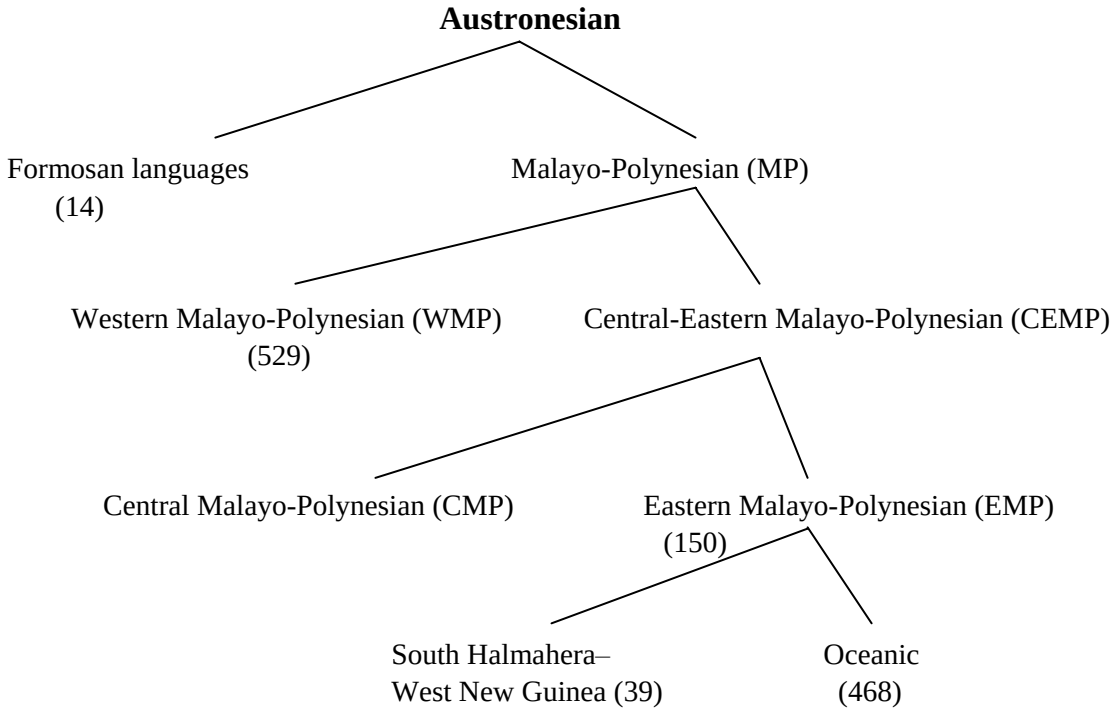


Figure 2: The Austronesian family tree with language numbers (from Tryon 1995)

This is fairly uncontroversial, although at this point it appears difficult to match the reconstructed classifiers for alienable possession in Proto SHWNG and POC. Compare Table 24.

Table 24: Possession in Proto SHWNG and POC

	Proto SHWNG	Proto Oceanic
inalienable	possessive suffix	possessive suffix
alienable general	classifier **ri*	classifier **na*
alienable edible	classifier **na*	classifier **ka*

While Proto SHWNG had **na* for edible possession, POC had **na* for general possession. Lichtenberk's question (1985:130) as to which EMP subgroup innovated is therefore still waiting for an answer.

The next question is whether we can go even higher up in the Austronesian tree. Can the alienable-inalienable distinction be assumed to have been present in Proto Central-Eastern Malayo-Polynesian (Proto CEMP)? Both Lichtenberk (1985:130) and Blust (1993) answer this question affirmatively. Blust (1993) even lists the alienable—inalienable distinction as one of two morphological innovations for Proto CEMP.[17] Both scholars base

[17] Adelaar (2005:25) has challenged the innovative status of the alienable–inalienable distinction in CEMP by pointing at a similar distinction in Puyuma (Formosan) and various Bidayuh (Land Dayak) languages in West Borneo (WMP), as well as at the lack of morphosyntactic uniformity in CEMP languages. However, the presence of the alienable–inalienable distinction in isolated locations in Formosa and Borneo appears to be simply another case of parallel development within the Austronesian family. The lack of morphosyntactic unity within SHWNG is undeniable, but the present article is an attempt to address this variety by positing a simple protosystem.

themselves on the fact that the alienable-inalienable distinction is present in several Central Malayo-Polynesian (CMP) languages, including Kaitetu (on Seram; Collins 1983). Laidig (1993) presents evidence from 18 CMP languages where the distinction is present, including languages on Ambon, Seram, Kei, Aru, Tanimbar and the southwestern islands such as Kisar and Wetar. It is also found in some languages of East Timor, including Keimak and Waimaha (Hull 2001:123-125). This spread over the CMP group is quite impressive and would seem to favour a reconstruction at the level of Proto CMP—if that was indeed a valid protolanguage (see Ross 1995). However, it should be borne in mind that the alienable-inalienable distinction is lacking in CMP languages further west, including all the languages of Bima, Sumba, Sumbawa, Flores as well as most of Timor, as shown in Klamer et al. (2007). The fact that the distinction is predominantly found in the eastern half of the CMP area, bordering on the SHWNG group and the Papuan languages, makes me very cautious in assigning it to Proto CMP and hence to Proto CEMP. A diffusion scenario from Papuan and SHWNG languages would offer a more adequate explanation of the current distribution. Another observation which points in this direction is that the distinction between general and edible alienables as reconstructable for Proto EMP appears to be lacking in all of the CMP languages. The only exception to this is Selaru (Laidig 1993), which, however, shows some unexpected morphology (*hina* is the classifier for edible possession and *wasi* for general possession). If I am correct in reconstructing two alienable strategies for Proto EMP, and if this was also true for Proto CEMP, the almost universal absence of this distinction in the CMP languages is truly puzzling.

I would therefore venture the double hypothesis that 1. the distinction between alienable and inalienable possession entered the Austronesian language family through Papuan contact at the time when Proto EMP was spoken somewhere in the Halmahera - West New Guinea area; and that 2. the distinction between alienable and inalienable possession in CMP languages is best explained as due to diffusion. On the second point I am in full agreement with what Klamer et al. (2007) posit, although they seem to argue that diffusion also occurred in the SHWNG subgroup. Rather than put forward a gradual diffusion of this feature throughout Eastern Indonesian (including SHWNG), I would posit a single contact-induced innovation at the level of Proto EMP, which would explain the presence of the distinction in virtually all the 500+ EMP languages. Ross (2001:138) says that '[I]t is also probable that the formal distinction between alienable and inalienable possession entered Proto Oceanic or an immediate precursor through Papuan contact.' I believe there is ample evidence that this 'immediate precursor' was Proto EMP.

References

Adelaar, K.A. 2005. The Austronesian languages of Asia and Madagascar: a historical perspective. In K.A. Adelaar and N.P. Himmelmann, eds *The Austronesian languages of Asia and Madagascar*, 1–42. London: Routledge/Curzon.

Adriani, N. and A.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 – Zuid-Halmahera*. Batavia: Landsdrukkerij.

Anceaux, J.C. 1961. *The linguistic situation in the islands of Yapen, Kurudu, Nau and Miosnum, New Guinea.* 's-Gravenhage: Martinus Nijhoff.

Blust, R.A. 1978. Eastern Malayo-Polynesian: a subgrouping argument, 181–234. Canberra: Pacific Linguistics.

—— 1993. Central and Central-Eastern Malayo-Polynesian. *Oceanic Linguistics* 32:241–293.

—— 2005. Squib: a note on the history of genitive marking in Austronesian languages. *Oceanic Linguistics* 44:215–222.

Bowden, J. 2001. *Taba: description of a South Halmahera language*. Canberra: Pacific Linguistics.

Collins, J. 1982. A short vocabulary of East Makian. In C.L. Voorhoeve, ed. *The Makian languages and their neighbours*, 99–128. Canberra: Pacific Linguistics.

—— 1983. *The historical relationships of the languages of Central Maluku, Indonesia*. Canberra: Pacific Linguistics.

Cowan, H.J.K. 1955. Notes on Windesi grammar. *Oceania* 26:42–48.

Donohue, M. 1999. *Warembori*. Languages of the World/Materials 341. Muenchen: Lincolm.

Esser, S.J. 1938. Talen. In *Atlas van tropisch Nederland*. Amsterdam: Koninklijk Nederlandsch Aardrijkskundig Genootschap.

Gordon, R.G. Jr., ed. 2005. *Ethnologue. Languages of the world. 15th edition*. Dallas: SIL.

Held, G.J. 1942. *Grammatica van het Waropensch*. Bandoeng: Nix.

Henning, T.A. n.d. The noun phrase in Wandamen. Unpublished typescript.

Hull, G. 2001. A morphological overview of the Timoric Sprachbund. *Studies in languages and cultures of East Timor* 4:98–205.

Klamer, M., G. Reesink, M. van Staden. 2007. East Nusantara as a linguistic area. In P. Muysken, ed. *From linguistic areas to areal linguistics*, 95–150. Amsterdam: John Benjamins.

Laidig, W.D. 1993. Insights from Larike possessive constructions. *Oceanic Linguistics* 32:311–352.

Laycock, D. 1978. A little Mor. In S.A. Wurm and L. Carrington, eds *Second International Conference on Austronesian Linguistics: Proceedings, Fascicle I*, 285–316. Canberra: Pacific Linguistics.

Lichtenberk, F. 1985. Possessive constructions in Oceanic languages and in Proto-Oceanic. In A. Pawley and L. Carrington, eds *Austronesian linguistics at the 15th Pacific Science Congress*, 93–140. Canberra: Pacific Linguistics.

Lynch, J., M. Ross and T. Crowley, eds. 2002. *The Oceanic languages*. Richmond: Curzon Press.

Maan, G. 1951. *Proeve van een Bulische spraakkunst*. ‘s-Gravenhage: Martinus Nijhoff.

Matsumura, T. and M. Matsumura. 1991. A preliminary grammar sketch of the Irarutu language. In M. Hartzler and L. Wood, eds *Workpapers in Indonesian languages and cultures*, vol.10, *Irian Jaya*, 75–110. Jayapura: Cenderawasih and SIL.

Ramar O., N. Ongkodharma and R. Flaming. 1983. *Vekavavo nana kavo Wondama. Percakapan-percakapan dalam bahasa Wandamen. Wandamen conversations*. N.p.: UNCEN-SIL.

Reesink, G.P. 1998. The Bird's Head as Sprachbund. In J. Miedema, C. Odé and R.A.C. Dam, eds *Perspectives on the Bird's Head of Irian Jaya, Indonesia*, 603–642. Amsterdam: Rodopi.

—— 1999. *A grammar of Hatam – Bird's Head peninsula, Irian Jaya*. Canberra: Pacific Linguistics.

—— 2005. West Papuan languages: roots and development. In A. Pawley, R. Attenborough, J. Golson and R. Hide, eds *Papuan Pasts: cultural, linguistic and biological histories of Papuan-speaking peoples.* Canberra: Pacific Linguistics.

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

—— to appear. Nouns and verbs in Magey Matbat. Available at http://www.ling.ed.ac.uk/~bert/

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

—— 1995. Some current issues in Austronesian linguistics. In Darrell T. Tryon, ed. *Comparative Austronesian dictionary* 1, 45–120. Berlin: Mouton de Gruyter.

—— 2001. Contact-induced change in Oceanic languages in north-west Melanesia. In A.Y. Aikhenvald and R.M.W. Dixon, eds *Areal diffusion and genetic inheritance*, 134–166. Oxford: University Press.

—— 2002. The history and transitivity of western Austronesian voice and voice marking. In F. Wouk and M. Ross, eds *The history and typology of western Austronesian voice systems*, 17–62. Canberra: Pacific Linguistics.

—— 2006. Reconstructing the case-marking and personal pronoun systems of Proto Austronesian. In. H.Y. Chang, L.M. Huang, and D. Ho, eds *Streams converging into an ocean. Festschrift in honor of Professor Paul Jen-Kuei Li on his 70th birthday*, 521–563. Taipei: Academia Sinica.

Silzer, P.J. 1983. Ambai: an Austronesian language of Irian Jaya, Indonesia. Unpublished PhD dissertation. The Australian National University, Canberra.

Tryon, D.T., ed., 1995. *Comparative Austronesian dictionary*. Part 1. Berlin: Mouton de Gruyter.

van den Berg, R. 1989. *A grammar of the Muna language*. Dordrecht: Foris.

van den Berg, R. and P. Bachet. 2006. *Vitu grammar sketch*. Data Papers on Papua New Guinea languages vol. 51. Ukarumpa: SIL. Available at http://www.sil.org/pacific/png/abstract.asp?id=47977.

van den Berg, R. and T. Matsumura. 2008. Possession in Irarutu. *Oceanic Linguistics* 47:213–222.

van den Heuvel, W. 2006. *Biak. Description of an Austronesian language of Papua*. Utrecht: LOT.

Voorhoeve, C.L. 1989. Notes on Irarutu. *Irian* 17:107–119.

Whisler, J. 1996. A grammar of Sawai. Unpublished typescript.

21 *Reassessing the reconstruction of plural affixes in PAn: evidence from the Formosan languages*

ELIZABETH ZEITOUN

1 Introduction[1]

Over the past decade, our understanding of plural formation and plural marking in Formosan languages has steadily improved as more synchronic studies have become available. In the 1970s and the 1980s, plural formation did not draw the attention of many linguists in the field. It was discussed as a specific morpho-syntactic device in only two grammars (Li 1973, Pecoraro 1979).[2] Li (1973:107) argued that in Tanan Rukai, 'the features "dual", "plural" and "collective" have little or no syntactic consequence aside from the semantic content.' Pecoraro (1979:61) mentioned that 'in discourse, the plural form is usually omitted and is used when something needs to be stressed upon …' (my

[1] Sections of this paper were presented or discussed in several occasions, most notably at (i) the *Tenth International Conference on Chinese linguistics*, June 22–24, 2001 at Irvine, (ii) the *International Symposium on Austronesian cultures: issues relating to Taiwan*, Dec. 8–11, 2001 at Academia Sinica, (iii) the *Tuesday seminar of the UH Linguistics Department* (University of Hawai'i at Mānoa), February 8, 2007 and (iv) the International Conference *The past meets the present: a dialogue between historical linguistics and theoretical linguistics*. Taipei, Academia Sinica, July 14–16, 2008. I am grateful to all the participants for their and comments, in particular Robert Blust, Lillian M. Huang, Lawrence Reid and Joy J. Wu. I would also like to thank Laurent Sagart, Stacy Teng and Raleigh Ferrell for their helpful suggestions on earlier versions of this manuscript as well as Hua Jia-jing and Chu Tai-hwa for discussions on Southern Paiwan and Tungho Saisiyat respectively. I am also indebted to two reviewers for their pertinent comments. This study was financially supported by a grant from the National Science Council NSC 92-2411-H-001-078 MD.

[2] As of June 2006 when a draft of this paper was revised for publication purposes, twenty grammars had been consulted; thirteen written in Chinese (Series on Formosan languages): Huang (2000a,b,c), A. Chang (2000), H-h. Chang (2000), H. Chang (2000a-b), Lin (2000), Wu (2000), Yeh (2000), Zeitoun (2000a,b,c) on nearly all the extant Formosan languages except Saaroa and Kanakanavu and seven written in English (see Huang 1995a on Atayal, Holmer 1996 and Pecoraro 1979 on Seediq, Tung et al. 1964 on Tsou, Wang 2004 on Thao, A. Chang 2006 on Paiwan, Li 1973 on Tanan Rukai). Two more grammars are now available that treat this issue much more extensively, Teng (2008) on Nanwang Puyuma and Zeitoun (2007) on Mantauran Rukai.

Alexander Adelaar and Andrew Pawley, eds
Austronesian historical linguistics and culture history: a festschrift for Robert Blust, 359–372.
Pacific Linguistics, 2009.
© This edition vested with Pacific Linguistics.

translation, EZ). In the late 1990s and early 2000s, studies on (i) the typology of personal pronouns in Formosan languages (Huang et al. 1999; Zeitoun 2001a), (ii) the notion of quantification in Tsou (H. Chang 2002), (iii) plurality in Mantauran (Rukai) (Zeitoun 2001a and Zeitoun 2007) and (iv) numeral classifiers in Northern Paiwan (Tang 2004) demonstrated that plural marking is overtly expressed on pronouns, nouns and/or verbs in many Formosan languages. Huang et al. (1999) and Zeitoun (2001b) claimed that plurality is overtly expressed at the morphological level in personal pronouns in many Formosan languages. In Zeitoun (2001b), I showed that in Mantauran Rukai, plural marking is expressed through morphological means (affixation and/or reduplication); syntactically, it is both subject-sensitive and object-sensitive, i.e. it is triggered by the quantification of a subject and yielded by the occurrence of a non-volitional/human/oblique participant; semantically, the notion of plurality is closely associated with that of 'humanness', i.e. (usually) only nouns referring to humans are overtly marked as plural. H. Chang (2002) investigated the syntax and semantics of distributive quantifiers in Tsou in relation to the notions of collectivity and plurality. Tang (2004) examined the morphological, syntactic and semantic behavior of numerals primarily in Northern Paiwan in correlation with plural marking.

More recently, the issue on whether plural marker(s) should be reconstructed in Proto Austronesian has been addressed again.[3] Blust (2005)[4] proposes the reconstruction of three genitive case forms in Proto Austronesian: *nu, *ni and *na and argues that '*nu marked the genitive of common nouns, while *ni and *na marked the genitive of singular and plural personal nouns respectively' (p.215). Two papers were written in response to Blust's (2005) hypothesis, one by Ross (2006) and the other by Reid (2007). Ross (2006:527, 530–531) questions Blust's reconstruction of *na as a plural form for personal nouns and posits the reconstruction of a plural marker *a based on the evidence of the nominative case marker *si-a (reflected in Amis *ca* [tsa] and Paiwan *ti-a*) and the genitive case marker *ni-a (> Paiwan *ni-a*), concluding that 'the derivation of *na* from *nia is straightforward: *nia > *ña > *na*.' (Ross 2006:513). He shows that this reconstructible plural marker *a also occurs in third-person pronominal forms in Pazeh and Saisiyat. He mentions, in passing, that Saisiyat and Proto Atayal reflect a plural marker *-la- in third-person pronouns but dismisses this form as 'unconnected with other data' (Ross 2006:537). Reid (2007) reassesses the reconstruction of the three genitive case markers *nu, *ni and *na by examining data from Central Philippine and East-Formosan languages and concludes that there is no strong evidence in support of such a reconstruction because the current forms appear to result from parallel drift and convergence of form and function. Because of the non-cognacity of the data at hand, he proposes different analyses to account for the development of the plural genitive case marker *na in the Philippine and in the East-Formosan languages. He shows that in Philippine languages the most plausible analysis is to assume that the third person plural pronouns *si=dá and *ni=dá were attached to the personal case markers *si (unmarked) and *ni (genitive), yielding respectively **si-dá + si NP and **ni=dá + ni NP. In the daughter languages, these forms later merged, (i) either giving rise to *sidá and *nidá or (ii) undergoing subsequent sound changes, with *d becoming *l* in *sila* and *nila* (*d > l) or, as in *da*, with loss of the unstressed initial syllable. He argues that in the East Formosan languages the formation of certain demonstratives

---

3 See Blust (2005) for a detailed literature review of this topic.

4 This paper is dedicated to Robert Blust as a token of my gratitude for his kindness and guidance to me as a newcomer in the field in the early 1990s and for his friendship.

developed through the cliticisation of the ligature *a* onto demonstrative forms, e.g. Sakizaya Amis *ina* 'this' < Proto Amis *ini=a. By the same process, *na* developed through (i) the cliticisation of *ni* to genitive plural pronouns beginning with the vowel *a* and (ii) the deletion of the vowel *i*, e.g., Amis namu '2P' < Proto Amis *ni=amu. By analogy, *na was reanalyzed as a marker for genitive plural nouns (Reid 2006:245).

The present paper re-examines plural formation and plural marking from a synchronic and a diachronic perspective. Synchronically, it discusses the plural marking of pronouns and nouns in nine Formosan languages (Plngawan Atayal, Central Amis, Isbukun Bunun, Southern Paiwan, Mantauran Rukai, Saaroa, Tungho Saisiyat, Thao, Tsou) and shows that these languages exhibit two plural affixes, viz. *la-*[5] and *a-/-a*. Diachronically, it suggests that these two plural affixes should be reconstructed in PAn as *Na- and *a-/-a, based on the fact that these languages, though all spoken in Taiwan, belong to various primary subgroups (see Blust 1999).

2 Morphological marking of plurality on pronouns

In a majority of Formosan languages, with the exception of Kavalan, Seediq, Tsou, Puyuma and Kanakanavu, plural marking is expressed through the affixation of *a-/-a* or *la-* to the base of third person pronouns.

2.1 Affixation of *a-/-a* to the pronominal base

The affixation of *a-/-a* to the pronominal base form to mark plural is found in four languages: Central Amis (1), Southern Paiwan (2), Thao (3) and Pazeh (4).

(1) Central Amis (Huang 1995b; Huang et al. 1999; Wu 2000)

a.	*ts-**a**-ŋra*	'they (NOM)'	vs	*tsiŋra*	's/he (NOM)'
b.	*ts-**a**-ŋraan*	'they (LOC)'	vs	*tsiŋraan*	's/he (LOC)'
c.	*n-**a**-ŋra*	'their (GEN)'	vs	*ni(ŋ)ra*	'his/hers (GEN)'

(2) Southern Paiwan

a.	*ti-**a**-maɲu*	'they (NOM)'	vs	*timaɲu*	's/he (NOM)'
b.	*ni-**a**-maɲu*	'their (GEN)'	vs	*nimaɲu*	'his/her (GEN)'

(3) Thao (Huang 2000b:82)

a.	*θ-**a**-yθuy*	'they (NEUTRAL)'	vs	*θiθuʔ*	's/he (NEUTRAL)'
b.	*θ-**a**-yθuy*	'their (GEN)'	vs	*θiθuʔ*	'his/her (GEN)'
c.	*θ-**a**-yθun*	'them (ACC)'	vs	*θiθun*	'him/her (ACC)'

(4) Pazeh (Lin 2000:109)

a.	*-(**a**)-misiw*	'they (NOM)'	vs	*-misiw*	's/he (NOM)'
b.	*-y-**a**-misiw*	'they (NOM/OBL)'	vs	*-imisiw*	's/he (NOM/OBL)'
c.	*n-**a**-misiw*	'them (GEN)'	vs	*n-imisiw*	'his/her (GEN)'

[5] This prefix has different realisation in the Formosan languages and I will refer to *la-* for convenience.

In Central Amis, Southern Paiwan and Thao, the pronominal forms consist of a (non-common/personal noun) case marker plus a pronominal base. In Southern Paiwan and in Thao, the plural marker *a-/-a* is added to the base; in Amis, it replaces the vowel /i/. In Pazeh, *a-/-a* is either attached to the base or it replaces the vowel *i*; compare (4a-b) with (4c).

2.2 Prefixation of *la-* to the pronominal base

The occurrence of *la-* is found in at least four languages: Tungho Saisiyat (5), Atayal (6), Saaroa (7), and Mantauran Rukai (8).

(5) Tungho Saisiyat

a.	***la***-*θia*	'they (NOM)'	vs	*θia*	's/he (NOM)'
b.	*hi*-***la***-*θia*	'them (ACC)'	vs	*hi*-*θia*	'him/her (ACC)'
c.	*kan*-***la***-*θia*	'them (LOC)'	vs	*kan*-*θia*	'him/her (LOC)'
d.	*ni*-***la***-*θia*	'their (GEN)'	vs	*ni*-*θia*	'his/her (GEN)'
e.	*ʔan*-***la***-*θia-a*	'theirs (POSS)'	vs	*ʔan*-*θia-a*	'his/hers (POSS)'
	ʔn-***la***-*θia-a*	'theirs (POSS)'	vs	*ʔn*-*θia-a*	'his/hers (POSS)'
f.	*ʔni*-***la***-*θia*	'them (BEN/DAT)'	vs	*ʔni*-*θia*	'him/her (BEN/DAT)'
g.	*ki*-***la***-*θia*	'with them (COM)'	vs	*ki*-*θia*	'with him/her (COM)'
	ki-***l***-*θia*	'with them (COM)'			

(6) Plngawan Atayal

a.	***la***han	'them (OBL)'	vs	*hiyan*	'him/her (OBL)'
b.	***la***ha*ʔ*	'they (NEUTRAL)'	vs	*hiya*ʔ	's/he (NEUTRAL)'

(7) Saaroa (Li 1997:285)

a.	*i-la-**la**-isa*	'they'	vs	*-i-la-isa*	's/he'
b.	*-**l**-isa*	'they'	vs	*-isa*	's/he'

(8) Mantauran Rukai

a.	*i*-***l***-*inə*	'them (OBL)'	vs	*-inə*	'him/her (OBL)'
b.	*i*-***l***-*iðə*	'them (OBL)'	vs	*-iðə*	'him/her (OBL)'
c.	***l***-*i*-*ni*	'their (GEN)'	vs	*-ni*	'his/her (GEN)'
d.	***l***-*i*-*ða*	'their (GEN)'	vs	*-ða*	'his/her (GEN)'

In Tungho Saisiyat, all the pronominal forms are inflected for case (*Ø* for Nominative, *hi* for Accusative, *kan* for Locative etc. …). Among these, *ʔan-la-θia-a ~ ʔn-la-θia-a* 'theirs (POSS)' and *ki-la-θia ~ ki-l-θia* 'with them (COM)' occur in free variation. In both Mantauran Rukai and Plgnawan Atayal, deletion is observed: in Plgnawan Atayal *iy* is deleted in the plural form, and in Mantauran Rukai, *la-* actually surfaces as *l-* (the circumfix *i-* …*-ə* represents the marking of the oblique case).

Isbukun Bunun pronominal forms are very neat and thus may have undergone restructuring. In any case, this language exhibits an opposition between the singular forms, marked by *s-* and the plural forms, marked by *n*.

(9) Isbukun Bunun

a.	***n***-*aiaʔ*	'they (NOM)'	vs	***s***-*aiaʔ*	's/he (NOM)'
b.	***n***-*aitiaʔ*	'they (ACC)'	vs	***s***-*aitiaʔ*	's/he (ACC)'
c.	*ʔi*-***n***-*aitiaʔ*	'they (GEN)'	vs	*ʔi*-***s***-*aitiaʔ*	's/he (GEN)'

2.3 Partial conclusions

The occurrence of the two affixes *a-/-a* and *la-* in so many languages cannot be treated as a mere coincidence or as a result of diffusion or borrowing as suggested in Li (1997) because they exhibit both cognacity and paradigmaticity. Ross (2006) has demonstrated that the contrast between Paiwan *ti-a* (NOM.pl) and *ti-* (NOM.sg), *ni-a* (GEN.pl) and *ni-* (GEN.sg) supports the reconstruction of the plural *a in PAn. Such a reconstruction is further evidenced by the pronominal paradigms of Central Amis, Thao and Pazeh. In Tungho Saisiyat, Mantauran Rukai, Plgnawan Atayal and Saaroa, *l*/*ɬ* are reflexes of PAn *N (see Ross 1992 and Li 1997) and the occurrence of *la-/l-/ɬa-* points toward the reconstruction of the PAn form *Na-. Though Bunun has merged PAn *N and *n into *n*, the data fits the paradigm given for the other Formosan languages. I agree that in general 'third-person personal pronouns in Formosan languages either are derived from or still are demonstrative pronouns' (Ross 2007:536) and believe that such forms were marked as plural at the PAn level.

Plural marking on pronouns in the nine Formosan languages discussed in this paper is summarised in Table 1, along with the reconstructed PAn forms.

Table 1: Morphological formation of Plural pronouns in nine Formosan languages

Plural marking on:	Paiwan	Amis	Thao	Pazeh	Saisiyat	Rukai	Atayal	Saaroa	Bunun
personal pronouns	-a	-a	-a	a-	la-/l-	l-	la-	ɬa-	n-
PAN reconstruction	*a-/-a				*Na-				

The affixes *a-/-a* and *la-* also occur in noun phrases (either on the noun or on the case marker) in at least six languages,[6] viz. Mantauran Rukai, Southern Paiwan, Central Amis, Nanwang Puyuma, Tungho Saisiyat and Tsou.

3 Overt plural marking on nouns and/or case markers

Two generalisations are worth mentioning at the outset: first, plural marking can be expressed either on the noun or on the case marker but never simultaneously on the noun and the case marker that precedes it; second, plural marking usually occurs on nouns with a human reference.

In Central Amis, plural marking is overtly indicated on case markers. In Mantauran Rukai and Tsou, it is specifically marked on nouns. In Tungho Saisiyat and Southern Paiwan, it occurs either on case markers (*-l* in Tungho Saisiyat and *-a* in Southern Paiwan) or on nouns (*la-* in Tungho Saisiyat and *ʎa-* in Southern Paiwan).

Tables 2 and 3 depict the distribution of the affixes *a-/-a* and *la-* in the Formosan languages.

---

[6] Further investigation may reveal even more complex systems of plural marking in the Formosan languages.

Table 2: Distribution of *a-/-a* in Formosan languages

Occurrence of /*a*/ in:		Pazeh	Thao	Amis	Paiwan	Rukai	Tsou
Pronouns		+	+	+	+	–	–
Case markers		–	–	+	+	–	–
Nouns	Proper and kinship	–	–	–	–	–	–
	Common	–	–	–	–	+	(+)

Table 3: Distribution of *la-* in Formosan languages

Occurrence of *la-* in:		Saisiyat	Atayal	Paiwan	Rukai	Saaroa
Pronouns		+	+	–	+	+
Case markers		+	–	–	–	?
Noun	Proper and kinship	+	+	+ (voc. only)	+	?
	Common		–		–	?

3.1 Occurrence of *a-/-a*

3.1.1 Occurrence of *-a* on case markers

In Central Amis and Southern Paiwan, the suffix *-a* occurs on the case markers preceding proper nouns and kinship terms and indicate associative plurality. As mentioned above, in Central Amis, the suffix *-a* replaces the vowel /i/, while it attaches to the base in Southern Paiwan. The case marking system of these two languages is outlined in (10) and (12) with examples illustrating the singular/plural contrast given (11) and (13).

(10) Southern Paiwan nominal case marking system

		Nominative	Genitive	Oblique
Common nouns		*a*	*n(u)a*	*t(u)a*
Proper nouns	sg.	*ti*	*ni*	*cay*
and kinship terms	pl.	*ti-a*	*ni-a*	*cai-a*

(11) Southern Paiwan

a. *na-k<əm>ə‿ɬəm* *ti kapi* *cay* *kivi.*
PRF-beat<AF>beat NOM Kapi OBL Kivi
'Kapi beat Kalalu.'

b. *na-k<əm>ə‿ɬəm* *tia kapi* *cay* *kivi.*
PRF-beat<AF>beat NOM:pl Kapi OBL Kivi
'Kapi (and his companions) beat Kivi.'

(12) Central Amis nominal case marking system (Wu 2000:64)

		Nominative	Genitive	Accusative	Neutral
Common nouns		*ku*	*nu*	*tu*	*u*
Proper nouns	sg.	*ci*	*ni*	*ci...-an*	*ci*
and kinship terms	pl.	*ca*	*na*	*ca…-an*	*ca*

(13) Amis

a. *mapaloʔ ni ina ci mayaw.*
PF:beat GEN mother NOM Mayaw
'Mayaw was beaten by mother.'

b. *mapaloʔ ni ina ca mayaw.*
PF:beat GEN mother NOM:pl Mayaw
'Mayaw (and his companions) were beaten by mother.'

3.1.2 Occurrence of *a-* on nouns

In Tsou and Mantauran Rukai, *a-* occurs on common nouns referring to a human participant to mark the plural. While this process seems to be fossilised in Tsou, it is very productive in Mantauran Rukai.

Examples of plural formation through the affixation of *a-* in Tsou is illustrated in (14); *a-* is usually inserted immediately before the base:[7]

(14) Plural formation in Tsou

a.	*ha-**a**-hocŋɨ*	'men'	vs	*hahocŋɨ*	'man'
b.	*ma-**a**-mespiŋi*	'girls, women'	vs	*mamespiŋi*	'girl, woman'
c.	*ma-**a**-meoi*	'old men'	vs	*mameoi*	'old man'

In Mantauran Rukai, nearly all nouns referring to a human participant can appear in a plural form. Different classes of nouns must be distinguished, however, which undergo different plural word formation processes. Nouns referring to a human entity, whether basic (e.g., 'child', 'woman', 'man') or derived through nominalisation (e.g., 'married ones') are marked by *a-* for plural, as illustrated in the paradigm given in (15) and (16).

(15) Mantauran Rukai (Zeitoun 2007:118–119)

a.	***a**-savasavarə*	'young men'	vs	*savarə*	'young man'
b.	***a**-tamatama*	'middle-aged men'	vs	*tamatama*	'middle-aged man'
c.	***a**-tomotomo*	'old (wo)men'	vs	*tomotomo*	'old (wo)man'
d.	***a**-vaɭovaɭo*	'young women'	vs	*vaɭovaɭo*	'young woman'

(16) Mantauran Rukai (Zeitoun 2007:119)

a.	*ta-**a**-ʔacakəlaə*	'who are married'	vs	*ta-ʔacakəlaə*	'who is married'
b.	*ta-ka-**a**-roðaŋə*	'who are old'	vs	*ta-ka-roðaŋə*	'who is old'

The contrast between (15a) and (15b-d) on the one hand and (16) on the other shows that the prefixation of *a-* yields the reduplication of the base form of underived nouns, e.g. *a-sava-savarə* 'young men' vs *savarə* 'young man' unless it is attached to a lexicalised reduplicated root as in *vaɭovaɭo* 'young woman'; in that case, the base form does not undergo further reduplication, e.g. *a-vaɭovaɭo* 'young women'. Such a restriction does apply to nouns derived through nominalisation, see (16).

---

[7] In the following examples, I tentatively treat the first syllable as an output of reduplication.

3.2 Occurrence of *la-*

Throughout the Formosan languages, *la-* is more productive with nouns (see §3.2.1) than with case markers (§3.2.2).

3.2.1 Occurrence of *la-* on nouns

To date, the occurrence of *la-* is evidenced by data from Mantauran Rukai, Southern Paiwan, Plngawan Atayal and Tungho Saisiyat. These languages vary in terms of the types of nouns that can be marked for plural.

In Mantauran Rukai, only vocative and non-vocative kinship nouns as well as family names can be marked for plural through the prefixation of *la-*, as illustrated in (17)–(18).

(17) Mantauran Rukai (Zeitoun 2007:120)

a.	*l-inakə*	'Mother and aunts!'	vs	*inakə*	'Mother/aunt!'
a'.	*l-iina*	'mother and aunts'	vs	*ina-*	'mother'
b.	*l-amakə*	'Father and uncles!'	vs	*amakə*	'Father/uncle!'
b'.	*l-aama*	'father and uncles'	vs	*ama-*	'father'

(18) Mantauran Rukai (Zeitoun 2007:121)

a. *ðonaʔi* *la-paðoma* *oðaacə* *ʔaəla* *moa*
that pl-family name DYN.FIN:leave DYN.SUBJ:move DYN.SUBJ:go

soŋao ...
Bunun
'As for the Lapadhoma, they left and moved to the Bunun tribe …'

b. *ʔəɬəŋə-ɬao* *Ø-paðoma.*
Elenge-1S.NOM Ø-family name
'I am Elenge Padhoma.'

In Southern Paiwan, Plngawan Atayal and Tungho Saisiyat, the use of *la-* was generalised to all nouns with a human reference.

In Southern Paiwan, common nouns referring to a human participant can be marked as plural either through the reduplication of the base and/or the prefixation of *ʎa-* (also a reflex of *Na) as shown in (19).

(19) a. *vavayavayan* '(a group) girls/women' vs *vavayan* 'girl/woman'
ʎa-vavayavayan 'girls/women'

b. *ʔuqaʎaqaʎay* '(a group of) boys/men' vs *ʔuqaʎay* 'boy/man'
ʎa-ʔuqaʎaqaʎay 'boys/men'

Reduplication yields a collective/plural meaning ('a group of'),[8] whereas prefixation of *ʎa-* yields a vocative/plural meaning; compare (20)–(21).

(20) a. *vavayan* *tiamaɲu*
girl/woman 3P.NOM
'We are girls.' (two at most)

[8] As mentioned in Tang (2004), the reduplication of *kakəɖian* 'child' yields *kakəɖikakəɖian* '(very) young', and not '(a group of) children' as expected.

b. *vavayavayan* *tiamaɲu*
RED:girl/woman 3P.NOM
'We are girls.' (above three)

c. **ʎa-vavayavayan* *tiamaɲu*
pl-RED:girl/woman 3P.NOM

(21) *təkəʃ-u,* *ʎa-vavayavayan.*
drink-IMP pl-RED:girl/woman
'Girls/women, drink!'

Kinship terms used vocatively are marked by *ʎa-*, as shown in (22)–(23):

(22) Southern Paiwan

a.	*ʎa-'ama*	'Father and uncles!'	vs	*'ama*	'father (±voc)'
b.	*ʎa-'ina*	'Mother and aunts!'	vs	*'ina*	'mother (±voc)'
c.	*ʎa-vuvu*	'Grandparents/grandchildren'	vs	*vuvu*	'grandparent/grandchild (±voc)'

(23) a. *idu,* *vuvu!*
come:IMP grandparent/grandchild
'Come, grandparent/grandchild!'

b. *idu,* *ʎa-vuvu!*
come:IMP pl-grandparent/grandchild
'Come, grandparents/grandchildren!'

In other words, the neutralisation in Southern Paiwan takes place among vocative nouns.

In Plngawan Atayal, the prefix *la-* can attach to proper nouns (24a), kinship terms (24b) and common nouns with a human reference (24c).

(24) a.	*yumin*	'Yumin'	~	*la-yumin*	'Yumin (and friends)'
b.	*la-yaki?*	'grandmothers/grandmother (and friends)'	~	*yaki?*	'grandmother'
c.	*kinsat*	'policeman'	~	*la-kinsat*	'policemen'

Associative plurality in Tungho Saisiyat is marked on nouns with a human reference through the prefixation of *la-*. Compare (25) and (26).

(25) Tungho Saisiyat

a. *βaʃi?* *ki* *?ataw* *ka-kotih* *noka* *hima?.*
Bashi' COM 'ataw RED:REC-pinch GEN hand
'Bashi' and Ataw pinch each other's hands.'

b. *tatini?* *ki* *korkoriŋ* *ka-kotih* *noka* *hima?.*
old man COM child RED:REC-pinch GEN hand
'The old man and the child pinch each other's hands.'

(26) a. *βaʃi?* *ki* *la-?ataw* *ka-kotih* *noka* *hima?.*
Bashi' COM pl-'ataw RED:REC-pinch GEN hand
Bashi' and 'ataw (and his friends/companions/relatives …) pinch one another's hands.'

b. *tatini?* *ki* *la-korkoriŋ* *ka-kotih* *noka* *hima?.*
old man COM pl-child RED:REC-pinch GEN hand
'The old man and the children pinch each other's hands.'

3.2.2 Occurrence of *-l* on case markers

The occurrence of the plural *-l* on case markers is found in only one language, Tungho Saisiyat: instead of appearing on the noun, the plural marking is overtly expressed on the case marker as in (27a). The ungrammaticality of (27b) shows that *la-* and *-l* cannot occur simultaneously on the noun and on the case marker. Besides, case markers preceding common nouns with a human reference cannot be overtly marked for plurality, as shown in (27c).

(27) a. *βaʃi?* *ki-l* *ʔataw* *ka-kotih* *noka* *hima?.*
Bashi' COM-pl 'ataw RED:REC-pinch GEN hand
Bashi' and 'ataw (and his friends/companions/relatives …) pinch one another's hands.'

b. **βaʃi?* *ki-l* *la-ʔataw* *ka-kotih* *noka* *hima?.*
Bashi' COM-pl pl-'ataw RED:REC-pinch GEN hand

c. **tatini?* *ki-l* *korkoriŋ* *ka-kotih* *noka* *hima?.*
old man COM-pl child RED:REC-pinch GEN hand

3.3 Summary

The distinction between *a-*/*-a* and *la-* (the function of these affixes overlaps somehow in the six languages compared above) can be captured as follows:

(i) In Central Amis and Southern Paiwan, the plural suffix *-a* occurs on case markers (preceding proper nouns and kinship terms) and marks (associative) plurality.

(ii) In Southern Paiwan, the prefix *ʎa-* attaches to vocative nouns with a human reference.

(iii) In Plngawan Atayal and Tungho Saisiyat, there is neutralisation between plural common and proper/kinship nouns: both types of nouns can be prefixed by *la-* to form the plural. In Tungho Saisiyat, the prefix *la-* can further be attracted to the case marker preceding proper/kinship nouns. It is then realised as *-l*.

(iv) In Mantauran Rukai, there is a distinction between plural common nouns marked by *a-* (as in Tsou), and plural non-common nouns marked by *la-*.

Table 4 provides a tabular summary of these generalisations.

Table 4: Morphological marking of plurality on nouns in six Formosan languages

Plural marking on:	Saisiyat	Atayal	Paiwan	Rukai	Amis	Tsou
Noun	*la-* (1/2)	*l-* (1/2)	*ʎa-*(2/3)	*a-* (2) *la-*(3)	–	*(a-)*(2)
case marker	*-l* (1)	–	*a-*(1)	–	*-a* (1)	–

Notes: 1. forms between parentheses indicate unproductive or fossilised marking.
2. (1) indicates associative plural, (2) plural meaning in human common nouns, and (3) plural meaning in proper nouns, kinship terms and/or family names.

4 Conclusion

Ross (2006) and Reid (2007) do not reconstruct any third-person plural pronouns for PAn: since 'Proto Austronesian probably did not have any third person personal pronouns, […] their presence in the daughter languages in Formosa and the Philippines is the result of independent innovations in a number of subgroups.' (Reid 2007:245) On the other hand, they both agree with Blust (2005) that at either a higher (i.e., PAn) or at a lower level (i.e., PCP), case markers could be marked as plural. Their respective analyses differ as to what should be reconstructed at one level or the other.

The present paper provides an alternative analysis, giving support to Ross's (2006) hypotheses. It shows that a close inspection of different Formosan languages leads to the reconstruction of two plural affixes *Na- and *a-/-a which occurred in nouns with a human reference (including proper and kinship nouns as well as common nouns) and third person pronouns (usually derived from demonstratives). It would be interesting to go further and try to find out whether PAN *Na- and *a-/-a also occurs in other Western Austronesian languages (among others, the Philippine languages), given that the only reconstructed PMP plural form so far is *da (Reid 2007). This issue goes, however, beyond the scope of this paper.

References

Blust, Robert. 1977. The Proto-Austronesian pronouns and Austronesian subgrouping: a preliminary report. *Working papers in Linguistics* 9.2:1–15.

—— 1999. Subgrouping, circularity and extinction: some issues in Austronesian comparative linguistics. In Zeitoun, Elizabeth and Paul Jen-kuei Li, eds *Selected Papers from the Eighth International Conference on Austronesian Linguistics*, 31–94. Symposium Series of the Institute of Linguistics (Preparatory Office), Academia Sinica, No.1. Taipei: Institute of Linguistics (Preparatory Office), Academia Sinica.

—— 2003. *Thao dictionary*. Language and Linguistics Monograph Series, No.A5. Taipei: Institute of Linguistics (Preparatory Office), Academia Sinica.

—— 2005. A note on the history of genitive marking in Austronesian languages. *Oceanic Linguistics* 44:215–222.

Chang, Hsiou-chuan. 2000. *A reference grammar of Paiwan*. Series on Formosan Languages, 9. Taipei: Yuanliou Pub. Co. (In Chinese)

Chang, A. Hsiou-chuan. 2006. A reference grammar of Paiwan. PhD dissertation, Canberra: The Australian National University.

Chang, Hsun-hui. 2000. *A reference grammar of Yami*, Series on Formosan Languages, 13. Taipei: Yuanliou Pub. Co. (In Chinese)

Chang, Henry Yung-li. 2000a. *A reference grammar of Seediq*, Series on Formosan Languages, 6. Taipei: Yuanliou Pub. Co. (In Chinese)

Chang, Henry Yung-li. 2000b. *A reference grammar of Kavalan*, 12. Series on Formosan Languages, 6. Taipei: Yuanliou Pub. Co. (In Chinese)

Chang, Henry Yung-li. 2002. Distributivity, plurality, and reduplication in Tsou. *Tsing Hua Journal of Chinese Studies* 32.2:1–30.

Ho, Dah-an. 1997. Kanakanavu. In Li, Paul Jen-kuei, ed. The Formosan languages of Kaohsiung county, 228–271. Kaohsiung: Kaohsiung county government. [In Chinese]

Holmer, A. 1996. *A Parametric grammar of Seediq*. Travaux de l'Institut de linguistique de Lund, 30. Lund: Lund University Press.

Huang, Lillian M.. 1995a. *A study of Mayrinax grammar*. Taipei: The Crane.

—— 1995b. The case markers and pronominal system in Amis. *Journal of National Chengchi University* 70:217–258.

—— 2000a. *A reference grammar of Atayal*. Series on Formosan Languages, 1. Taipei: Yuanliou Pub. Co. (In Chinese)

—— 2000b. *A reference grammar of Thao*. Series on Formosan Languages, 4. Taipei: Yuanliou Pub. Co. (In Chinese)

—— 2000c. *A reference grammar of Puyuma*. Series on Formosan Languages, 10. Taipei: Yuanliou Pub. Co. (In Chinese)

Huang, Lillian M., Elizabeth Zeitoun, Marie M. Yeh, Joy J. Wu and Anna H. Chang. 1999. A typological study of pronouns in the Formosan languages. In Wang, Samuel H., Tsao Feng-fu and Lien Chin-fa, eds *Selected papers from the 5th International Conference on Chinese Linguistics*, 165–198. Taipei: The Crane.

Li, Paul Jen-Kuei. 1973. *Rukai Structure*. Institute of History and Philology, Special Publications No.64. Taipei: Institute of History and Philology, Academia Sinica.

—— 1977. The internal relationships of Rukai. *Bulletin of the Institute of History and Philology*, 48.1:1–92.

—— 1996. The pronominal systems in Rukai. In the *Festschrift in honour of Professor Isidore Dyen*, 209–230. Hamburg: Abera Publishing House.

—— 1997. Saaroa. In Li, Paul Jen-kuei, ed. *The Formosan languages of Kaohsiung county*, 272–297. Kaohsiung: Kaohsiung county government. [In Chinese]

Li, Paul Jen-Kuei and Shigeru Tsuchida. 2001. *Pazih dictionary*. Taipei: Academia Sinica. Language and Linguistics Monograph Series, No.A2. Taipei: Institute of Linguistics (Preparatory Office), Academia Sinica.

Lin, Ying-chin. 2000. *Pazeh*, Series on Formosan Languages, 3. Taipei: Yuanliou Pub. Co. (In Chinese)

Pecoraro, F. 1979. *Eléments de grammaire taroko*. Précédés de la présentation de la vie et de la culture des taroko. Paris: Association Archipel.

Reid, Lawrence. 2007. Another look at the marking of plural personal noun constructions in Austronesian languages. *Oceanic Linguistics* 46:232–252.

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

—— 1995. Reconstructing Proto-Austronesian verbal morphology: evidence from Taiwan. *Austronesian studies relating to Taiwan*. In Paul Jen-Kuei Li, Cheng-hwa Tsang, Ying-kuei Huang, Dah-an Ho, and Chiu-yu Tseng, eds 727–791. Taipei: Institute of History and Philology, Academia Sinica.

—— 2006. Reconstructing the case-marking and personal pronoun systems of Proto Austronesian. In Chang, Yung-li, Lillian M. Huang and Dah-an Ho, eds *Streams converging into an ocean*, 521–563. Language and Linguistics Monograph Series No.W-5. Taipei: Academia Sinica.

Shelley, George. 1979. Wudai Dekai, the language, the context and its relationships. PhD Hatford University.

Starosta, Stanley. 1994. Proto-Rukai-Tsouic: subgroup or treetop? Paper presented at the Seventh International Conference on Austronesian Linguistics, Leiden, The Netherlands, August 22–27.

—— 1995. A grammatical subgrouping of Formosan languages. In Paul Jen-kuei Li, Cheng-hwa Tsang, Ying-kuei Huang, Dah-an Ho and Chiu-yu Tseng, *Austronesian studies relating to Taiwan*, 683–756. Symposium Series of the Institute of History and Philology, Academia Sinica No.3. Taipei: Academia Sinica.

Tan, Cindy Ro-lan. 1997. A study of Puyuma simple sentences. Taipei: National Taiwan Normal University MA thesis.

Tang, Chih-Chen Jane. 2004. Two types of classifier languages. *Language and Linguistics* 5.2:377–407.

Tang, Chih-Chen Jane, Yung-li Chang and Dah-an Ho. 1998. On noun phrase structures in Paiwan. *Tsing Hua Journal of Chinese Studies, New Series*, 28.3:335–384.

Teng, Stacy F. 2008. *A grammar of Puyuma: an Austronesian language of Taiwan*. Canberra: Pacific Linguistics.

Tung, T'ung-ho, S.H. Wang, T.K. Kuan, T.F. Cheng, and M. Yan. 1964. *A descriptive study of the Tsou language, Formosa*. Institute of History and Philology, Special Publications No.48. Taipei: Institute of History and Philology, Academia Sinica.

Wang, Shan-shan. 2004. An ergative view of Thao syntax. PhD dissertation. Honolulu, University of Hawai'i at Mānoa.

Wu, Jing-lan. 2000. *A reference grammar of Amis*. Series on Formosan Languages, 11. Taipei: Yuanliou Pub. Co. (In Chinese)

Yeh, Mei-li. 2000. *A reference grammar of Saisiyat*. Series on Formosan Languages, 2. Taipei: Yuanliou Pub. Co. (In Chinese)

Zeitoun, Elizabeth. 1995. Problèmes de linguistique dans les langues aborigènes de Taiwan [English version: Issues on Formosan linguistics] PhD dissertation, Paris = Université Denis Diderot (Paris 7).

—— 2000a. *A reference grammar of Bunun*. Series on Formosan Languages, 5. Taipei: Yuanliou Pub. Co. (In Chinese)

—— 2000b. *A reference grammar of Tsou*. Series on Formosan Languages, 7. Taipei: Yuanliou Pub. Co. (In Chinese)

—— 2000c. *A reference grammar of Rukai*. Series on Formosan Languages, 8. Taipei: Yuanliou Pub. Co. (In Chinese)

—— 2001a. Plurality in Mantauran (Rukai). Paper presented at the Tenth International Conference on Chinese Linguistics. Irvine, University of California at Irvine, June, 22–24, 2001.

—— 2001b. Typology of the Formosan languages. Paper read at the *International Symposium on Austronesian cultures: issues relating to Taiwan*. Taipei, Academia Sinica, December 8–11.

—— 2007. *A grammar of Mantauran Rukai.* Language and Linguistics Monograph Series A4–2. Taipei: Academia Sinica.

Zeitoun, Elizabeth and Hui-chuan Lin. 2003. *We should not forget the stories of the Mantauran: memories of the past.* Language and Linguistics Monograph Series A4. Taipei: Academia Sinica.

Part 4

Subgrouping

22 *Austronesian language phylogenies: myths and misconceptions about Bayesian computational methods*

SIMON J. GREENHILL and RUSSELL D. GRAY [1]

Historical linguistics has never been particularly intimate with computers. The first wave of computational historical linguistics—lexicostatistics—was developed in the 1950s (Swadesh 1952; Lees 1953) and quickly applied to language groups around the world from Indo-European to Austronesian (Lees 1953; Hymes 1960; Embleton 1986). However, critics were quick to point out the problems caused by assuming a single constant rate of lexical replacement and repeatedly noted the erroneous results that this produced (Hoijer 1956; Bergsland and Vogt 1962; Blust 1981; McMahon and McMahon 2006). As a consequence of these critiques lexicostatistics has been widely rejected by mainstream historical linguists (Campbell 2004).

The last few years have seen a second wave of computational approaches entering historical linguistics: phylogenetic methods. These techniques, drawn from evolutionary biology, have been used to investigate some provocative and controversial claims about human prehistory. For example, we have applied phylogenetic methods to lexical data compiled by Bob Blust to test hypotheses about the settlement of the Pacific (Gray and Jordan 2000; Greenhill and Gray 2005; Gray et al. 2009). Our results reflected a settlement pattern through Island South-East Asia, New Guinea and then into Oceania, consistent with the 'Out of Taiwan' scenario (e.g. Blust 1999; Pawley 2002; Diamond and Bellwood 2003). We have also used these methods to investigate the origins of the Indo-European (Gray and Atkinson 2003) and Bantu languages (Holden 2002; Holden and Gray 2006). Other groups have applied phylogenetic methods to investigate the internal subgrouping of these families (Ringe et al. 1998; Rexovà et al. 2003, 2006). The application of

[1] We would like to thank Andreea Calude, Andy Pawley, and Malcolm Ross for comments on this paper. We would like to note that some of the analyses reported in this paper have been superseded by those conducted using a better fitting model of cognate evolution reported in Gray et al. (2009).

Alexander Adelaar and Andrew Pawley, eds.
Austronesian historical linguistics and culture history: a festschrift for Robert Blust, 375–397.
Pacific Linguistics, 2009.
© This edition vested with Pacific Linguistics.

computational phylogenetic methods has not been restricted to just lexical data. Phylogenetic analyses of structural features have revealed historical signals in Papuan languages that may stretch back around 10,000 years (Dunn et al. 2005). Nor have phylogenetic methods been restricted to just building trees. Phylogenetic network methods have been used to investigate conflicting signals in Indo-European (Bryant et al. 2005), Bantu (Holden and Gray 2002), Chinese dialects (Hamed and Wang 2004), and Polynesian (Bryant 2006; Gray 2007). Finally, phylogenetic methods have recently been used to investigate general claims about the factors that affect the rate of language change. Pagel et al. (2007) used phylogenetic methods to estimate the rates of lexical replacement in Indo-European languages and showed an almost hundred-fold difference between the rates of rapidly evolving words (e.g. 'dirty') and the slowly evolving words (e.g. 'tongue'). They then calculated the frequency at which these words were currently used in four large language corpora. Their results showed a strong correlation between the frequency with which words are used today and their stability over time: the more a word is used, the slower it evolves. This striking result suggests that over the 9000 years of Indo-European language history, there have been consistent underlying mechanisms controlling lexical replacement. A second study (Atkinson et al. 2008) used phylogenetic methods to test claims that speakers often use their language as a social tool for increasing group cohesion and demarcating groups (Labov 1994). The results showed a strong relationship between the total amount of lexical change and the number of language splitting events along the tree: between 10% to 33% of the total lexical change in the Bantu, Indo-European, and Austronesian languages occurred as a rapid burst of change shortly after languages diverged. This punctuational change (e.g. Bowern 2006) is consistent with rapid language change in small founder populations and differentiation as a cultural marker.

Given the combination of strong claims, new techniques, and the high-profile reporting of results, it is not surprising that these studies are often controversial. Responses have ranged from the positive: 'Computational methodologies of this kind can only be helpful for historical linguistics' (April McMahon in Balter 2003:1491), to the skeptical: 'There is no reason whatsoever to assume that vocabulary would behave the same way that organisms do.' (Alexander Lehrman in Balter 2004:1326), to the negative: '… have ignored the fatal shortcomings of glottochronology …' (Eska and Ringe 2004:569), and the painfully incorrect; 'sledg(ing) the dead horse of the Swadesh algorithm' (Holm 2007:201).

Sadly many of these criticisms are mired in misunderstanding. Computational phylogenetic methods are not just lexicostatistics redux, but a powerful supplement to the comparative method used in historical linguistics. On several occasions Bob Blust has challenged us to specify exactly how phylogenetic methods differ from lexicostatistics and explain why they are superior. Here we respond to his challenge. To do this, we will focus on one of the great battlegrounds between lexicostatistics and the traditional comparative method: the Austronesian language family. First, we will describe how Bayesian phylogenetic methods work, and then give a step-by-step explanation of an analysis of a large lexical dataset for 400 Austronesian languages (Gray et al. 2009; Greenhill et al. 2008).

1 The Austronesian language family

The Austronesian language family is one of the two largest in the world, containing around 1000 to 1200 languages (Gordon 2005). Before Columbus, these Austronesian languages were also the most widely dispersed with speakers in Mainland and Island South-East Asia, Madagascar, Micronesia, Melanesia, and Polynesia (Bellwood et al.

1995). The groundwork that identified this family began in the 16th and 17th centuries as European scholars began to compare word lists that trickled back from early explorers and missionaries (e.g. Houtman 1603; Reland 1708; Forster 1778; Brandes 1884; Kern 1886). Dempwolff (1934, 1938) systematically reconstructed early Austronesian phonology and lexicon, and identified a large subgroup, Oceanic, to which he assigned the languages of Melanesia, Polynesia and (most of) Micronesia (Dempwolff 1937). The evidence that all these Oceanic languages formed a subgroup of Austronesian implied that they stem from a single Austronesian settlement of this region from the west (Grace 1961, 1964a; Pawley and Green 1973; Pawley and Ross 1995).

A major challenge to this hypothesis came from Dyen's lexicostatistical analyses of vocabulary from 352 Austronesian languages (1962, 1965). Lexicostatistics had previously been applied to subgroups within Austronesian (an early paper by Elbert (1953) explored Polynesia), but Dyen's was by far the largest in scale. At the time, Dyen's analysis was an impressive computational feat; his program compared 7,000,000 pairs of words. The lexicostatistical results suggested a tree with 40 first-order branches, no fewer than 30 of which were located in Melanesia. Dyen took this to indicate that the most probable area of origin of the Austronesian languages was in Melanesia, possibly in the Bismarck Archipelago north of New Guinea, with subsequent expansions east into Polynesia, and west into Indonesia then to the Philippines and Taiwan. This study was hailed by Murdock (1964:117) as '… a significant work—one which may conceivably be as revolutionary for Oceanic linguistics and culture history as was the work of Greenberg (1949–54) for the interpretation of African languages and cultures'.

This enthusiasm was short-lived. Grace (1964b, 1966) was quick to to suggest that the difference between the lexicostatistical view of Austronesian relationships and that of the traditional view may be a consequence of faster rates of lexical replacement in Melanesia. Blust (1981, 2000) quantitatively demonstrated that the Austronesian languages varied markedly in their retention rates across a 200-item basic vocabulary word-list. Retention rates in Malayo-Polynesian languages ranged from 5% to 60% in the interval between Proto Malayo-Polynesian and the present, a time period of around 4000 years. Moreover, Blust (2000) argued that the inability of lexicostatistics to discriminate between shared retentions and innovations—a distinction that had been critical in historical linguistics since Brugmann (1884)—exacerbated the effect of different rates. These differences in retention rates, especially in regions such as Melanesia where there have been high levels of language contact and borrowing (Ross 1996) rendered the lexicostatistical conclusions invalid.

In contrast to a Melanesian origin for Austronesian languages suggested by lexicostatistics, the comparative method has provided strong evidence that all languages outside Taiwan belong to a single sub-group (Dahl 1973; Blust 1977), which Blust (1977) named Malayo-Polynesian. In a series of publications Blust (e.g. 1977, 1978, 1982, 1999) marshalled a large array of evidence for the claim that the Proto Austronesian (PAn) homeland lay in Formosa (Taiwan). First, Blust (1999) concluded there are at least nine primary subgroups of Austronesian within Taiwan, whereas all Austronesian languages spoken outside of Taiwan fall into a single first order subgroup. There are a number of phonological and morphological innovations that are shared by the Malayo-Polynesian subgroup but are not found in the Formosan languages. If we assume that the region with the most primary subgroups is likely to be the primary dispersal centre Taiwan is thus strongly favoured as the Austronesian homeland. Blust (1982) also used the distribution of flora and fauna lexicon to delimit the range of possible Austronesian homelands. The

distribution of cognate words for placental and marsupial mammals in Austronesian languages suggests that ancestral Austronesian society was located in the Asiatic faunal zone to the west of the Wallace line. Archaeological evidence indicates that the spread of Neolithic cultures from Taiwan parallels the directions and dates of the Austronesian linguistic expansion. This conjunction of different lines of evidence has convinced most specialists in Austronesian historical linguistics that the Austronesian-speaking people were present in Taiwan around 5500 years ago, before spreading into the Philippines, Indonesia and through the Pacific (e.g. Shutler and Marck 1975; Bellwood 1997; Blust 1995; Kirch 2000; Kirch and Green 2001; Pawley 2002).

The failure of lexicostatistics to get Austronesian 'right' is not surprising—computing Austronesian language relationships is a very difficult problem. First, the rapid expansion of the Austronesian family means that it is likely to be difficult to resolve the fine branching structure of the Austronesian language tree as there is little time for the internal branches on the tree to develop numerous shared innovations (Pawley 1999). Second, as these languages moved across the Pacific they encountered new environments and the consequent need for new terminology may have increased the rates of language replacement. This acceleration in rates is likely to be exacerbated by the effects of language contact—particularly within Near Oceania (Ross 1996). Additionally, many Austronesian languages have small speech communities, which are also likely to speed up the rates of language evolution (Nettle 1999). The effects of these factors can be seen in the substantial variation in cognate retention rates in Austronesian languages (Blust 1981, 2000; Pawley this volume). Finally, the sheer scale of the Austronesian language family is daunting—with around 1000 to 1200 languages there are more than 10^{2864} possible rooted family trees. In the following section we will outline a Bayesian phylogenetic analysis on the Austronesian languages.

2 A phylogenetic approach

Much of biology and linguistics is historical. That is, to understand these systems properly we need to know their history. Where did particular languages or species come from? When did they arise and diverge? What sequence of changes took place? Are two characteristics similar because they share common ancestry or are they similar because they've evolved to fill the same function? To investigate these questions biologists have developed a large collection of tools collectively known as phylogenetics. Biologists initially constructed phylogenetic trees with clustering algorithms such as UPGMA ('Unweighted Pair-Group Method using Arithmetic averages', Sneath and Sokal 1963), that analysed pairwise similarity matrices (just like the lexicostatistical percentage shared cognacy matrices). Not surprisingly, this approach also produced inaccurate results when there were substantial differences in the rates of genetic change between lineages (Felsenstein 1978). However, rather than abandon a computational approach when confronted with this difficulty, biologists improved the computational methods. In the last few decades phylogenetic methods have revolutionised biology and have become the dominant way of testing historical evolutionary hypotheses (Huelsenbeck and Rannala 1997; Pagel 1999). Currently, the Bayesian phylogenetic approach is seen as the most powerful and robust approach available (Lewis 2001; Huelsenbeck et al. 2001, 2002). In the section below we will outline the major components of Bayesian phylogenetic analysis: dataset construction, maximum likelihood modeling, and the search for the most probable evolutionary trees.

2.1 Data

For successful phylogenetic analysis we need a large amount of well-sampled data with sufficient historical information to resolve the aspects of the phylogeny we are interested in. The comparative method commonly used in historical linguistics takes a sample of lexicon and proceeds to reconstruct systematic sound correspondences between the languages in order to uncover historically related 'cognate' forms (Durie and Ross 1996). This information about cognate sets can easily be coded as binary characters. An example of this is shown in Table 1. The data, in this case the words meaning 'bone' in a number of Austronesian languages (Column A) are divided into cognate sets on the basis of systematic sound correspondences (Column B). Once the cognate sets have been determined and any known loan words removed, then the data can be coded into a binary matrix showing the presence or absence of each cognate set for every language (Column C). In the 400-language dataset used in this paper, the cognate sets in a 210-item word-list produced 34,440 binary characters.

It is worth emphasising that whilst most recent work computing language phylogenies has primarily been based on cognate datasets (e.g. Atkinson et al. 2008; Gray and Atkinson 2003; Gray and Jordan 2000; Greenhill and Gray 2005; Gray et al. 2009; Holden 2002, Holden and Gray 2006; Pagel et al. 2007; Rexovà et al. 2003), other linguistic characters could also be used as long as there is sufficient data and an appropriate way of modeling the changes in these characters. Indeed, some studies have used combinations of lexical and grammatical data (Rexovà et al. 2006) and typological information (Dunn et al. 2005).

Table 1: Cognate data coding from original lexical data (A), to cognate set information (B), to binary characters (C)

Language	(A) Item	(B) Cognacy	(C) Binary Coding
Paiwan	tsuqela	1	1000
Itbayaten	tuqgan	1	1000
Bare'e	wuku	2	0100
Mangarrai	toko	2	0100
Numfor	kor	3	0010
Motu	turia	4	0001
Fijian (Bau)	sui-na	4	0001
Tongan	hui	4	0001
Samoan	ivi	4	0001
Maori	iwi	4	0001

2.2 Maximum likelihood models

The next step is to analyze the data. Bayesian phylogenetic inference builds on an older tradition of Maximum Likelihood methods (Fisher 1922; Edwards 1964; Felsenstein 1981; Pagel 1999). In this framework the data is treated as a fixed and given observation, and the analysis aims to find the values of model parameters that explain this data well (Pagel 1999; Steel and Penny 2000). To do this we need a stochastic model of language evolution that specifies how the changes between the character states should be counted. In modeling

language evolution in this way we make simplifying assumptions about relevant processes and explicitly build these into the model. For example, a very simple model of lexical evolution would require one parameter—the rate of change between the absence of a specific cognate and the presence of that cognate. In this simplest model, this rate would be symmetrical in the sense that the rate at which any cognate was gained would be equal to the rate at which a cognate was lost. Obviously, this is not very realistic. Once a cognate set has arisen it is much more likely to be lost than for another language to independently derive it. A more realistic model would accommodate the differential ease of losing a cognate over gaining it by adding a second parameter, so there is now one rate for cognate gain and one rate for cognate loss (we will refer to this as the two-parameter mode below).

What other important parameters could be added? One of the major problems with lexicostatistics is that it assumed a constant rate of cognate loss of around 19% every thousand years in the 200-item Swadesh list (Lees 1953). This fixed rate did not allow for differences in rates of change between cognate sets, or for differences in rates of change between languages. Both of these types of rate variation are common in Austronesian languages (Blust 1981, 2000). Site-specific rate heterogeneity (different sites in DNA sequences evolving at different rates) was also a problem for early phylogenetic methods (Posada and Crandall 2001). More recent approaches, however, have solved this by enabling a distribution of rates instead of a single rate. One common method is to estimate a gamma distribution of rate changes from the data (Yang 1994). This method gives each character an *inherent* rate of change so that some cognates are gained or lost rapidly, whilst others are more resistant to change. Modeling lexical change in this way allows for the differences between highly persistent characters like reflexes of 'hand—Proto Austronesian **(qa)-lima* (Blust 1999)—and highly unstable characters such as words meaning 'dirty'.

The full model with two rate parameters and gamma-distributed rate heterogeneity can then be used to calculate a numeric value known as the likelihood. The likelihood measures how well the data are explained by the tree under this model. Our aim is to find the set of trees that explain the data well, or in other words, find those trees with the maximum likelihood. The general approach to finding trees here is to take a tree and then permute it in some fashion (e.g. by changing the tree shape, or the amount of change along a branch, or model parameters, etc.) to give a second tree. The likelihood of both those trees under the given model of language evolution can then be compared to find the better tree. Here the search algorithm (usually a Markov Chain Monte Carlo approach—described below), preferentially selects the tree with the better likelihood, and iterates over this procedure many times, to find a set of good trees.

Critics of early language studies using Bayesian phylogenetic methods claimed that the models were 'inappropriate' as they had been designed for biological analyses rather than linguistic change (Eska and Ringe 2005; Naklekh et al. 2005). This criticism demonstrates a misunderstanding of the rationale behind model-based inference. Whilst it is true that language change is complex, and the model employed here and elsewhere (e.g. Gray and Atkinson 2003; Gray et al. 2009) is simple, this simplicity does not necessarily discredit or invalidate the methodology. Developing a model is a trade-off between over- and under-fitting model parameters (Burnham and Anderson 1998). Typically the fit will improve as parameters are added to the model, especially if the new parameters capture an important

aspect of the process. More complex models are not uncommon in biology; one of the most popular models used for genetic data is the General Time-Reversible model (Yang et al. 1994). This model has six parameters: one for each of the rates of change between each combination of the four bases found in DNA. This is often coupled with gamma distributed rate heterogeneity, and an allowance for invariant sites, giving a total of eight parameters.

However, as parameters are added the sampling error also increases and therefore it becomes difficult to reliably estimate the model parameters (Swofford et al. 1996). Therefore, the goal of modeling language evolution is not to build a complex model that captures every aspect of language change, but rather to construct the simplest model that provides reliable estimates of the parameters with finite amounts of data. Choosing the most appropriate model is not an issue for armchair speculation. We can evaluate the performance of the model by analysing the data with a range of models, and then selecting the best model with a standard model comparison test such as the Likelihood Ratio Test (Goldman 1993), or Bayes Factor Comparison (Suchard et al. 2001).

2.3 An example of a likelihood calculation

To clarify the way in which likelihood scores are calculated we have outlined a simple example in Figure 1 (adapted from Swofford et al. 1996, and Atkinson and Gray 2006). This figure shows the basic procedure on a set of data coded in a binary matrix as described above (1A). We will follow the process of likelihood calculation for one of these characters: character ‘a’. Character ‘a’ represents a cognate set found in the Oceanic languages Motu, Fijian, Samoan and Maori, and absent from the other languages in our example dataset. To show how the likelihood can measure how well a topology describes the data we will compare two different trees (1B). The tree on the left represents the accepted linguistic history of these languages, whilst the tree on the right does not. First, character ‘a’ is mapped onto both the trees, and all the possible ancestral states of this character are enumerated. The likelihood of this distribution of character state change on the tree is then calculated using the chosen model of cognate evolution that specifies the probabilities of transitions between cognate presence and absence (1C). The likelihood of the distribution of character ‘a’ on the tree is the product of all possible ancestral state reconstructions for this character (1D). Finally, the overall likelihood of each tree can be calculated by repeating this process for all the characters in the data, giving rise to a single score for each tree. Note that, in contrast to lexicostatistics, actual character state changes are inferred on the tree. This means that the distinction between retentions and innovations is part of the analysis. The overall likelihood score, generally reported as the log of the likelihood (lnL), represents how well the data is explained by the tree given the model. Better trees are characterized by less negative log likelihoods (1E). In figure 1, the tree on the left has a log likelihood of -4178, whilst the second tree scores -4627. Thus, the former tree is a better explanation of the data.

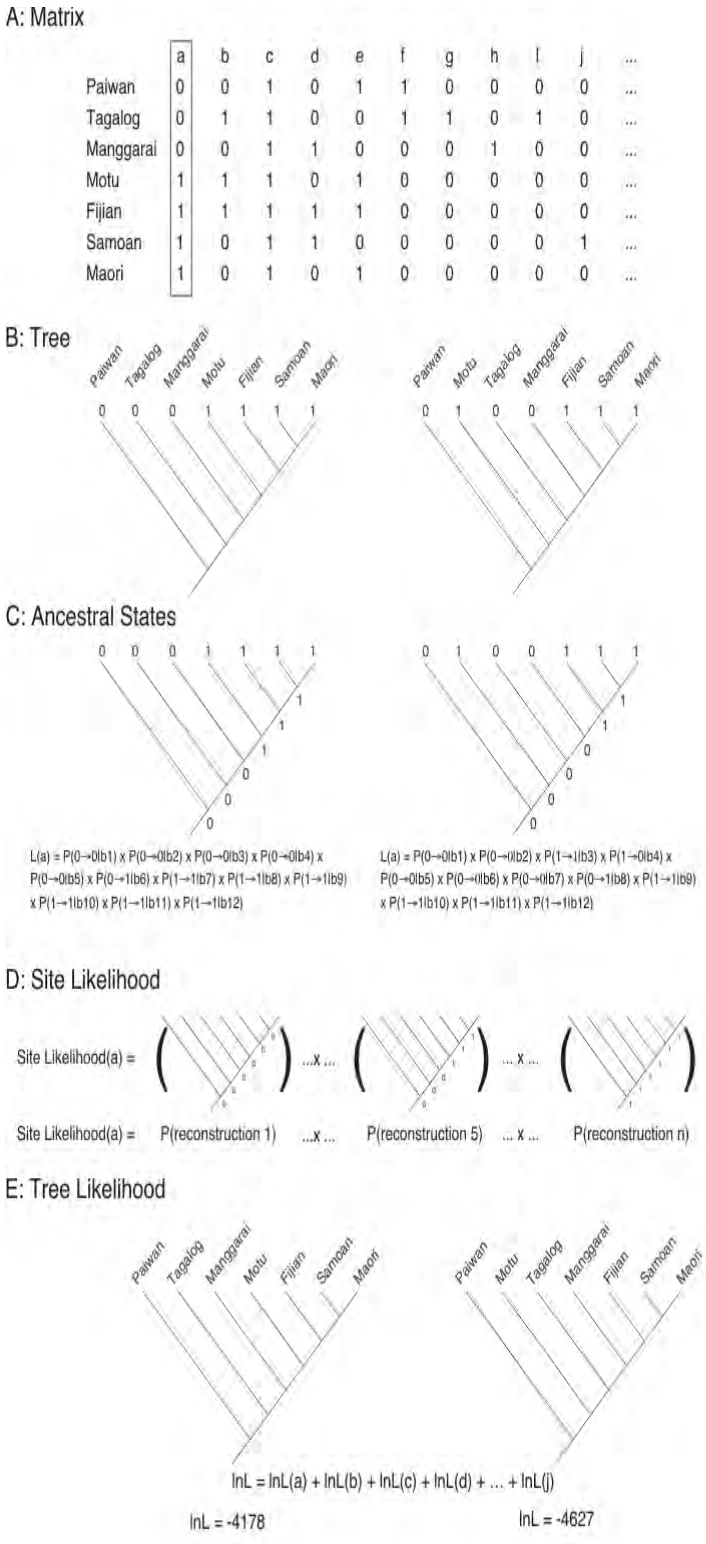


Figure 1: The calculation of the log likelihood of a tree. (A) A hypothetical cognate presence/absence matrix for seven Austronesian languages. (B) Two different trees for these languages with character 'a' mapped onto them. (C) An example of one possible ancestral state reconstruction of character 'a' on these topologies. (D) The site likelihood for character 'a' on the tree is calculated as the product of the probability of all possible ancestral state combinations for that character. (E) The overall tree likelihood is calculated as the sum of

the likelihoods of each site, where the tree with the lower (less negative) log likelihood fits the data better. Here, under a two-parameter model of cognate gain/loss (with no gamma distribution), the tree on the left is a better fit to the data with a log likelihood of -4178, whilst the other tree fits the data less well with a likelihood of -4627.

2.4 Finding the most probable trees

Once we've chosen an appropriate model we then have an explicit *optimality criterion* with which to measure how good a tree is. This means that we can search through the range of possible trees until we find the one(s) with the highest likelihood under this optimality criterion. However, as the number of languages analysed increases so does the number of possible trees. If a tree is strictly bifurcating (i.e. each node can only have two daughter languages), then the number of trees can be calculated as shown in (1) where *n* is the number of languages (Graham and Folds 1972).

(1) $$\frac{(2n-3)!}{2^{n-1}(n-1)!}$$

Thus, when there are four languages there are 15 possible trees. Adding one more language increases the number of possible trees to 105. When the data contains more than 50 languages, there are more possible trees than there are atoms in the universe. If Austronesian has around 1000 languages, then there are an intimidating 3.8×10^{2864} possible combinations. This unfortunately means that it is not possible to search through all the trees in any non-trivial dataset. A systematic technique for finding a subset of the good trees from this huge space of possible trees is therefore required. Moreover, as with any statistical estimate, we need some way of evaluating how robust our inferences are.

To do this we use a Bayesian inferential approach that combines the likelihood with our prior knowledge of the trees to give the *posterior probability distribution* of trees. This can be calculated using Bayes's theorem as in (2) (Huelsenbeck et al. 2001).

(2) $$\mathrm{P[Tree \mid Data]} = \frac{\mathrm{P[Data \mid Tree]} \times \mathrm{P[Tree]}}{\mathrm{P[Data]}}$$

The posterior distribution contains the trees that have high likelihoods and fit the data well, given the data and the priors. Priors are the initial values of the model parameters. Often the prior distribution of the parameters is 'flat'; that is all values are considered equally probable. However, if there is strong external evidence supporting some hypothesis, then this can be taken into account explicitly (Lewis 2001). For example, if one wanted to assume that new languages were born at a constant rate across the tree, then a 'Yule' prior on branching rate could be implemented. The ability to incorporate extra information using priors is very powerful—but must be justified. Calculating the posterior probability distribution is hard as it involves the integration of all model parameters, across all branch length combinations, over every single tree (Huelsenbeck et al. 2001). However, using Markov Chain Monte Carlo methods (MCMC, Metropolis et al. 1953; Huelsenbeck et al. 2001), we can sample from the posterior probability distribution. The phrase 'Monte Carlo' refers to a random sampling method, and a 'Markov Chain' is a process which draws each sample from the probability distribution of the previous state (Larget 2005). To find trees this method starts with a tree (usually randomly generated) and permutes it in some fashion (e.g. changing the topology, branch lengths or model parameters)—this is the Markov

Chain process. The chain preferentially samples trees from this distribution according to their likelihood scores—the Monto Carlo process. If run long enough the chain provides a representative sample of the most probable trees. There are two further considerations in the use of Bayesian MCMC methods. First, the initial trees sampled are heavily contingent on the model's starting parameters (i.e. the priors). To avoid this early samples in an MCMC run are usually discarded as 'burn-in'. Second, each successive tree in an MCMC run is a permutation of the previous one due to the nature of the Markov Chain process (i.e. tree 2 is tree 1 with a branch moved or a change in branch length, etc). This means that each tree is highly correlated with its neighbors. To avoid this auto-correlation, and thus make each sample statistically independent, it is common to only keep every 1,000th or 10,000th tree from the post-burn-in set of trees.

3 Using phylogenetic trees

Using this procedure we will be left with a collection of trees sampled from the posterior probability distribution that should explain the data well. The results we present here are drawn from an analysis using the two-parameter model of cognate gain/loss and gamma-distributed rate variation (Pagel and Meade 2004). This was run for 100,000,000 generations on a cluster of over 150 processors (over 21 years of computer time). The trees were sampled every 10,000 generations after a burn-in of 20,000,000 generations. This gave us a final sample of 8000 trees. However, the endpoint of a phylogenetic analysis is not finding the trees: trees by themselves are boring. Instead, the rationale is to use them to test hypotheses and to investigate the process of evolution. There are many things one can do with trees (Gray et al. 2007). Here we will describe how this set of 8000 most probable trees from the MCMC run can be used to test hypotheses about subgrouping, to date events on the trees, and to trace character change.

3.1 Subgrouping

In historical linguistics it is common to use a family tree to depict the groupings (families, groups, clades, etc) once the groups have been identified using the comparative method. However, there is no formal way of quantifying the support for subgroups. The phylogenetic trees provide a statistical *estimate* of the sub-groupings in the data, and provide a measure of the uncertainty in this estimate. A common way of doing this is to use a Majority Rule 'consensus' tree. This combines the groupings present in all trees in the posterior tree sample. The percentage of trees containing a certain group can be taken as a measure of the support for that grouping in the data. Figure 2 shows an example majority rule consensus tree from our Austronesian data. Subgroups with posterior probability values close to 1.0 are well-supported. For example, the grouping of the Philippine languages is strongly supported by the data (0.99). More surprisingly, the branch grouping the languages of Vanuatu and New Caledonia is also well-supported (0.98). These values mean that 99% and 98% of the 8000 trees in the posterior tree distribution contain those respective groupings. In contrast, other regions of the tree are more poorly supported (e.g. the branch placing the Admiralties languages inside Oceanic after the New North Guinea/Papuan Tip languages has only 0.58 support). Groups with very weak support (<0.50) are not shown. Weakly supported groups could either be the consequence of little signal in the data due to rapid population expansions, or conflicting stronger signals (perhaps produced by borrowing), or non-tree-like descent processes such as dialect chains and linkages.

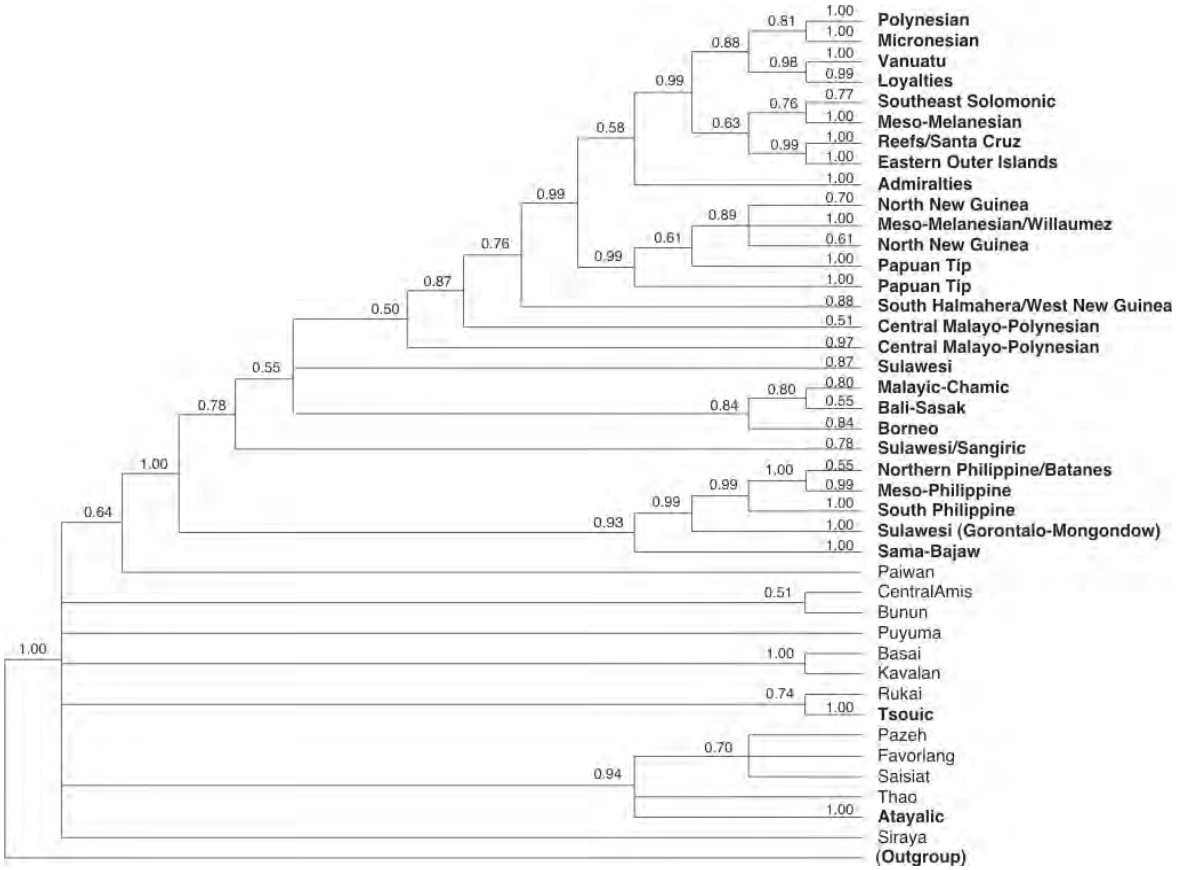


Figure 2: The majority-rule consensus tree of all post burn-in Austronesian trees. Labels in bold represent subgroups of languages, normally-weighted labels denote languages. Where subgroups appear twice in the tree this indicates that they are not monophyletic (e.g. Central Malayo-Polynesian). The numbers on the branches denote the posterior probability of each node. For example, the split between the Northern- and Meso-Philippine languages is strongly supported (1.00). Posterior probability values below 0.50 are considered weak and are not included.

Recall that the lexicostatistical analyses incorrectly 'rooted' the Austronesian languages in East New Guinea. Our phylogenetic analyses, however, support Blust's (1999) rooting of the Austronesian languages in Taiwan. The Formosan languages are placed at the base of the tree after the outgroup languages. There is no unified Formosan subgroup but at least seven higher-order branches of Formosan derived from Proto Austronesian. Moreover, whilst the Tsouic and Atayalic subgroups of Formosan languages are robust, there is little support for other higher-order sub-groupings within Formosan. These results are all concordant with Blust (1999).

Not only do the phylogenetic trees support a Formosan origin of the Austronesian languages, the sequence of the higher-order subgroups closely conforms to the 'Out of Taiwan' scenario of Austronesian settlement. Moving down the tree, after Formosan languages we find the languages of Island South-East Asia, with strong support for the Philippine and Malayic-Chamic language groups. This is followed by two weakly supported groups of Central Malayo-Polynesian languages, and then the well-supported South Halmahera/West New Guinea group. Finally, there is a well-supported Oceanic subgroup, with strong support for the recognized subgroups within Oceanic (Polynesian, Micronesian, Southeast Solomonic, Eastern Outer Islands, Admiralties). Our results split

Oceanic into two major groups, both strongly supported (0.99). The first of these Oceanic subgroups is comprised of the Papuan Tip, North New Guinea and Meso-Melanesian languages. This represents the Western Oceanic group identified by Ross (1988). However, only the Willaumez languages of Meso-Melanesian are in this subgroup, the remainder is located in our second Oceanic grouping. This second Oceanic group contains the Remote Oceanic language subgroups and the majority of the Meso-Melanesian languages. Interestingly, we show strong support (0.99) for the recently identified subgroup Temotu containing the languages from the Eastern Outer Islands and the Reefs–Santa Cruz region (Ross and Næss 2007). In contrast to Blust (1998), the Admiralties subgroup is not at the base of the entire Oceanic subgroup, but is situated—albeit very weakly (0.58)—between Western and Remote Oceanic. Some of the higher-order nodes within our two Oceanic groupings are only weakly supported, such as the cluster grouping Temotu to Southeast Solomonic (0.63). These low values may reflect the rapid dispersal of languages through this region (Pawley 1999), or the large amounts of contact induced change in large-scale dialect networks found in this region (Ross 1996).

3.2 Dating

One of the great attractions of lexicostatistics was its apparent ability to calculate absolute dates of language divergence times through a method known as *glottochronology* (Lees 1953). This technique calculated absolute ages by assuming that as languages split they lost vocabulary at a constant rate. Accordingly, a simple decay curve of cognate loss could be used to calculate divergence times by solving the equation in (3) where C is the percentage of shared cognates between the two languages, r is the retention rate, and t is the estimated time depth.

(3) $$t = \frac{\log C}{2 \log r}$$

Over 1000 years the retention rate r was often assumed to be 81% for the 200 item Swadesh list (Lees 1953). Therefore, if two languages shared 90% of their basic vocabulary, they should have diverged 250 years ago, whilst languages that were 75% similar should have diverged around 680 years ago. However, these glottochronological calculations magnified all the shortcomings of lexicostatistics. Languages vary substantially in their retention rates, and this rate variation produced some obviously inaccurate dates (Bergsland and Vogt 1962; Blust 2000). For example, Icelandic shares over 95% of its core vocabulary with Old Norse. According to glottochronology Old Norse and Icelandic would have diverged less than 200 years ago. This is incorrect—Old Norse was spoken around 1000 years ago (Bergsland and Vogt 1962). Problems such as this led to such a strong rejection of glottochronology that over fifty years later we are still being cautioned about its inaccuracy (McMahon and McMahon 2006).

The age of the Indo-European language family has been a topic of considerable interest and much debate. There are two main theories. The first proposes that Proto Indo-European broke up 5000–6000 years ago when Indo-European languages spread with the expansion of the archaeological culture known as Kurgan (Gimbutas 1973). The main alternative account suggests that Indo-European spread with the advent of farming technology around 8000–9000 years ago (Renfrew 1987). Naturally, one of the first uses we put phylogenetic methods

to was dating the divergence of particular branches of Indo-European (Gray and Atkinson 2003, Atkinson and Gray 2006). Our results showed strong support for an initial breakup of the Indo-European family around 8000–9000 years ago, with a subsequent breakup of 'Nuclear Indo-European' (Indo-European minus Anatolian and Tocharian) around 6000 years ago. The results were robust to different calibrations, cognate coding, and likelihood models (Atkinson et al. 2005). However, we were promptly criticized for merely, 'reintroducing glottochronology by the back door' (Gamble et al. 2005:208), and 'ignor(ing) the fatal shortcomings of glottochronology' (Eska and Ringe 2004:569). These are unfortunate misunderstandings. Phylogenetic dating methods, such as the Penalized Likelihood rate smoothing approach (Sanderson 1997, 2002) used by Gray and Atkinson (2003), as well as newer methods which can 'relax the clock' (Drummond et al. 2006), do not have the fatal shortcomings of glottochronology. These approaches need not assume that there is a single 'clock-like' rate of lexical change (Atkinson and Gray 2006).

To demonstrate how divergence date estimation can be obtained without a strict 'glottoclock' we will estimate the age of Proto Austronesian on the (expanded) tree from Figure 2. The branches on the trees in our posterior sample are proportional to the amount of change along that lineage. This is usually expressed as the rate of substitutions (in this case the gain or loss of cognates in a language). These branch lengths can be converted to time by adding historically attested calibration points. For example, the Eastern Polynesian subgroup can be constrained to around 1200 to 1300 years ago on the basis of initial settlement times (Green and Weisler 2002). Similarly, the Chamic subgroup can also be calibrated based on the fact that Chamic speakers were mentioned in Chinese records around 1800 years ago, and probably entered Vietnam around 2600 years ago (Thurgood 1999). This calibration of nodes on the tree within a historical time range allows the method to estimate how fast the changes measured by the branch lengths are occurring. The Penalized Likelihood rate-smoothing approach can then convert branch lengths into time estimates by smoothing the rates of change across the tree. Instead of assuming a constant retention rate, this allows certain parts of the tree to change faster or slower than others. We applied this approach to one tree from the posterior distribution of trees for our analysis. The resulting dated tree (Figure 3) shows an age of around 5310 years for Proto Austronesian, and an age of 4240 years for Proto Malayo-Polynesian. We must emphasise at this point that the date estimates should be done on all trees in the posterior sample and not just a single one. Calculating divergence dates on all the trees would produce a distribution of the most probable age of Proto Austronesian. This distribution can then be used to provide a confidence interval on any date estimate. As our aim in this paper is to illustrate the overall approach rather than to test specific hypotheses, we have just dated one tree for illustrative purposes. However, dates from this tree support the emergence of Proto Austronesian in Taiwan around 5500 years ago (e.g. Blust 1995; Pawley 2002). Note also the presence of pauses and rapid pulses of expansion as has been argued by Blust (1999), Green (1999) and Pawley (1999, 2002). In this tree, we see a pause of around 1000 years before Proto Malayo-Polynesian arises, and a subsequent rapid pulse of expansion through to Proto Oceanic. Another pause then expansion pulse occurs after the initial settlement of the Central Pacific region.

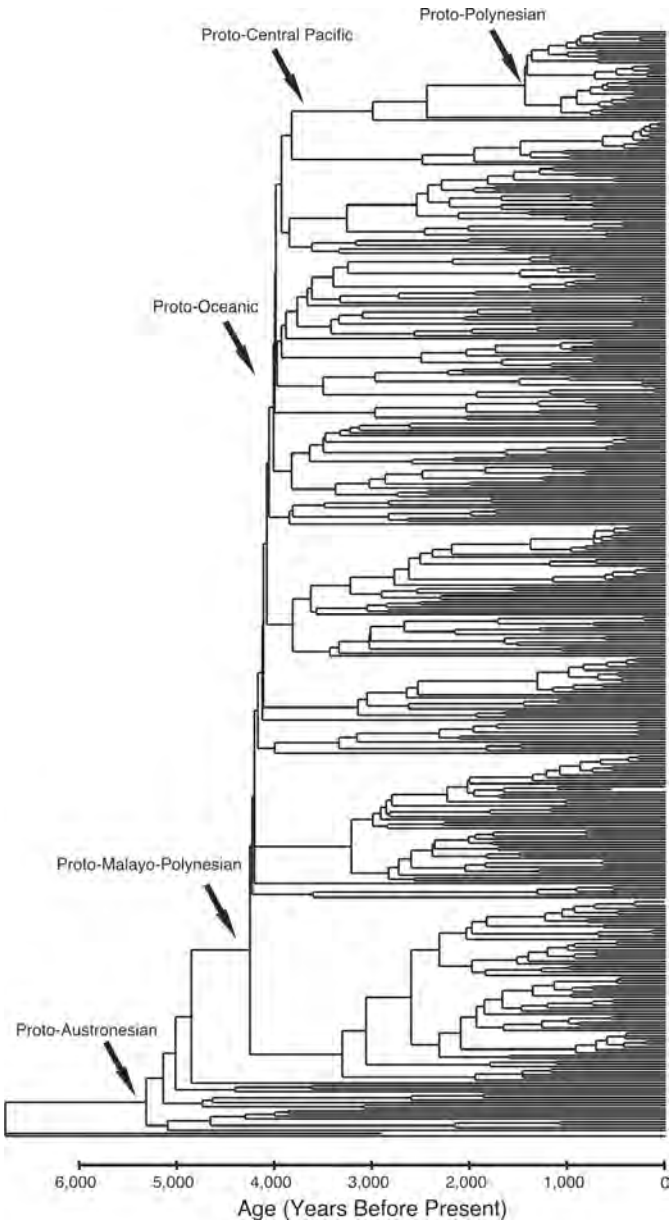


Figure 3: A dated tree of 400 Austronesian languages showing the age of a number of protolanguages estimated using Penalized Likelihood rate smoothing. On this tree Proto Austronesian is estimated to be 5310 years old and Proto Malayo-Polynesian 4240 years.

3.3 Tracing character history

Much of historical linguistics is concerned with the reconstruction of protoforms. This is done both as a means to subgrouping and as a way of making inferences about society and culture of ancestral speech communities. Biologists have also developed phylogenetic methods to reconstruct ancestral states. These methods have been used to tackle problems such as identifying the origin of ancestral genes in the eukaryote genome (Lester et al. 2006). One common phylogenetic approach essentially 'maps' a character of interest onto the posterior tree sample using a continuous-time Markov model of trait evolution (Pagel et al. 2004). Under this model a character can change between a finite number of states over infinitesimally small time periods. The rates of change between these states along the

branches can be estimated directly from the posterior tree sample. These model parameters can then be used to calculate the probability of a certain state at any given node. For example, one might want to evaluate how the words for 'earth, soil' had evolved in the Polynesian languages, and infer what variant was spoken by Proto Polynesian. Figure 4 shows three cognate sets for words meaning 'earth/soil' mapped onto a tree of the Central Pacific subfamily (the expanded form of Figure 2). Cognate set A (colored white) reflects forms like Tongan *kelekele*, Samoan *'ele'ele* and Fijian (Bau) *qele*. Cognate set B (colored gray) reflects forms like the Tahitian *repo* and Hawaiian *lepo*. Cognate set C (colored black) reflects forms like Vaeakau-Taumako's *pela*. Using the Bayesian ancestral state reconstruction method (Pagel et al. 2004) we can estimate that, on this tree, the probability that Proto East Polynesian and Proto Tahitic had cognate set B was 0.99. This is concordant with the comparative method, where the reconstructed Proto East Polynesian form is **repo* (Biggs and Clark 2000). Deeper in the tree, the Proto Polynesian and Proto Central Pacific nodes reflect cognate set A with a probability very close to 1. Again, this matches the reconstructed Proto Central Pacific form **g(w)ele* (Ross et al. 1998). Cognate set C presumably reflects Prot Oceanic **pela* 'muddy' (Biggs and Clark 2000) with semantic change. We emphasise again that ideally this estimation should be integrated over the set of trees in the posterior sample, not just a single tree.

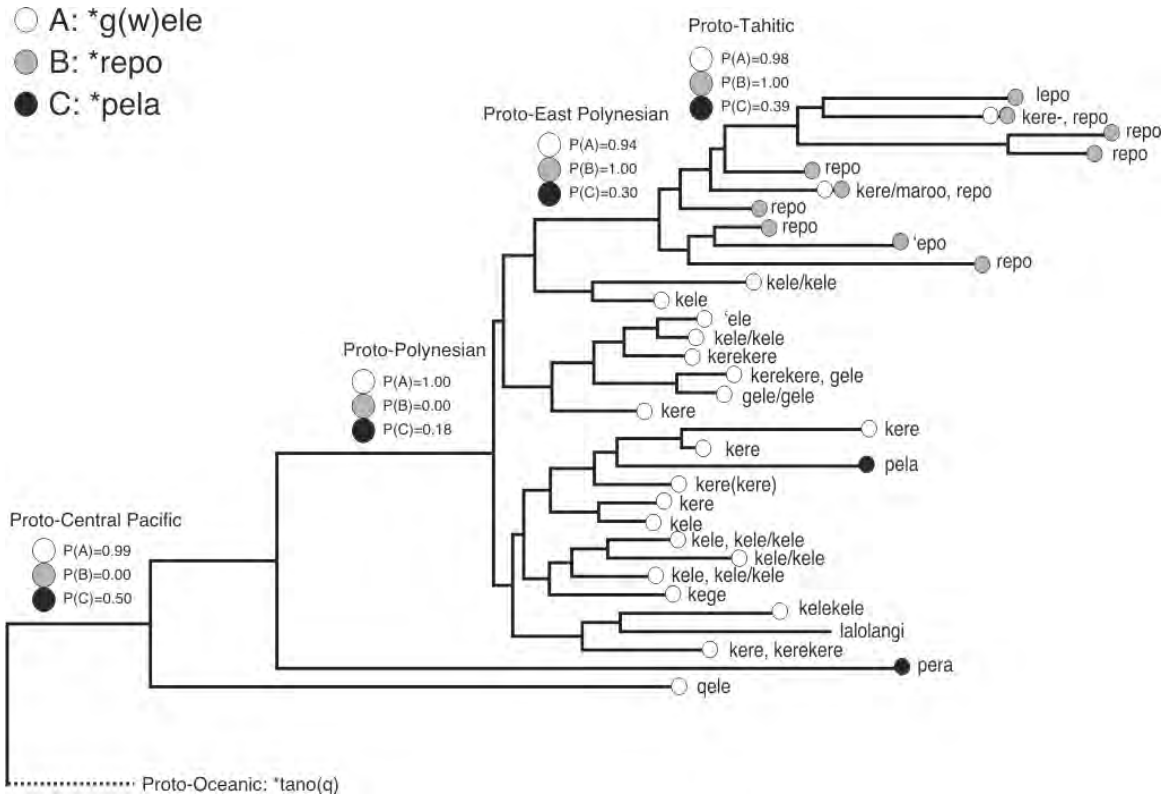


Figure 4: Tree of Central Pacific languages showing the distribution of three cognate sets A (in white), B (in gray), and C (in black) containing words for 'earth/soil'. Branch lengths are proportional to amount of change along the lineage. The probability of the ancestral states are marked for a number of protolanguages. The probability of Proto Tahitic and Proto East Polynesian belonging to cognate set B (*repo) is >0.99. Proto Polynesian and Proto Central Pacific instead contain cognate set A (*g(w)ele) also with a probability of >0.99.

4 Conclusion

We hope that this chapter has corrected most of the persistent myths and misconceptions about the application of computational phylogenetic methods to historical linguistics. Let us be very clear. Phylogenetic methods do not make the flawed assumptions of lexicostatistics or glottochronology. They do not count cognates to calculate pairwise similarity measures. Instead, the likelihood calculations are based on each cognate set and how it fits onto the tree. Phylogenetic methods do not require a single 'one size fits all' rate of lexical replacement. These methods can allow for different rates of change both between cognate sets and between different lineages. Moreover, this framework can explicitly take into account external evidence such as archaeological dates and known historical events to make robust inferences about divergence dates. In marked contrast to lexicostatistics, the phylogenetic methods we have detailed here perform exceptionally well on the very difficult problem of the Austronesian subgrouping and dating. First, the trees are rooted in Taiwan, in line with the results of the comparative method. Second, the sequence and subgrouping of these phylogenetic trees strongly reflect the structure of the family tree suggested by the comparative method, at least in those cases where there is a consensus among comparative linguists. Third, the timing of events on these trees again corresponds extremely well to the 'Out-of-Taiwan' scenario.

We also hope to have laid to rest a final vexing misconception about phylogenetic linguistics: 'this method is not giving anything new' (Jasanoff in Wade 2004:1). Not only do phylogenetic methods work well and outperform lexicostatistics, they also provide a range of new tools that can be of great benefit to linguistics. First, phylogenetic methods provide an explicit optimality criterion for evaluating how well different trees (i.e. historical scenarios) are supported by the data. Second, they provide an empirical way of assessing the statistical robustness of any subgroup in those trees. We have shown here a number of Austronesian examples where the support values on our trees coincide well with linguistic intuitions about the strength of support for these groupings. Third, despite the failure of glottochronology to provide robust date estimates, the attraction of absolute dating is strong. Dates are critically important for inferences about human prehistory. They provide a powerful way of linking linguistic, archaeological, cultural, and genetic evidence. It is not uncommon to still see glottochronological age estimates cited in publications, along with the standard disclaimer that this method cannot be trusted (e.g. Campbell 1997; Comrie 2002; Pawley 2002). Phylogenetic dating methods can, when used carefully and appropriately, help integrate our inferences about human prehistory without these glaring disclaimers. Fourth, these methods enable us to investigate how linguistic and cultural traits have evolved in families by tracing their history. These tools can infer ancestral states and can even be used to infer functional dependency between linguistic characters (Gray et al. 2007). Far from being lexicostatistics-redux, Bayesian phylogenetic methods provide exciting new tools for historical linguistics.

References

Atkinson, Q.D., G. Nicholls, D. Welch and R.D. Gray. 2005. From words to dates: water into wine, mathemagic or phylogenetic inference? *Transactions of the Philological Society* 103:193–219.

Atkinson, Q.D. and R.D. Gray. 2006. Are accurate dates an intractable problem for historical linguistics? In C.P. Lipo, M.J. O'Brien, M. Collard and S.J. Shennan, eds *Mapping our ancestors: phylogenetic approaches in anthropology and prehistory*, 269–298. New Brunswick: Aldine.

Atkinson, Q.D., A. Meade, C. Venditti, S.J. Greenhill and M. Pagel. 2008. Languages evolve in punctuational bursts. *Science* 319:588.

Balter, M. 2003. Early date for the birth of Indo-European languages. *Science* 302:1490–1491.

Balter, M. 2004. Search for the Indo-Europeans. *Science* 303:1323–1326.

Bellwood, P. 1997. *Prehistory of the Indo-Malaysian Archipelago*. Honolulu: University of Hawai'i Press.

Bellwood, P., J.F. Fox and D. Tryon. 1995. *The Austronesians: historical and comparative perspectives*. Canberra: Research School of Pacific and Asian Studies, The Australian National University.

Bergsland, K. and H. Vogt. 1962. On the validity of glottochronology. *Current Anthropology* 3:115–153.

Blust, R.A. 1977. The proto-Austronesian pronouns and Austronesian subgrouping: a preliminary report. *University of Hawai'i Working Papers in Linguistics* 9:1–15.

—— 1981. *Variation in retention rate among Austronesian languages*. Talk given to the Third International Conference on Austronesian Linguistics. Bali.

—— 1982. The linguistic value of the Wallace line. *Bijdragen tot de taal-, land- en. volkenkunde* 138:231–250.

—— 1995. The prehistory of the Austronesian-speaking peoples: the view from language. *Journal of World Prehistory* 9:453–510.

—— 1998. A note on higher-order subgroups in Oceanic. *Oceanic Linguistics* 37:182–188.

—— 1999. Subgrouping, circularity and extinction: some issues in Austronesian comparative linguistics. In E. Zeitoun and P. Jen-kuei Li, eds *Selected papers from the Eighth International Conference on Austronesian Linguistics* vol. 1, 31–94. Taipei, Taiwan: Symposium Series of the Institute of Linguistics, Academia Sinica.

—— 2000. Why lexicostatistics doesn't work: the 'universal' constant hypothesis and the Austronesian languages. In C. Renfrew, A. McMahon and L. Trask, eds *Time depth in historical linguistics*, 311–331. Cambridge: The McDonald Institute for Archaeological Research.

Bowern, C. 2006. Punctuated equilibrium and language change. In K. Brown, ed. *Encyclopedia of language and linguistics*, 286–289. Oxford: Elsevier.

Brandes, J.L.A. 1884. *Bijdrage tot de vergelijkende klankleer der westersche afdeeling van de Maleisch-Polynesische taalfamilie*. Utrecht.

Brugmann, K. 1884. Zur Frage nach den Verwandtschaftsverhältnissen der Indogermanischen Sprachen. *Internationale Zeitschrift für allgemeine Sprachewissenschaft* 1:226–256.

Bryant, D. 2006. Radiation and Network Breaking in Polynesian Language Evolution. In P. Forster and C. Renfrew, eds Phylogenetic Methods and the Prehistory of Languages, 111–118. Cambridge: McDonald Institute Press, University of Cambridge.

Bryant, D., F. Filimon and R.D. Gray. 2005. Untangling our past: languages, trees, splits and networks. In R. Mace, C.J. Holden and S. Shennan, eds *The evolution of cultural diversity: phylogenetic approaches*, 67–84. London: UCL Press.

Burnham, K.P. and D.R. Anderson. 1998. *Model selection and inference — a practical information-theoretic approach*. New York: Springer.

Campbell, L. 1997. *American Indian languages: the historical linguistics of Native America*. Oxford: Oxford University Press.

—— 2004. *Historical linguistics: an introduction.* 2nd edition. Edinburgh: Edinburgh University Press.

Comrie, B. 2002. Farming dispersal in Europe and the spread of the Indo-European language family. In P. Bellwood and C. Renfrew, eds *Examining the farming/ language dispersal hypothesis*, 409–419. Cambridge: The McDonald Institute for Archaeological Research.

Dahl, O.C. 1973. *Proto-Austronesian.* Scandinavian Institute of Asian Studies Monograph Series No.15. Studentlitteratur: Lund, Sweden.

Dempwolff, O. 1934. *Vergleichende Lautlehre des austronesischen Wortschatzes. Zeitschrift für Eingeborenen-Sprachen*. 1 Induktiver Aufbau einer indonesischen Ursprache, 15, Berlin: Deitrich Reimer.

—— 1937. *Vergleichende lautlehre des austronesischen Wortschatzes. Zeitschrift für Eingeborenen-Sprachen*. 2, Deductive Anwendung des urindonesischen auf austronesische Einzelsprachen, 17, Berlin: Deitrich Reimer.

—— 1938. *Vergleichende lautlehre des austronesischen Wortschatzes. Zeitschrift für Eingeborenen-Sprachen*. 3 Austronesisches Wörterverzeichnis, 19, Berlin: Deitrich Reimer.

Diamond, J. and P. Bellwood. 2003. Farmers and their languages: the first expansions. *Science* 300:597–603.

Dixon, R.M.W. 1997. *The rise and fall of languages*. Cambridge: Cambridge University Press.

Dunn, M., A. Terrill, G. Reesink, R.A. Foley and S.C. Levinson. 2005. Structural phylogenetics and the reconstruction of ancient language history. *Science* 309:2072–2075.

Durie, M. and M. Ross. 1996. *The comparative method reviewed: regularity and irregularity in language change*. Oxford: Oxford University Press.

Drummond, A.J., S.Y.W. Ho, M.J. Phillips and A. Rambaut. 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biology* 4:e88.

Dyen, I. 1962. The lexicostatistical classification of the Malayopolynesian languages. *Language*, 38:38–46.

—— 1965. A lexicostatistical classification of the Austronesian languages. In *Indiana University Publications in Anthropology and Linguistics: Memoir 19 of the International Journal of American linguistics*. Indiana: Indiana University.

Edwards, A.W.F. and L.L. Cavalli-Sforza. 1964. Reconstruction of evolutionary trees. In J. McNeill, ed. *Phenetic and phylogenetic classification*, 67–76. Cambridge: Cambridge University Press.

Elbert, S. H. 1953. Internal relationships of Polynesian languages and dialects. *Southwest Journal of Anthropology* 9:147–173.

Eska, J.F. and D. Ringe. 2004. Recent work in computational linguistic phylogeny. *Language* 80:569–582.

Felsenstein, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Systematic Zoology* 27:401–410.

—— 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Evolution* 17:368–376.

Fisher, R.A. 1922. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London* Series A, 222:309–368.

Forster, J.R. 1778. *Observations made during a voyage round the world.* London.

Gamble, C., W. Davies, P. Pettitt, L. Hazelwood and M. Richards. 2005. The archaeological and genetic foundations of the European population during the Late Glacial: implications for 'agricultural thinking'. *Cambridge Archaeological Journal* 15:193–223.

Gimbutas, M. 1973. Old Europe c. 7000–3500 B.C., the earliest European cultures before the infiltration of the Indo-European peoples. *Journal of Indo-European Studies* 1:1–20.

Goldman, N. 1993. Statistical tests of models of DNA substitution. *Journal of Molecular Evolution* 36:182–198.

Gordon, Raymond G., Jr. 2005. *Ethnologue: languages of the World*, Fifteenth edition. Dallas, Texas: SIL International. Online version: http://www. ethnologue. com/.

Grace, G.W. 1961. Austronesian linguistics and culture history. *American Anthropologist* 57:359–368.

—— 1964a. Movement of the Malayo-Polynesians 1500 BC to AD 500: the linguistic evidence. *Current Anthropology* 5:361–368.

—— 1964b. The linguistic evidence. *Current Anthropology*, 5:361–368.

—— 1966. Austronesian lexicostatistical classification: a review article. *Oceanic Linguistics* 5:13–31.

Graham, R.L. and L.R. Foulds. 1982. Unlikelihood that minimal phylogenies for a realistic biological study can be constructed in reasonable computational time. *Mathematical Biosciences* 60:133–142.

Gray, R.D. 2007. *Tangled trees: what do phylogenetic networks reveal about Oceania linguistic history?* Talk given to the Seventh International Conference on Oceanic Linguistics. Noumea, New Caledonia.

Gray, R.D. and Q.D. Atkinson. 2003. Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature* 426:435–439.

Gray, R.D. and F.M. Jordan. 2000. Language trees support the express-train sequence of Austronesian expansion. *Nature* 405:1052–1055.

Gray, R.D., A.J. Drummond and S.J. Greenhill. 2009. Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science*, 323:479–483.

Gray, R.D., S.J. Greenhill and R.M. Ross. 2007. The pleasures and perils of darwinizing culture (with phylogenies). *Biological Theory* 2:360–375.

Green, R.C. 1999. Integrating historical linguistics with archaeology: insights from research in remote Oceania. *Indo-Pacific Prehistory Association Bulletin* 18 (Melaka Papers), vol. 2:3–16.

Green, R.C. and M.I. Weisler. 2002. The Mangarevan sequence and dating of the geographic expansion into Southeast Polynesia. *Asian Perspectives* 41:213–241.

Greenberg, J.H. 1949–54. Studies in African linguistic classification. *South-western Journal of Anthropology* 5:79–100, 190–198, 309–317; 6:47–63, 143–160, 223–237, 388–398; 10:405–415.

Greenhill, S.J. and R.D. Gray. 2005. Testing population dispersal hypotheses: Pacific settlement, phylogenetic trees and Austronesian languages. In R. Mace, C.J. Holden and S. Shennan, eds *The evolution of cultural diversity: phylogenetic approaches*, 31–52. London: UCL Press.

Greenhill, S.J., R. Blust and R.D. Gray. 2008. *The Austronesian basic vocabulary database*: From Bioinformatics to Lexomics. *Evolutionary Bioinformatics,* 4:271–283. http://language.psy.auckland.ac.nz

Hamed, M.B. and F. Wang. 2006. Stuck in the forest: Trees, networks and Chinese dialects. *Diachronica* 23:29–60.

Hoijer, H. 1956. Lexicostatistics: a critique. *Language* 32:49–60.

Holden, C.J. 2002. Bantu language trees reflect the spread of farming across sub-Saharan Africa: a maximum-parsimony approach. *Proceedings of the Royal Society of London, B Biological Sciences* 269:793–799.

Holden, C.J. and R.D. Gray. 2006. Rapid radiation, borrowing and dialect continua in the Bantu languages. In P. Forster and C. Renfrew, eds *Phylogenetic methods and the prehistory of languages*, 19–31. Cambridge: McDonald Institute for Archaeological Research.

Holm, H.J. 2007. The new arboretum of Indo-European ‘trees’. Can new algorithms reveal the phylogeny and even prehistory of Indo-European? *Journal of Quantitative Linguistics* 14:167–214.

Houtman, F. de. 1603. *Spraeck ende Woord-Boeck, Inde Maleysche ende Madagaskarsche Taln met vele Arabische ende Turcsche Woorden*. Amsterdam.

Huelsenbeck, J.P. and B. Rannala. 1997. Phylogenetic methods come of age: testing hypotheses in an evolutionary context. *Science* 276:227–232.

Huelsenbeck, J.P., F. Ronquist, R. Nielsen and J.P. Bollback. 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* 294:2310–2314.

Huelsenbeck, J.P., B. Larget, R.E. Miller and F. Ronquist. 2002. Potential applications and pitfalls of Bayesian inference of phylogeny. *Systematic Biology* 51:673–688.

Hymes, D.H. 1960. Lexicostatistics so far. *Current Anthropology* 1:3–44.

Kern, H. 1886. *De Fidji-taal vergeleken met hare verwanten in Indonesië en Polynesië*. Verhandelingen der Koninklijk Akademie van Wetenschappen, 16, Amsterdam.

Kirch, P.V. 2000. *On the road of the winds: an archaeological history of the Pacific Islands before European contact*. Berkeley: University of California Press.

Kirch, P. and R. Green. 2001. *Hawaiki, Ancestral Polynesia: an essay in historical anthropology.* Cambridge: Cambridge University Press.

Labov, W. 1994. *Principles of linguistic change: internal factors*. Oxford: Blackwell.

Larget, B. 2005. Introduction to Markov Chain Monte Carlo methods in molecular evolution. In R. Nielsen, ed. *Statistical Methods in Molecular Evolution*, 45–62. New York: Springer.

Lees, R.B. 1953. The basis of glottochronology. *Language* 29:113–127.

Lester, L., A. Meade and M. Pagel. 2006. The slow road to the eukaryotic genome. *BioEssays* 28:57–64.

Lewis, P.O. 2001. Phylogenetic systematics turns over a new leaf. *Trends in Ecology and Evolution* 16:30–37.

Lynch, J., M.D. Ross and T. Crowley. 2002. *The Oceanic languages*. Richmond: Curzon Press.

McMahon, A. and R. McMahon. 2006. Why linguists don't do dates: evidence from Indo-European and Australian languages. In P. Forster and C. Renfrew, eds *Phylogenetic methods and the prehistory of languages*, 153–160. Cambridge: McDonald Institute for Archaeological Research.

Murdock, G.P. 1964. Genetic classification of the Austronesian languages: a key to Oceanic culture history. *Ethnology* 3:117–126.

Nakhleh, L., T. Warnow, D. Ringe and S.N. Evans. 2005. A comparison of phylogenetic reconstruction methods on an Indo-European dataset. *Transactions of the Philological Society* 103:171–192.

Nettle, D. 1999. Is the rate of linguistic change constant? *Lingua* 108:119–136.

Nichols, J. 1997. Modeling ancient population structures and movement in linguistics. *Annual Review of Anthropology* 26:359–384.

Pagel, M. 1999. Inferring the historical patterns of biological evolution. *Nature* 401:877–884.

Pagel, M. and A. Meade. 2004 A phylogenetic mixture model for detecting pattern-heterogeneity in gene sequence or character-state data. *Systematic Biology* 53:571–581.

Pagel, M., Q.D. Atkinson and A. Meade. 2007. Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. *Nature* 449:717–720.

Pawley, A. 1999. Chasing rainbows: implications of the rapid dispersal of Austronesian languages for subgrouping and reconstruction. In E. Zeitoun and P. Jen-kuei Li, eds *Selected papers from the Eighth International Conference on Austronesian Linguistics* vol. 1, 95–138. Taipei, Taiwan: Symposium Series of the Institute of Linguistics, Academia Sinica.

—— 2002. The Austronesian dispersal: languages, technologies and people. In P. Bellwood and C. Renfrew, eds *Examining the farming/language dispersal hypothesis*, 251–274. Cambridge: McDonald Institute for Archaeological Research.

Pawley, A. and R.C. Green. 1973. Dating the dispersal of the Oceanic languages. *Oceanic Linguistics* 121–2:1–67.

Pawley, A. and M. Ross. 1995. The prehistory of the Oceanic languages: a current view. In P. Bellwood, J.J. Fox and D.T. Tryon, eds *The Austronesians: historical and comparative perspectives*, 39–74. Canberra: Research School of Pacific and Asian Studies, The Australian National University.

Penny, D., B.J. McComish, M.A. Charleston and M.D. Hendy. 2001. Mathematical elegance with biochemical realism: the covarion model of molecular evolution. *Journal of Molecular Evolution* 53:711–723.

Posada, D. and K.A. Crandall. 2001. Selecting the best-fit model of nucleotide substitution. *Systematic Biology* 50:580–601.

Renfrew, C. 1987. *Archaeology and language: the puzzle of Indo-European origins*. London: Cape.

Reland, H. 1708. *Dissertatio de linguis insularum quarundem orientalium*. Trajecti ad Rhenum.

Rexovà, K., Y. Bastin and D. Frynta. 2006. Cladistic analysis of Bantu languages: a new tree based on combined lexical and grammatical data. *Naturwissenschaften* 93:189–194.

Rexovà, K., D. Frynta and J. Zrzavy. 2003. Cladistic analysis of languages: Indo-European classification based on lexicostatistical data. *Cladistics* 19:120–127.

Ringe, D., T. Warnow and A. Taylor. 2002. Indo-European and computational cladistics. *Transactions of the Philological Society* 100:59–129.

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

—— 1996. Contact-induced change and the comparative method: cases from Papua New Guinea. In M. Durie and M.D. Ross, eds *The comparative method reviewed: regularity and irregularity in language change*, 180–217. New York: Oxford University Press.

—— 1997. Social networks and kinds of speech-community events. In R. Blench and M. Spriggs, eds *Archaeology and language*, vol. 1, 209–261 London: Routledge.

Ross, M. and A. Næss. 2007. An Oceanic origin for Äiwoo, the language of the reef islands? *Oceanic Linguistics* 46:456–498.

Sanderson, M.J. 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. *Molecular Biology and Evolution* 14:1218–1231.

—— 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Molecular Biology and Evolution* 19:101–109.

Shutler, R. and J.C. Marck. 1975. On the dispersal of the Austronesian horticulturalists. *Archaeology and Physical Anthropology in Oceania* 10:81–113.

Sokal, R.R. and P.H.A. Sneath. 1963. *Principles of numerical taxonomy*. San Francisco: W.H. Freeman.

Steel, M. and D. Penny. 2000. Parsimony, likelihood, and the role of models in molecular phylogenetics. *Molecular Biology and Evolution* 17:839–850.

Suchard, M.A., R.E. Weiss and J.S. Sinsheimer. 2001. Bayesian selection of continuous-time Markov Chain evolutionary models. *Molecular Biology and Evolution* 18:1001–1013.

Swadesh, M. 1952. Lexico-statistic dating of prehistoric ethnic contacts. *Proceedings of the American Philosophical Society* 96:453–463.

Swofford, D.L., G.J. Olsen, P.J. Waddell and D.M. Hillis. 1996. Phylogenetic inference. In D.M. Hillis, C. Moritz and B.K. Mable, eds *Molecular Systematics*, 407–514. Sinauer Associates, Sunderland, MA.

Thurgood, G. 1999. *From ancient Cham to modern dialects: two thousand years of language contact and change.* Hawaii: University of Hawaii Press.

Wade, N. 2004, 16 March. A biological dig for the roots of language. *The New York Times*.

Yang, Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *Journal of Molecular Evolution* 39:306–314.

Yang, Z., N. Goldman and A.E. Friday. 1994. Comparison of models for nucleotide substitution used in maximum likelihood phylogenetic estimation. *Molecular Biology and Evolution* 11:316–324.

23 *Linguistic differences among Siraya, Taivuan and Makatau*

PAUL JEN-KUEI LI

1 Introduction

Siraya, Taivuan and Makatau were the Formosan languages or dialects formerly spoken in the southwestern plains of Taiwan.[1] Roughly speaking, Siraya was spoken in the coastal area of Tainan Plain and Taivuan mostly in the inland of Tainan Plain to the north, while Makatao was spoken in Kaohsiung and Pingtung prefectures to the south. The languages or dialects probably became extinct in the first half of the 19th century (Li 2002). Dutch missionaries left behind three main written documents, namely *The Gospel of St. Matthew in Formosan Sinkang Dialect* (Gravius 1661, henceforth *St. Matthew)*), *'t Formulier des Christendoms* (Gravius 1662, henceforth *Formulary*), and the *Utrecht Manuscript* (unknown author, published in Van der Vlis 1842).[2] Ever since then only short wordlists have been recorded in various villages in the southwest plains at different times between 1717 and 1917 by the Chinese, Europeans, and Japanese. Ogawa (1917) assembled those wordlists and classified them into three main groups: Siraya, Makatau and Taivuan. There are altogether 75 villages or sources of language data and 163 lexical entries represented in his comparative wordlist (see Tsuchida et al. 1991). Due to the paucity of language data in that area, his comparative wordlist is extremely valuable, especially for Makatau. Tsuchida (Tsuchida et al. 1991:ix) prepared a map, which shows the location of 39 villages. It uses three different signs to indicate the three different groups of languages or dialects, which gives us an idea about the geographical distribution of the erstwhile linguistic communities in the southwestern plains.

[1] An earlier version of this paper in Chinese (Li 2006) appeared in a conference proceedings. In this version I have up-dated the language data, revised the internal relationships of the three groups and added some new findings. I also discuss the affiliation of Dutch missionary documents. In preparing this paper, I benefited from Ogawa's pioneering work on Siraya as well as from Tsuchida's and Adelaar's valuable suggestions and the help of my assistants, Hsiu-min Huang and Amy Minnuan Chen. This work was supported in part with a grant from the National Science Council (NSC95-2411-H-001-010-H).

[2] Adelaar (1997:364ff.) also discusses dialect variations between the *Utrecht Manuscript* on the one hand, and *St. Matthew* and the *Formulary* on the other.

Alexander Adelaar and Andrew Pawley, eds
Austronesian historical linguistics and culture history: a festschrift for Robert Blust, 399–409.
Pacific Linguistics, 2009.
© This edition vested with Pacific Linguistics.

Based on Ogawa's comparative wordlist, Tsuchida pointed out that the three groups have different reflexes of PAn *l and *N, as shown below.

Table 1: Sirayaic Reflexes of PAn *l and *N

	PAn	Siraya	Taivuan	Makatau	
(1)	*l	r	Ø~h	r	
e.g.	*telu	turu	too, toho	toru	'three'
	*lima	rima	hima	rima	'five'
(2)	*N	l	l	n	
e.g.	*(qa)Nuang	luang	lowan	noang	'cow'
	*puNi	mapuli	mapuli	mapuni	'white'

But as Tsuchida was aware, there are many exceptions to the rules, perhaps due to the poor or inaccurate transcriptions of the language data from various sources. Assuming that the phonological correspondences given above are correct, each group differs from the other two by only one phonological innovation. These might be regarded as dialectal differences, as commonly found in other Formosan languages, such as Rukai (Li 1977) and Atayal (Li 1981).

2 Evidence from Sinkang manuscripts

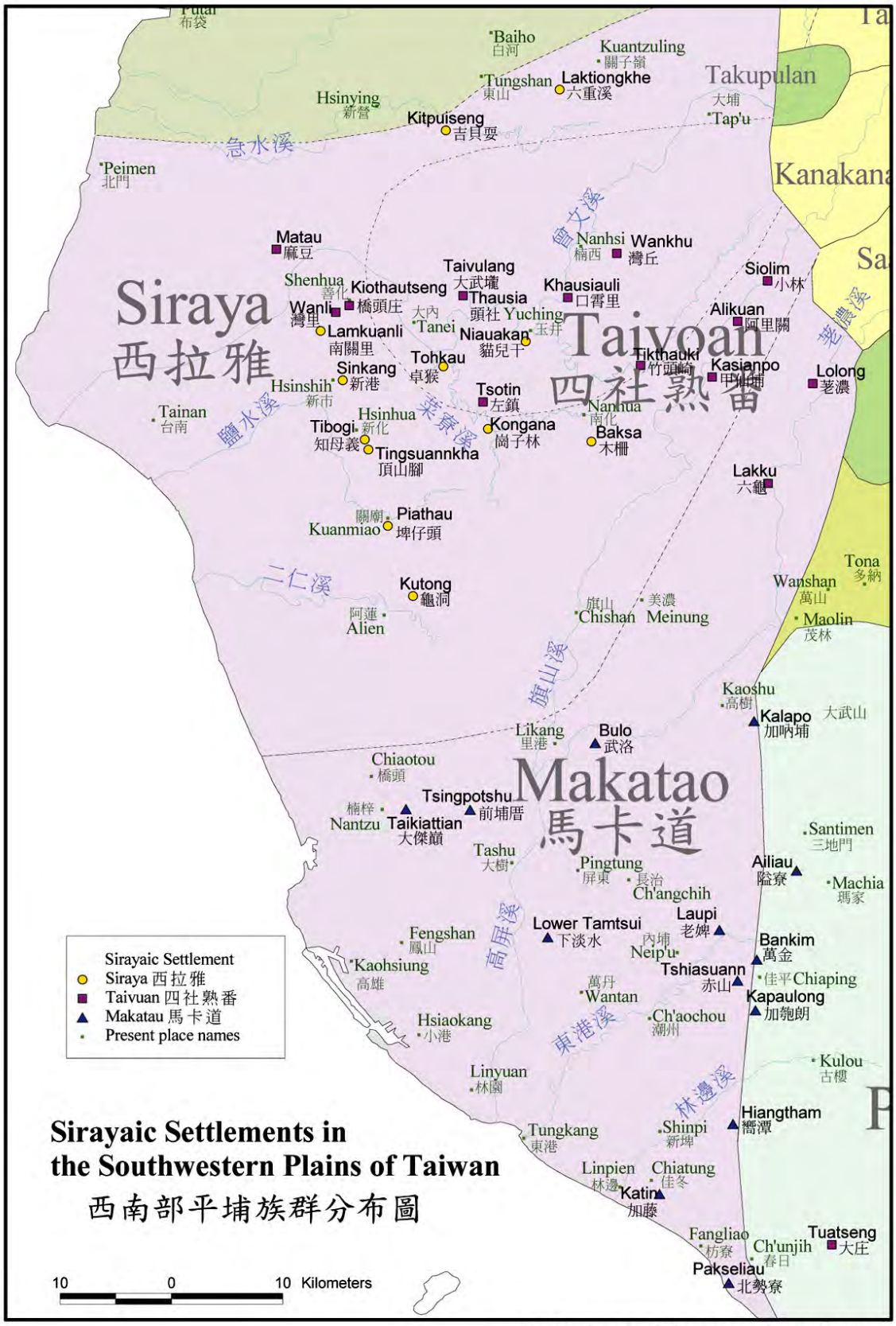
Aside from the Dutch missionary documents and the short wordlists for the language data of the southwestern plains, there is a third type of language data available: the so-called 'Sinkang manuscripts' are contracts written in Romanised script. These manuscripts or texts were found in various villages in the southwestern plains that belong to the three different groups. For example, Sinkang, Tohkau and Kongana belong to Siraya; Wanli, Matau (see below) and Taibulang belong to Taivuan; Lower Tamtsui and Katin belong to Makatau. The earliest text is dated 1663 and the latest 1818. Murakami (1933) collected 101 manuscripts.[3] My colleagues and I have accumulated 170 manuscripts.[4] The great majority came from Siraya villages, only 23 came from Taivuan villages and four from Makatau villages. Among these, 67 are written in both Chinese and a native language, while the remaining ones are monolingual. A careful study of these texts may reveal significant linguistic differences, not found in wordlists.

My assistants Hsiu-min Huang, Chin-wen Chien, and I have worked on Sinkang manuscripts in the past eight years (since 2001). Although they are extremely hard to decipher, we have tried to decipher and transcribe all of them, determine word and sentence boundaries, identify each lexical item, and give interlinear glosses and free translation for each sentence whenever possible. All 170 texts exist in the form of computer files. These texts do reveal some interesting facts about the language or dialects in the southwestern plains.

---

[3] Of the 101 manuscripts collected by Murakami, 87 are from the village of Sinkang, six from Matau, three from Tohkau, one from Taibulang, one from Lower Tamtsui, and three from Katin. The 87 Sinkang manuscripts are treated as the main body of his monograph, while the other 14 are given in appendices.

[4] Of the 170 manuscripts, one came from Backloan (Siraya, not found in Murakami), seven from Tohkau, ten from Matau, one from Taibulang (Taivuan), eleven from Wanli (Taivuan, not found in Murakami), one from Lower Tamtsui, three from Katin, and two from Gutiaupo 牛稠埔 (Siraya or Makatau, not found in Murakami).



Map Designed by GIS Joint Lab, Academia Sinica

2.1 Phonological evidence

In addition to the two phonological innovations observed by Tsuchida (1991), I have found two additional ones based on the language data in the Sinkang manuscripts, as illustrated below.

Table 2: Sirayaic Reflexes of PAn *D, *-k- and *-S-/*-R-

	PAn	Siraya		Taivuan		Makatau	
		Sinkang	Tohkau	Wanli	Matau	Lower Tamtsui	
(3)	*D	s	s	r~d	r~d	r~d	
e.g.	*Daya	saija	saija	raija	–	–	‘east’
	*lahuD	raos	raos	raur	–	–	‘west’
	*DapaN	sapal	–	rapan	–	–	‘foot’
		sa	sa	ra, da	ra, da	ra, da	‘and’
		hiso	hiso	hairo, ro	haijro, do	–	‘if, as’
		posoh	–	poroh	–	–	‘land’
		maisisang	–	–	–	maeraerang	‘magistrate’
(4)	*-k-	-k-	-k-	Ø	Ø	-k-~Ø	
	*-S-, *-R-[5]	-g-~-h-	-g-	Ø	Ø	–	
e.g.		akosaij	akusiuo	ausaij	ausaij	akusai[6]	‘not have’
		tarokaij	–	taroaej	–	tarauwei[7]	‘name’
	*DuSa	soso(h)a	–	–	–	–	‘two’
	*baqeRu	vaho	–	–	–	–	‘new’
		dagogh	dagogh	daoh	daoh	–	‘price’
		ligig	–	liih	–	–	‘sand’
		matagi-vohak	–	mataij-vohak	–	–	‘regret’

As shown in the examples above, *s* in Siraya corresponds to *r* or *d* in Taivuan and Makatau in word-initial or final position, derived from PAn *D or *d, as illustrated in (3). As shown in (4), *k* or *g~h* in Siraya is lost in Taivuan in word-medial position. The *k* in Siraya is derived from PAn *k, and the *g~h*, which is interpreted as velar fricative *x* by Adelaar (1999), is derived from PAn *S or *R (Adelaar 1999:334).

Rule (3) shows Siraya in contrast with Taivuan and Makatau, while Rule (1) shows Taivuan in contrast with Siraya and Makatau. Rule (4) shows that the medial velar obstruents *k* and *g* [x] are lost in Taivuan, but retained in both Siraya and Makatau.[8] It is an innovation in Taivuan.

---

5 The symbol ‘g,’ ‘gh’ or ‘h’ [x] in Siraya is historically derived from PAn *S or *R; see Table 6 below for examples. There is extremely limited vocabulary in all Sinkang manuscripts, and I can identify few certain PAn cognates for such a derivation in these Sinkang manuscripts , such as *sosoha* or *sosoa* ‘two’ and *v(ah)o* ‘new’ found only in Sinkang.

6 Although the term ‘not have’ is unavailable for Lower Tamtsui, the form *akusai* is cited for Taikiattian, another village of Makatau in Ogawa’s list, and the form *akosaij* appears in a manuscript from Gutiaupo, which might be another village of Makatau or Siraya (see Tsai 2002, Appendix 1, p.3).

7 The personal name *tarauwei* appears in a Katin text (Murakami 1933:144). This shows that *-k-* is occasionally lost in certain Makatau subdialects.

8 Ogawa investigated the Piathau dialect of Siraya in 1921. But he (Ogawa 2006:354) cited the form *li’igh* ‘sand’ for Siraya dialects, probably taken from *St. Matthew*; see Table 6 in §3 below.

Matau was considered to belong to the Siraya group by Chinese and Japanese scholars, as indicated in Ogawa's (1917) grouping and Tsuchida et al.'s map (1991:ix). However, the phonological innovations in Matau generally indicate that it belongs to the Taivuan group rather than the Siraya one: Matau *l > Ø or *h*, as in *telu > *tao* 'three', *lima > *hima* 'five'; and *D > *r*, *d* (see Table 2, above).

Both Matau and Wanli villages of Taivuan are in the coastal and transitional area, geographically close to Sinkang and Tohkau villages of Siraya. The phonological differences, especially (3) and (4), between Siraya, Taivuan, and Makatau, are quite regular.

2.2 Morphological evidence

In addition to the phonological differences, a type of morphological difference can be observed, as shown below:

Table 3: Sirayaic future markers

Siraya		Taivuan		Makatau	
Sinkang	Tohkau	Wanli	Matau	Lower Tamtsui	
-ali, -ili	-ati, -ili	-ah	-ah	-ani	'future'

The verbal suffix indicating future also shows that Taivuan differs from both Siraya and Makatau. Note that *-ani* in Makatau regularly corresponds with *-ali* in Siraya, another bit of morphological evidence indicating that Matau is Taivuan instead of Siraya.

2.3 Lexical evidence

Tsuchida et al. (1991:7–8) pointed out the following lexical differences among the three groups, as based on Ogawa's comparative wordlist:

Table 4: Lexical differences in Sirayaic

Siraya	Taivuan	Makatau	
siraya	taivoan	makatao	'aborigines (self-appelation)'
it	tau	lihu	'liquor, wine'
pak	harau	buka	'rice'

However, there has also been a fair amount of lexical borrowing between the groups. The term for 'aborigines' in Matau is *siraiya* instead of the anticipated form *taivoan*. The term for 'wine' in Kongana (Siraya) is *diho* instead of the anticipated form *it*; see Texts 5 and 11 published in Li (2002). Similarly the term for 'wine' in *St. Matthew* is also *dihou*. The term for 'rice' in Laupi (Makatau) is *pak* instead of the anticipated form *buka*. In short, unlike phonological innovations, lexical evidence is not very reliable for language subgrouping.

2.4 Summary

For the sake of convenience, let's summarise what we have discussed so far in Table 5.

Table 5: Summary of Sirayaic Reflexes of PAn *l, *N, *D, *-k- and *-S-/*-R-

	PAn	Siraya	Taivuan	Makatau
(1)	*l	r	Ø~h	r
(2)	*N	l	l	n
(3)	*D, *d	s	r~d	r~d
(4)	*-k-	-k-	Ø	-k-~Ø
	*-S-, *-R-	-g-	Ø	–

In addition, Taivuan has a suffix *-ah* indicating 'future', which is different from both Siraya *-ali/-ili* and Makatau *-ani*.

Two phonological innovations, (*l > Ø or *h*, and loss of medial consonants *-k, *-S- and *-R-), plus one piece of morphological evidence, show that Taivuan is in maximal contrast with Siraya and Makatau. Makatau differs from both Siraya and Taivuan only in Rule (2), which is not significant. It shares Rules (1), (4) and the correspondence seen in Sirayaic future markers with Siraya, but it shares only Rule (3) with Taivuan. As a preliminary conclusion, it seems that Taivuan constitutes a first split from the Sirayaic group, while Siraya and Makatau are more closely related, as shown below:[9]

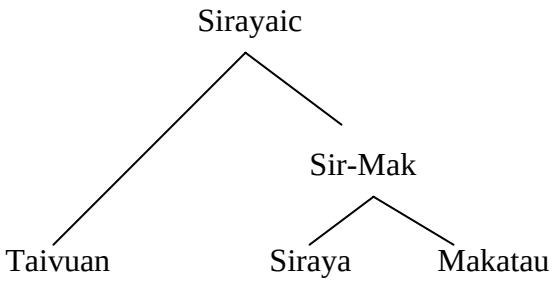


Figure 1: Sirayaic tree diagram based on numbers of shared phonological innovations

It seems reasonable to classify Taivuan as a separate language, but it is not clear whether Siraya and Makatau can be treated as separate languages until we find more linguistic difference.

3 The Dutch Missionary documents

Is *St. Matthew* based on Siraya or Taivuan? The following vocabulary found in *St. Matthew* seems to indicate that it is Taivuan:

[9] Adelaar's (pers. comm.) interpretation of Table 5 is that there is no clear subgrouping pattern emerging. He also notes that in Rule (2) in Table 5, Makatau *n* would be an innovation, rather than a retention, if PAn is reconstructed as *L.

(5) *D, *Z > r, d, cf. (3) above

> *Daya > *reia* ‘east’, *lahuD > *raour* ‘west’, *DapaN > *rahpal* ‘foot’, *likuD > rikour ‘back’, *ZaNum > *raloum* ‘water’, *da* ‘and’ (*sa* in Sinkang), *pourough* ‘land’ (*posoh* in Sinkang), *haouroung* ‘to steal’ (*haosong* in Sinkang), *laumari* ‘coins’ (*lomasi* in Sinkang), *ka-harim-auh* ‘will be forgiven’ (*ka-hasim-ing* in Sinkang)

(6) Loss of *-k-* and *-g-* [x] < *S, cf. (4) above

- Loss of *-k-*: *(i)aku > *jau*, *-au* ‘1sg’; *aousi*, *akousi* ‘not have’

The loss of intervocalic *-k-* seems to be restricted to a few special grammatical categories, namely personal pronoun and negative, and note the free variant of *aousi ~ akousi*. It does not apply to ordinary vocabulary, e.g. *bukeS > *voukugh* ‘hair’, *likuD > *rikour* ‘back’ and *takut > *takout* ‘to fear’.

- Loss of *-g-*: *daœugh*, *daœuh* ‘price’ (*dagogh* in Sinkang); *li'igh*, *liih* ‘sand’ (*ligig* in Sinkang)

Exceptions (*-S- is retained as *-h-*): *CaSiq > *t<m>ahy* ‘to sew’, *DuSa > *rouha*, *douha* ‘two’

Both the phonological innovations, *D, *Z > *r, d* and loss of intervocalic *-k-* and *-g-* [x], and the suffix *-au* or *-auh* ‘future’ (cf. Table 3 above) indicate that *St. Matthew* was most likely based on some dialect of Taivuan, rather than Siraya proper. However, that there are exceptions to the loss of intervocalic *-k- and *-S- > -g-, -h- [x] seem to indicate there might be mixture of dialects in *St. Matthew*, as suggested by Adelaar (pers. comm.), it is ‘most likely that *the Gospel* text [=*St. Matthew*] was not the product of one person only: this is clear from the text itself, and ... that there was a committee deciding over the final edition.’ In fact, it is stated in the introduction to *St. Matthew*:

> Hence, too, it follows that the present Translation can be of service to only a few, though populous Villages, such as Soulang, Mattauw, Cinckan (Sinkang), Bacloan, Tavokan, Tevorang, and perhaps also to some of the people in Dorko and Tilocen. These are the places where the work has been carried on for the longest time (p.xiii).

Similarly, it is most likely that *Formulary* (Gravius 1662) was also based on some dialect of Taivuan, whereas *Utrecht Manuscript* was based on a dialect of Siraya, as shown in the following comparison:[10]

[10] Both Adelaar (1997) and Tsuchida (1998) make a comparison of these three Dutch missionary documents. Adelaar (2006) notes two main dialects, the ‘Gospel dialect’ and the ‘Utrecht Manuscript dialect.’ He considers *St. Matthew* and *Formulary* to represent the same Gospel dialect.

Table 6: A comparison of Utrecht, Matthew and Formulary

		Utrecht Ms	St. Matthew	Formulary	
(7)	*D	s	r, d	r, d	
e.g.	*Daya	taga-seia	reia	–	‘east’
	*lahuD	taga-raos	raour	–	‘west’
	*DapaN	sapal	rahpal	rahpal	‘foot’
	*DuSa	so-soa	dou-rouha	rou-rouha	‘two’
	*likuD	ricos	rikour	rikour	‘back’
		sama	rama	dama	‘father’
		sa	ra	ra	‘but’
		soo, sou	rou	dou, rou	‘if, as, when’
		isang	irang	irang	‘great, large’[11]
		sasim	rarim	mou-rarim	‘down, below’
		pesanach	paeraenaeh	paeraenaeh	‘tree’
		massou	–	marou	‘corn’
		ka-pousoch-ang	pourough	pourogh	‘land’
(8)	*-k-	-k-, Ø	-k-, Ø	Ø	
e.g.		acoussey	akousi, aousi	aoussi	‘not have’[12]
	*(i)aku	iau, -au	jau, -au	jau	‘I’
(9)	*S/*R	-h-, Ø/-g-, -h-	-h-, Ø/-h-, ', Ø	-h-, Ø/h, ', Ø	
e.g.	*kaSu	cau [kaw]	kow	kow	‘thou’
	*DuSa	so-soa	dou-rouha	rou-rouha	‘two’
	*CaSiq	t<m>ahy	–	–	‘sew’
	*waRi	wagi	wæ'i	wæ'i	‘sun, day’
	*wiRi	ougi	u-i	ou-i	‘left’
	*baqəRu	vacho	vahæu, va'æu	vahæu,va'æu	‘new’
	*kiRim	k<m>igim	k<m>i'im	k<m>i'ym	‘seek’
	*kaRaC	k<m>agat	–	–	‘bite’
		ligig	li'igh	–	‘sand’
		ma-dagoa	–	dæeua	‘entirely’
(10)		-a, -al, -ale	-ah, -auh	-a, -ah, -al	‘future’

The examples in (7) show that the language of *Utrecht Manuscript* is similar to that of the Siraya-speaking villages of Sinkang and Tohkau, while the language of *St. Matthew* and *Formulary* is similar to that spoken in the Wanli and Matau villages of Taivuan or Lower Tamtsui village of Makatau, as illustrated in (3) above. However, the pronominal forms, first person and second person singular, and the negative in (8) and (9) do not show much

[11] *isang/irang* is the root of the derivation *maisisang/maeraerang* in Table 2. It is cognate with *ma-‘iDang* ‘big’ in Puyuma, as suggested by Tsuchida (pers. comm.).

[12] As based on Adelaar’s (pers. comm.) research, these are two different negatives: *assei, assey* or *assi* means ‘no(t),’ while *akousi* means ‘not have,’ which is derived from *akou-* ‘to have’ + *(a)ssi* ‘not.’ Still another negative is *inang* ‘will not.’

difference between the groups. The medial *-k-* is kept or lost in the negative forms in *Utrecht Manuscript* and *St. Matthew*, but lost in *Formulary*. All the three groups have zero reflex of *S in the form of *kaSu. On one hand, *Utrecht Manuscript* has zero reflex of *S in the form of *DuSa, while *St. Matthew* and *Formulary* have retained *h* as its reflex. On the other hand, while *-g-*, the reflex of *S or *R is retained in the lexical forms *wagi* 'sun,' *ougi* 'left,' *k<m>igim* 'seek,' *ligig* 'sand' and *ma-dagoa* 'entirely' in *Utrecht Manuscript*, it is lost in the forms *wœ'i, u-i, k<m>i'im*, and *li'igh* in *St. Matthew* and the similar forms in *Formulary*. That is to say, there is some conflicting evidence. But several examples seem to indicate that *Utrecht Manuscript* is based on a Siraya dialect, while *St. Matthew* and *Formulary* are based on a Taivuan dialect. The suffix *-ah* 'future' in (10) also indicates that *St. Matthew* and *Formulary* are based on a Taivuan dialect. It seems clear that there is some dialect mixture in both *St. Matthew* and *Formulary*.

4 Relative chronology and subgrouping

The sound change PAn *D, *d, *Z > *s* in Siraya, > *r* ~ *d* in Taivuan and Makatau must have taken place prior to the Dutch occupation of Taiwan (1624–62), as the phonological difference is manifested in the Dutch missionary documents: *s* is found in the *Utrecht Manuscript* vs *r* ~ *d* in *St. Matthew* and the *Formulary*, as discussed in the preceding section.

The change *-S-, *-R- > *x* (written as *g, gh, h*) or Ø may have started at the beginning of the 17th century because the rule applies to some lexical forms, but not to the others containing the consonant even in the same set of language data as recorded by the Dutch missionaries; see Table 6 above.

PAn * l is retained as *r* rather than *h* or Ø (see (1) above) in *St. Matthew*, e.g. *lahuD > *raour* 'west,' *piliq > *peri* 'to choose,' *kalih > *k<m>ari* 'to dig'. That is to say, PAn *l was still retained as *r* in mid 17th century when *St. Matthew* was translated. In fact, it was still retained in a Wanli text dated 1770, as in *lahuD > *raur* 'west,' and as *h* in a Matau text dated 1781, as in *likuD > *mi-likoh* 'to return.' It was not lost until much later when the Japanese started to investigate the languages of the southwestern plains in 1897. So the sound change *l > *h* or Ø was a late innovation limited to Taivuan.

If we take the relative chronology of the sound changes into consideration, then the first split of the Sirayaic group should be Siraya, as shown in Figure 2 below.

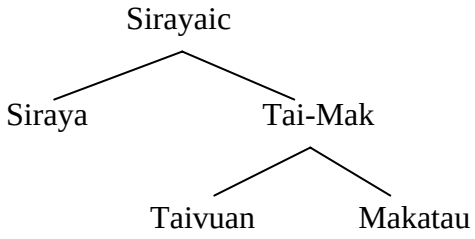


Figure 2: Sirayaic tree diagram based on the chronological order of phonological innovations

However, if we compare the number of shared phonological innovations, then the first split would be Taivuan, as shown in Figure 1 above. Which type of evidence should carry more weight: the earliest phonological innovation or the number of shared phonological innovations? It seems to me the former should carry more weight in a subgrouping hypothesis.

References

Adelaar, K. Alexander. 1997. Grammar notes on Siraya, an extinct Formosan language. *Oceanic Linguistics* 36.2:362–397.

—— 1999. Retrieving Siraya phonology: a new spelling for a dead language. In Elizabeth Zeitoun and Paul Li, eds *Selected papers from the Eighth International Conference on Austronesian Linguistics*, 313–354. Institute of Linguistics (Preparatory Office). Taipei: Academia Sinica.

—— 2006. Siraya dialogues. In Henry Y. Chang, Lillian M. Huang and Dah-an Ho, eds *Streams converging into an ocean: festschrift in honor of Professor Paul Jen-kuei Li on his 70th Birthday*, 665–686. Language and Linguistics Monograph Series Number W-5. Institute of Linguistics. Taipei: Academia Sinica.

—— 2007. Siraya, Taiwan's oldest written language. In *The margins of becoming: identity and culture in Taiwan*, ed. by Carsten Storm and Mark Harrison, 19–34. Wiesbaden: Harrassowitz.

Chen, Bien-horn. 2001. *Notes on vocabulary of Siraya version of Matthew* (in Chinese). Association of the Plain Tribes in Tainan.

—— 2005. *Notes on formulary of Christianity in Formosan Siraya dialect—with translation in English, Chinese and Dutch* (in Chinese). Private Circulation.

Gravius, Daniel. 1661. *Het Heylige Euangelium Matthei en Johannis Ofte Hagnau Ka D'llig Matiktik, Ka na sasoulat ti Mattheus, ti Johannes appa.* Amsterdam: Michiel Hartoch.

—— 1662. *Patar Ki Tna-'msing-an Ki Christang, ka Taukipapatar-en-ato tmaeu'ug tou Sou KA MAKKA-SIDEIA, 't Formulier des Christendoms Met de Verklaringen van dien, Inde Sideis-Formosaansche Tale*. Amsterdam: Michiel Hartogh.

Li, Paul J. K. 1977. The internal relationships of Rukai. *Bulletin of the Institute of History and Philology, Academia Sinica* [BIHP] 48.1:1–92.

—— 1981. Reconstruction of proto-Atayalic phonology. *BIHP* 52.2:235–301.

—— 1993. New data on three extinct Formosan languages. *BIHP* 63.2:301–323.

—— 2002. Preliminary interpretations of the 15 recently uncovered Sinkang manuscripts (in Chinese). *Taiwan Historical Research* 9.2:1–68.

—— 2006. The languages of the plain tribes in southern Taiwan—with remarks on the linguistic position of Matau (in Chinese). In Chuen-rong Yeh, ed. *Constructing Siraya: Selected Conference Papers*, 17–38. Tainan: Government of Tainan.

Murakami, Naojirô. 1933. Sinkan manuscripts. Memoirs of the Faculty of Literature and Politics, Taihoku Imperial University, vol. 2, no.1. Formosa: Taihoku Imperial University.

Ogawa, Naoyoshi. [1917]. Siraia, Makatao, Taivoan. Manuscripts.

—— 2006. *A comparative vocabulary of Formosan languages and dialects*, ed. by Paul Li and Masayuki Tohoshima. Asian and African Lexicon Series No.49. Research Institute for Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies.

Steere, Joseph Beal. 1874. [The aborigines of] Formosa. *Journal of the American Geographical Society of New York* 6:302–334. New York.

—— 2002. *Formosa and its inhabitants*. Edited by Paul Li. Institute of Taiwan History (Preparatory Office), Academia Sinica.

Tsai, Cheng-ui. 2002. Transfer of land ownership of the Sinkang villagers in Tianliao district during the Ching Dynasty (1736–1895) (in Chinese). Unpublished MA thesis, National Tainan Normal University.

Tsuchida, Shigeru. 1996. Personal pronouns of Siraya (Formosa). In Bernd Nothofer, ed. *Reconstruction, classification, description. Festschrift in honor of Isidore Dyen*, 231–247. Hamburg: Abera Verlag.

—— 1998. English index of the Siraya vocabulary by Van der Vlis. *Studies of Taiwan Aborigines* 3:281–310.

Tsuchida, Shigeru, Yukihiro Yamada and Tsunekazu Moriguchi. 1991. *Linguistic materials of the Formosan Sinicized populations I: Siraya and Basai.* University of Tokyo.

Utrecht Manuscript. n.d. Vocabularium Formosanum. MS. Utrecht University Library.

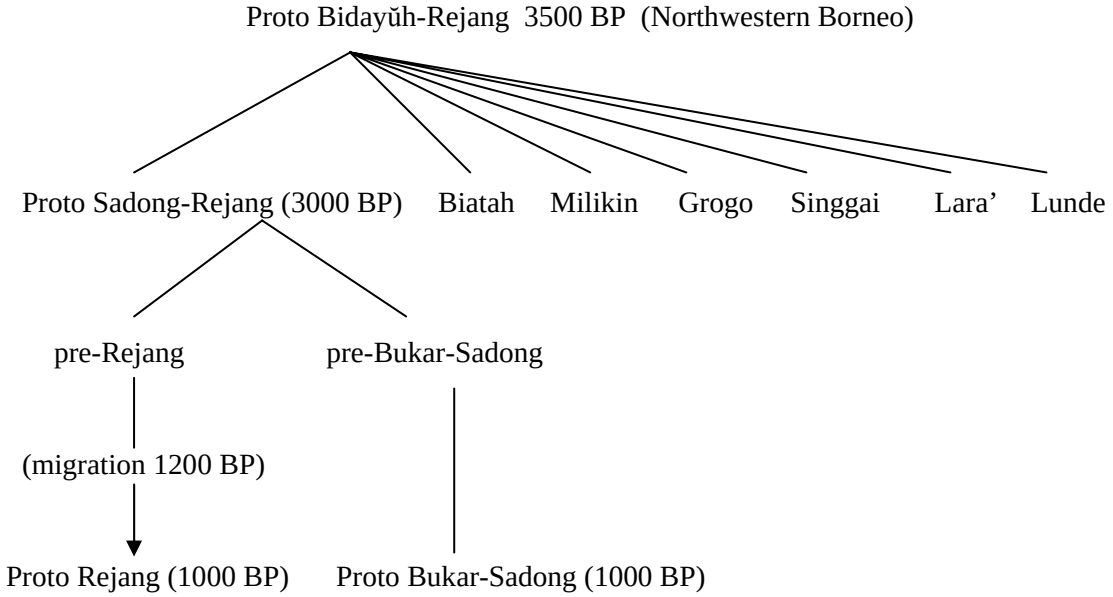
Van der Vlis, C.J. 1842. Formosaansche woorden-lijst, volgens een Utrechtsch Handschrift. Voorafgegaan door eenige korte aanmerkingen betreffende de Formosaansche taal. *Verhandelingen van het Bataviaasch Genootschap* 18:437–452 (Notes), 453–483 (Glossary), 484–488 (Conversation).

24 *Out-of-Borneo subgrouping hypothesis for Rejang: re-weighing the evidence*

RICHARD McGINN

This paper revisits a subgrouping hypothesis for the Rejang language of Sumatra presented in McGinn (2003), and offers a new hypothesis based on new evidence and on criticisms received from Austronesianists. The hypothesis to be defended is shown in (1).

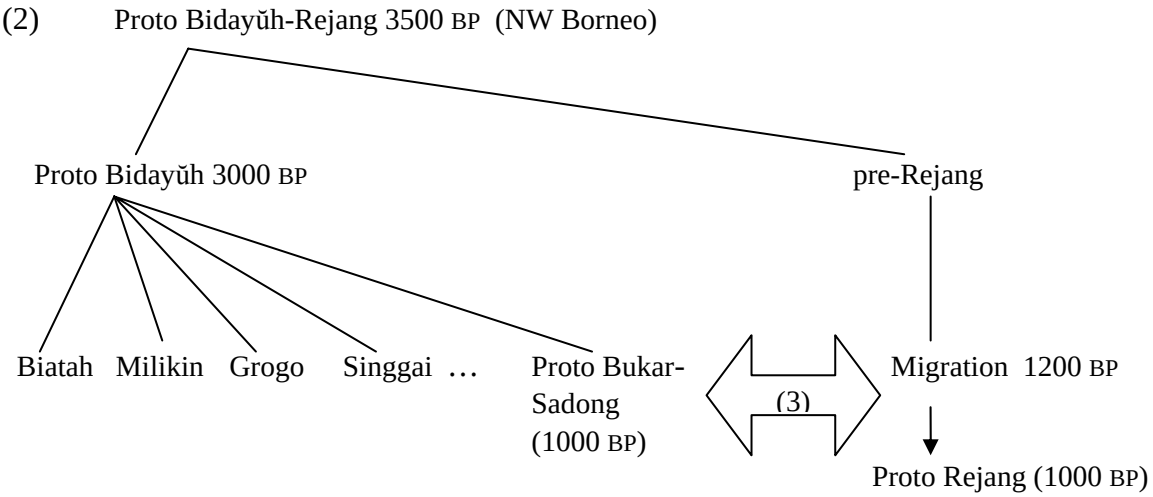
(1)



The subgrouping hypothesis shown in (1) above replaces the proposal by this writer (2003) illustrated in (2).

Alexander Adelaar and Andrew Pawley, eds
Austronesian historical linguistics and culture history: a festschrift for Robert Blust, 411–424.
Pacific Linguistics, 2009.
© This edition vested with Pacific Linguistics.

(2)



In earlier publications (McGinn 1999, 2000, 2003), we attempted to construct an 'out-of-Borneo' subgrouping hypothesis for Rejang, and in particular, McGinn (2000) suggested the possibility that Generalised PMP *a Raising, illustrated in (3a-c) below, might constitute a set of innovations shared by Rejang and one or more Land Dayak languages. The following changes were among the first to distinguish pre-Rejang from PMP.

(3)	a.	*-ˈVCaC# > ˈVCəC[-velar]#	PMP unstressed *a > /ə/ except before velars.
	b.	*-ˈVCa# > ˈVCə#	PMP unstressed *a > /ə/ in word-final position.
	c	*-ˈVCaC# > VCˈVC#	PMP stress pattern (symbolised by preceding ˈ) shifted to the word-final syllable.

These three changes were originally posited by McGinn (1997) solely for the sake of supporting the regularity hypothesis against the challenges posed by Blust's pioneering article (1984), and only later were those same changes used as the basis for subgrouping arguments (McGinn 2000, 2003). The Stress Shift change (3b) was motivated by several factors, the most compelling being that it served to increase the regularity of the other sound changes, thereby explaining many apparent irregularities. The most important change for subgrouping purposes, however, is (3a) because the conditioning factor ('except before velars') is unusual and phonetically unmotivated. (See §4). Moreover, a comparable change had been recorded for the Tapŭ and Mawang (labeled Měntu) dialects of Bukar-Sadong Bidayŭh by Court (1967a) and Topping (1990), and confirmed by this writer during field work on six dialects[1] of Bukar-Sadong conducted in 2000 and 2001.

Exploring the possibility that (3a) represented a shared innovation linking Rejang and Bukar-Sadong, McGinn (2003) provided further evidence to recommend a subgrouping hypothesis, including counterparts of (3a,b,c) (see (7) below). But there was also contradictory evidence in the form of an apparently regular sound change, namely *l > r, which affected all Bukar-Sadong dialects and several other Bidayŭh languages, but not Rejang. Adjusting the hypothesis to accommodate the contrary evidence, McGinn (2003) concluded that *l > r must have preceded *a Raising in Bukar-Sadong, and if so, *a Raising was not a shared innovation after all; therefore, *a Raising must have been the result of rule borrowing after a long period of language contact (presumably in Borneo).

---

[1] The six dialects surveyed were Tibakang, Ranchan, Mujat, Tapŭ, Mawang and Badŭp. Not included was Kampung Bunun, a Bukar-Sadong language with a six-vowel system as described by Asmah Haji Omar (1983:445), who claims that 'the phoneme /a/ is phonetically [ʌ].'

Problems with the borrowing theory were soon pointed out by Robert Blust (personal communication)[2] and David Zorc (2006:509); and Adelaar (2007) questioned my assumption that sometimes Bidayŭh /l/ directly reflects PMP *l. However, the criticisms were not wholly satisfactory, since taken together, they posed a paradox for the Comparative Method: the basic comparison (3a) seemed to resist explanation by either chance, borrowing or common inheritance. In particular:

(4) a. Chance is ruled out by the unusual conditioning of (3a).
    b. Borrowing of (3a) is ruled out because unsupported by independent lexical evidence.
    c. Direct inheritance of (3a) is ruled out owing to the contradictory evidence of *l > /r/ in many Bidayŭh languages but not Rejang.[3]

If (4a-c) offers a paradox, then one or more of the statements must be hiding a false assumption. It is theoretically impossible for all three to be valid.

1 The status of PMP *l > /r/ in Oceanic

A possible way out of the dilemma posed by (4a-c) follows from a comment by John Lynch which implies that (4c) may be erroneous.[4] The following was his reaction to the treatment of *l > r in McGinn (2003).

(5) At least in Oceanic, the *l > r change is very common, and in many cases not diagnostic of subgrouping. In other words, it seems to be a natural change which could easily occur independently. Looking at Tryon's New Hebrides (Vanuatu) survey, for example, there are a couple of cases in Malakula where groups of languages show *l > r but where, on other criteria, these languages subgroup with other languages which show *l > l. In Tanna, *l and *r (and *R when reflected) have merged, in some languages as l, in others as r. Similar kinds of things have occurred, as far as I am aware, in other areas of Oceanic.

    While I would not dismiss *l > r as an innovation, I think it is a weak one — rather like palatalisation of *t before *i: something that doesn't surprise you when you see it, as it happens so often; and therefore something not to be given much weight if there are other, less expected, innovations which would support a different subgrouping theory.

    (John Lynch pers. comm.)

---

[2] Blust (pers. comm.) advised the following experiment: 'Line up the basic vocabularies of Rejang and any two or three L(and)D(ayak) languages and pull out all of the exclusively shared lexical innovations. Do you find any? If not, this is reason to suspect that Rejang is not a recent arrival in Sumatra from the LD area.' In fact, so far we have found vanishingly few shared lexical innovations (listed in McGinn 2003), to which can be added Tibakang lʌcaʔ 'soaked' and Rejang ləcəaʔ 'soaked'.

[3] This last point bears repeating for clarity's sake: the Bukar-Sadong version of (3a), namely (7ii), while shared with Rejang, is not shared with other Bidayŭh languages; and Bukar-Sadong *l > /r/, while shared with many (but not all) Bidayŭh languages, is not shared by Rejang.

[4] The following is the statement that Lynch was addressing: 'The Bukar-Sadong version of PMP *a 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 (*a underwent raising 'except before velars') in exactly these two languages. Therefore, I shall argue for borrowing as the more likely explanation.' (McGinn 2003:49)

If the above comment is relevant to the concerns of this paper, then what is needed is evidence that *l > r in Proto Bidayŭh is indeed a weak innovation. Accordingly, we shall investigate the second conjunct in the following statement to determine if perhaps it might be too strong.

(6) The characteristic features of the Land Dayak languages which Hudson mentions include distinctive numerals for 'eight', 'nine' and 'ten' *and /r/ as the reflex of PAN *l*
(Kroeger 1998:150–151, citing Hudson 1978; emphasis ours)

According to Paul Kroeger (1998:139), 'The Land Dayak languages do not appear to be closely related to any other language in Sarawak, but they do form a linguistic subgroup with the many Land Dayak languages spoken across the border in West Kalimantan (Indonesian Borneo)'. This statement provides the context for Hudson's assertion, quoted above, claiming that PMP *l > /r/ is an important diagnostic feature of Land Dayak (= Bidayŭh) languages. Examination of the available evidence within this language group suggests, however, that the pattern of regular *l > /r/ is not as widespread throughout the family as was initially thought; therefore, any attempt to construct a protolanguage for the Bidayŭh group must do so on the basis of something other than regular *l > *r. This claim will be substantiated below.

2 The status of PMP *l > /r/ in Bidayŭh languages

In fact, the reflexes of PMP *l are problematic among Bidayŭh languages. Five sources of evidence support this claim.

First, the data displayed in Ray (1913) indicates that whereas most of the thirteen Land Dayak languages surveyed consistently show expected PMP *l > r, two of the languages—Grogo and Milikin—regularly retain PMP *l as /l/.

Second, in the Biatah-Bidayŭh language spoken in Sarawak, whereas most dialects regularly reflect PMP *l as /r/, the Mbaan dialect regularly retains PMP *l as /l/. (Kroeger Ms, 1994:22)

Third, Adelaar (2007), citing unpublished field notes, suggests that in another Bidayŭh language of West Kalimantan (Sungkung), /l/ appears to reflect intermediate *r which itself reflects the merger of PMP *R, *r and *l, implying that PMP *l > *r was ancient and /l/ a recent innovation. If this model is proven by future research to account for the organic /l/s in other Bidayŭh languages, it would strengthen the case for *l > /r/ as a diagnostic for Bidayŭh languages, and weaken if not break the case for a Rejang connection. On the other hand, logically and phonetically it is just as easy to assume that *R, *r and *l merged as intermediate *l in some language, then split into /l/ and /r/ in daugher languages, or changed unconditionally to /r/. Clearly, what are needed are detailed historical phonologies for individual dialects, where the results can be tested against the strict demands of the regularity hypothesis (see McGinn 1997:91–92 for discussion). In the meantime, Lynch's comments about Oceanic, cited above, remain as a valid cautionary note.

Fourth, among the six Bukar-Sadong dialects surveyed by this writer during field work in 2000 and 2001 and partially displayed in McGinn (2003), the reflexes of PMP *l seem to vary unpredictably between /l/ and /r/ for five of the dialects; and for the sixth dialect (Mujat) there is a three-way alternation between /l/, alveolar /r/ and uvular /ɣ/. In fact, out of a total of thirty-one potential reflexes of PMP *l, Proto Bukar-Sadong shows fifteen cases of *r (see McGinn (2003) examples 38, 39, 46, 103, 107, 108, 136, 158, 163, 168, 186, 187, 200, 242) and fifteen of *l (see the same source for examples 12, 19, 27, 50, 66,

93, 101, 105, 109, 135, 149, 153, 189, 233, 243). Whereas it is likely that some (perhaps all) putative B-S /l/ < PMP *l are Malay borrowings, this remains to be demonstrated in a definitive historical phonology of the Bukar-Sadong language group. For starters, a candidate for borrowing includes one member or the other of the doublet /jʌrʌn/ 'road' and /jalʌn/ 'walk' in the Tapŭ, Ranchan, Bedŭp and Mawang dialects, which show contrastive /r/ and /l/ corresponding to PMP *l; another may be the PBS outcome *m-ilih 'choose' corresponding to PMP *piliq 'choose'. PBS *milih contains suspicious /l/ and suspicious -/h/ (expected -/ʔ/), exactly like Malay *pilih* 'choose'. Other possibly problematic examples include PMP *bales = PBS *maləs 'reply'; PMP *gatel = PBS *gatəl 'scratch', PMP *palaqepaq = PBS *kilapa 'palm frond', and PMP *balaŋa = PBS *b[a,i]laŋaʔ 'clay pot'.

Fifth, Bukar-Sadong dialects overwhelmingly agree with respect to /r/ and /l/ as apparent reflexes of PMP *l. Therefore, even if all the unexpected /l/s are the result of massive borrowing, the borrowing would have occurred very early, in Proto Bukar-Sadong; otherwise the dialect uniformity is unexplained.

These comparative problems justify the decision to re-weigh the subgrouping value of *l > /r/ in relation to the Bidayŭh languages in general, and Bukar-Sadong in particular. We now have grounds to set the problem to one side; it is an anomaly to be investigated, not yet a counterexample; it is too weak to bear any weight for subgrouping purposes. By implication, much more importance can and should be given to the set of comparisons (3a,b,c), especially (3a). Just how much weight to assign to (3a) will be taken up in the next section.

3 Back to the basic comparison

As mentioned above, the most important change for subgrouping purposes is (3a) because the conditioning factor ('except before velars') is unusual and phonetically unmotivated. (See next section for arguments.) If so, then the discovery that both the Rejang dialect group in Sumatra and the Bukar-Sadong dialect group in Sarawak, show unmistakable traces of a similar change in their phonological histories, constitutes *prima facie* evidence for a shared innovation.

However, change (3a) is directly relevant only for Rejang, because it resulted in partial merger of PMP *a and *e as *ə (schwa). By contrast, the Bukar-Sadong counterpart did not result in merger. Therefore, to be consistent with hypothesis (1), the Bukar-Sadong facts require an additional (and perfectly natural) assumption to be added to the phonological history of Rejang, namely, that (3a) occurred in two steps as shown in (7ii) and (7iv) below.

(7)	i.	(3b)	*-ˈVCa# > ˈVCə#	Unstressed *a > /ə/ in word-final position.
	ii.	(3a-1)	*-ˈVCaC# > ˈVCʌC[-velar]#	Unstressed *-aC > *-ʌC except before velars.
	iii.	(3c)	Stress Shift (Language split)	Vowels in final syllables became stressed
	iv.	(3a-2)	*-VCˈʌC# > VCˈəC[-velar]#	Stressed *ʌC > -əC in Rejang (partial merger)

Thus, only the first step (7ii) is claimed to be a shared innovation, whereas the partial merger (7iv) occurred in pre-Rejang after language split.

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

The following formula represents five early changes in the historical phonology of pre-Rejang.

(8) i.

PMP		pre-Rejang	(Unstressed *a > *ə except before velars.)
*a	>	*ə / ˈVC__(C[-velar])#	

ii.

	PMP		pre-Rejang	PMP	Kebanagung	Gloss
a.	*a	>	*ə / ˈVC __ #	*duha	dui	‘two’
b.	*a	>	*ʌ / ˈVC __ C[-velar] #	*taŋan	*taŋʌn	‘hand’
c.	*-aw	>	*əw	*danaw	danəa	‘lake’
d.	*-ay	>	*əy	*matay	matəe	‘die’
e.	*ʌ	>	*ə	*taŋan	taŋən	’hand’

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

The set of pre-Rejang changes shown by the formula in (8i) *almost* works for reconstructed pre-Bukar-Sadong as well—only the partial merger of *ʌ and *ə is missing (8ii,e). Consider the following set of changes in Bukar-Sadong, illustrated by the Tibakang dialect.

(9)

	PMP		pre-Bukar-Sadong	PMP	Tibakang	Gloss
a.	*a	>	*ə / ˈVC __ #	*duha	duəh	‘two’
b.	*a	>	*ʌ / ˈVC __ C[-velar] #	*taŋan	tʌŋʌn	‘hand’
c.	*-aw	>	*əw . . . > u	*danaw	danu	‘lake’
	*-ay	>	*əy . . . > i	*punay	puni	‘dove’

To help explain all of these changes, we 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.

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

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

(10)

PMP	Pre-Rejang	Pre-Bukar-Sadong	Tibakang	Gloss
*duha	*ˈduə	*ˈduə	duˈəh	‘two’
*mata	*ˈmatə	*ˈmatə	baˈtəh	‘eye’
*naŋa	*ˈnaŋə	*ˈnaŋə	naˈŋəh	‘fork of river’
*limə	*ˈlimə	*ˈlimə	liˈməh	‘five’
*nia	*ˈniʔə	*ˈniʔə	niˈʔəh	‘he/she’

3.2.2 Neutralisation of PMP word-final *-a in diphthongs

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

(11)	PMP	Pre-Bukar-Sadong	Proto Bukar-Sadong and Tibakang	Proto Rejang and Rawas	Gloss
	*danaw	*danəw	danu	daniw	‘lake’
	*punay	*punəy	puni	punuy	‘dove’
	*qatey	*atəy	ati	atuy	‘liver’
	*hapuy	*apuy	apuy	apuy	‘fire’
	*kahiw	*kaiw	kayu	kiiw	‘wood’

3.2.3 Raising of PMP *a in closed final syllables ‘except before velars’

Data like that shown below is what first drew my attention to the comparison of Rejang and Bukar-Sadong. The unusual conditioning of PMP *a *except before velars* was first reported for the Měntu-Tapŭ dialect of Bukar-Sadong by Christopher Court (1967), and for the Musi dialect of Rejang by Robert Blust (1984).

(12)	PMP	Rejang (Rawas)	Bukar-Sadong (Tibakang)	Gloss	McGinn (2003) (Appendix)
	*bulan	buˈlən	buˈrʌtn	‘moon’	38
	*quzan	uˈjən	uˈjʌtn	‘rain’	214
	*tawad	taˈwəa	tawˈʌr	‘haggle’	184
	*anak	aˈnak	aˈnak	‘child’	3
	*bataŋ	baˈtaŋ	baˈtakŋ	‘trunk (of tree)’	15
	*hasaq	aˈsah	ŋ-aˈsaʔ	‘sharpen’	6

In the same Appendix, see also items: 13, 34, 44, 46, 93, 112, 146, 147, 165, 173, 182, 186, 203, 204, 217, 232 242.

This comparison is the strongest evidence of a greater-than-chance subgrouping relationship between Rejang and Bukar-Sadong. Section 4 provides the justification for this claim.

Finally, there are drift-theoretical comparisons between Rejang and Bukar-Sadong dialects which were listed in McGinn (2003) and repeated in Sets I and II of (13) below. David Zorc (2006:509) reviewed this evidence and judged it as ‘indeed plausible’ with respect to an out-of-Borneo migration theory for Rejang. Much more specifically, the following comparisons are also consistent with hypothesis (1) of this paper.

(13) Rejang and Bukar-Sadong	Widespread in Borneo	Shared by Malay
Set I		
*Ca- > *Cə- in trisyllables	YES	YES
*qa- > Ø in trisyllables	YES	NO
*z > *j (except Rejang d- in ‘road’ and ‘needle’)	YES	YES
*-eq > -aC; elsewhere *-eC > *-əC	YES	POSSIBLY
*-q > *-ʔ	YES	NO
*-mb-, -nd- > -m^b-, -n^d-, etc. (‘barred nasals’)[5]	YES	NO
*-m, *-n > ^{p}m, ^{t}n, etc. (pre-stopped nasals)[6]	YES	NO

[5] The distinction between Plain and Barred (post-stopped) nasals precludes the necessity of recognizing phonemic nasalised vowels. See Court (1967) and Scott (1964) for discussion in relation to Bukar-Sadong; see Coady and McGinn (1983) for analysis of a similar issue in Rejang.

Set II = (7)		
Stress on final syllable	YES	NO
*-a > *-ə	PROBLEMATIC[9]	PROBLEMATIC[7]
*-aC > *-ʌC except before velars	NO	NO
Set III (morphology)		
Loss of suffixes in a language with word-final stress	YES	NO
Retention of PMP completive infix *-in- reanalysed as passive morpheme	YES	NO
Loss of *p- and *b- in transitive active verbs, e.g. *piliq > *m-ilih* 'choose'; *pinzem > *m-inj̃əm*; *bili > *m-irih* 'buy'	YES	NO

In the aggregate, the evidence presented in McGinn (2000, 2003) and this paper offers compelling reasons to believe that the Rejangs originated in Borneo (rather than, say, Taiwan, the Philippines, Sulawesi, or the Malay peninsula). Furthermore, the evidence is consistent with the much stronger claim represented by (1), namely, that Rejang belongs in a lower-order subgroup with Bukar-Sadong Bidayŭh.[8]

4 Arguments against a drift-theoretical explanation of the basic comparison

The crucial sound change (7ii) can and must be reconstructed independently for pre-Proto Rejang and pre-Proto Bukar-Sadong, and for no other languages (including no other Bidayŭh languages), as far as is known at present. Therefore, the research question concerns how this comparison should be explained. The only possibilities are chance (phonetic drift) and shared innovation—implying direct inheritance from a common ancestor language, as illustrated in (1) above. (Borrowing has been excluded owing to the paucity of lexical evidence showing intimate contact between the two languages.) This section presents arguments against the drift-theoretical explanation.

The basic claim is that the conditioning of change (7ii), namely that the change occurred except before velars, is phonetically unexpected and therefore unlikely to have occurred in both Rejang and Bukar-Sadong as the result of mere chance. The crucial issue

---

6 For discussion of Rejang's pre-stopped nasals, see Coady and McGinn (1983:442) and Voorhoeve (1955).

7 Tadmor (2003) argues that *-a > *-ə* spread by borrowing from Sanskrit via Javanese during the Majapahit period (1293–1520), and subsequently affected scores of Malay dialects and numerous other western Austronesian languages, regardless of whether the affected vowel was stressed or unstressed, under the poliical sway of the Majapahit empire. Our claim is that Rejang and Bukar-Sadong underwent a similar change indpendently, and much earlier, which affected only unstressed vowels.

8 A note on reflexes of PMP *j is in order owing to astute comments by a reviewer. Rejang dialects reflect PMP *j as -/g/- between vowels and -t/-k word-finally, with -/k/ being the most frequent; however, -/t/ < *-j is the outcome for the dialect judged by McGinn (2005) to be the most conservative (Rawas). The solution to this problem adopted by McGinn (2003, 2005) was to retain *-j at the level of Proto Rejang, as a direct retention from PMP. The reviewer noted that whereas most Bidayŭh languages reflect PMP *j as -/d/- and -/d/, this is not the case for Lara', Bekati' and possibly Lundu, which reflect *j as -/g/- between vowels. It is apparent to me that the dialect/language splits in the Rejang and Bidayŭh groups point to a shared retention. PMP *j was simply retained as *j at the highest level (Proto Sadong-Rejang). After pre-Rejang split from Proto Sadong-Rejang, the pre-Bidayŭh languages developed independent reflexes for PMP *j as noted by the reviewer, and much the same thing happened in the Rejang group. This assumption accounts for the comparisons and preserves the hypothesis.

is whether the comparison (7ii) is *rare enough* to indicate a subgrouping hypothesis, given that the supporting evidence consists of common sound changes. Theoretically, the subgrouping value of (7ii) falls towards zero to the extent that its basis is phonetic and natural, but by the same token, if the basis is phonological and arbitrary, then its subgrouping value rises accordingly.[9] So why did final velar consonants block raising/neutralisation of PMP *a in pre-Bukar-Sadong and pre-Rejang?

Throughout, we shall assume the generalised version of PMP *a Raising shown in (3) above and analyzed further in (7) above, which involves three shared innovations in Rejang and Bukar-Sadong. Crucially, all of the raised reflexes of PMP *a that underwent Raising were unstressed; and the Raising of *-aC to *-ʌC occurred in all environments except before velars. Our proposal is that Stress Shift occurred after *a Raising had begun to spread, but before the spreading process was complete. In other words, Stress Shift (7iii) interrupted the spread of *a Raising (7i) and (7ii).

Recall that *a Raising only affected unstressed vowels—a phonetically well-motivated assumption. The primary change was probably (7i) affecting unstressed word-final position *-a before word boundary. Next, *a Raising spread to include word-final *-aC except when the final -C was a velar (*-q, *-k, *-ŋ). Left unchecked, the spreading should have generalised totally; so why, indeed, did the spreading stop? We doubt very much that it had anything to do with phonetic naturalness. The relevant question is: why did velar consonants check the spread of PMP *a neutralisation? Is it because velars offer more resistance to airflow from the lungs than, say, labials and alveolars? Does *a neutralisation require more air than the anticipation of a velar can provide? Such a line of questioning seems unlikely to lead to a satisfactory explanation.

A more likely explanation is that the spread of *a Raising was blocked by a competing sound change, namely, (7iii)—Stress Shift. This rule altered the stress pattern from trochaic to iambic, and in the process, would have affected negatively *any* rule in the process of spreading among unstressed vowels. This introduces the element of arbitrariness which is so important in a subgrouping argument. The outcome was 'unnatural' in the sense of Blevins (2004), since the synchronic rule shows neutralisation of a stressed vowel; however, the individual (and sometimes competing) sound changes which produced the outcome were all perfectly natural.

5 Phonetic and phonological effects in sound change

The explanation just offered has in part a phonetic basis and in part a phonological one. Phonetically, it is necessary to assume that *a neutralisation rules, such as (7i) and (7ii), affected only unstressed vowels. What cannot be motivated phonetically is the actual form of (7ii), namely, the fact that velar consonants blocked the spread of the change, whereas labials, dentals, alveolars, and even semivowels and zero, did not. See (11)–(12) above.

The phonological part of the argument benefits from the assumption that the changes raising PMP *a affected a phonological system, and that the system was disrupted by a competing prosodic change, Stress Shift. This is the assumption behind (3) and (7) above. If changes (7i-iv) were systemically connected, then it is convenient to assume that change (7i)—raising of *-a in open final syllables—was the primary change. After all, this change was clearly *phonologically* motivated; it completed the distribution of PMP *ə (schwa),

[9] This is essentially the form of the argument put forward by Blust (2006 and earlier work) in support of a subgroup he called Proto North Sarawak. See §7 and fn 11.

which did not occur word-finally in PMP. Second, a classic structuralist assumption holds that sound changes are regular because they tend to generalise (or spread) within a phonological system, allophone by allophone (Bloomfield 1929). The model allows for the situation that any generalizing sound change may compete with other sound changes, producing unexpected effects and even sometimes 'crazy rules' in contemporary languages (Bach and Harms 1972; Blevins 2004). Such rules are not caused by any lack of regularity of sound change, but by the effects of competing sound changes. As expressed by Blevins (2004:44–45), 'Changes which occur in the course of evolution are random … and (do) not necessarily result in a more symmetrical, more stable, or generally improved phonological system.'

6 After-effects of rule (7ii) in Bukar-Sadong and Rejang

The Stress Shift change (7iii) had important drift-theoretical effects in the two languages. Most importantly for this paper, after the protolanguage split into pre-Rejang and pre-Bukar-Sadong, the output of (7ii), namely *-ʌC from PMP *-aC, developed differently. At one and the same time, however, both languages developed seven-vowel systems,[10] and vowel harmony rules which appear to have operated regressively at first, but evolved into synchronic phonological rules operating progressively (and somewhat unnaturally in the sense of Blevins (2004)).

6.1 Bukar-Sadong: a new vowel phoneme /ʌ/

In Bukar-Sadong, rule (7ii) added a new allophone [ʌ] which subsequently evolved into a new phoneme /ʌ/ (contemporary orthographic ĕ). *Ex hypothesi* the new phoneme /ʌ/ began as an allophone in word-final (unstressed) position before Stress Shift, and later, after becoming a stressed vowel, gained phonemic status owing to the effects of a vowel harmony rule. In particular, after Stress Shift had converted allophonic [ʌ] into a stressed vowel, it served as trigger for a harmony rule which targeted the destressed reflexes of PMP *a, e.g. *zalan > *jarʌn > /jʌrʌn/ [jʌrʌtn] 'road' (all dialects). A full analysis of this harmony rule remains for future research.

6.2 Rejang: merger of *[ʌ] with *ə

By contrast, in Rejang the outcome of (7ii), namely *-ʌC, merged with the reflex of PMP *-eC after the break-up of the protolanguage, becoming Rejang -əC in all dialects. This outcome converged with the outcome of rule (7i), which also partially merged PMP *-a and *-e as *-ə before splitting into Proto Rejang *-əy, *-i, and *-o (McGinn 1997, 2005). These changes yielded two further, and closely-related, effects: (a) schwa came to bear a heavy functional load in the inherited four-vowel system, and (b) the lexicon became governed by height harmony based on the feature [±low] (McGinn 1999:226), as follows. Firstly, all words containing the neutral vowel (schwa) became harmonised by default, since schwa was harmonic with every vowel. Secondly, words lacking a schwa underwent eight harmonic changes, e.g. *manuk > *monok chicken'; *laŋit > *läŋät 'sky';

[10] Court (1967)) and Topping (1990) ascribe seven-vowel systems to a number of Bukar-Sadong dialects, including the two non-peripheral (central) vowels we have transcribed as /ʌ/ and /ə/ (traditional orthographic ĕ and ŭ respectively). However, Topping uses the symbol ə to represent ĕ (my /ʌ/) and the symbol + to represent ŭ (my /ə/).

*sapu > *supu ‘broom’; *tali > *tili ‘rope’ (McGinn 1997, 2005; cf. Blust 1984). In the process, Rejang added two new vowels to the phonemic inventory: mid-back /o/ and low front /ä/, owing to the effects of vowel harmony. (A third new vowel, mid front /e/, was added via borrowing from unknown sources (McGinn 2005), resulting in a seven-vowel system for Proto Rejang, and attested in contemporary Rawas.

An interesting twist is that Rejang’s harmony rules applied more or less simultaneously with Stress Shift, affecting the newly stressed final vowels and de-stressed penult vowels. But phonologically, the pattern evolved into a set of inviolable ‘crazy’ rules. In contemporary Rejang, as first noted by Blust (1984), penultimate mid-vowels /e/ and /o/ always co-occur with like vowels in ultimate syllables. According to McGinn (1997, 2005), Rejang’s synchronic mid-vowel harmony rule, which applies progressively, evolved from a historical rule that applied regressively. The synchronic rule is ‘unnatural’ in the sense of Blevins (2004), because unstressed vowels trigger harmony in stressed vowels; but as expected, the historical explanation consists entirely of natural changes.

7 Conclusion

This paper has attempted to explain an unusual comparison by hypothesizing that it was a shared innovation between the Rejang dialect group of Sumatra, Indonesia, and the Bukar-Sadong dialect group of Sarawak, Malaysian Borneo. The comparison involves a change neutralizing PMP *a in word-final syllables ‘except before velars’. In our current state of knowledge, only these two languages show evidence of this comparison, which we have attributed to a common ancestor consisting of just these two languages, named Proto Sadong-Rejang, which was a daughter of Proto Bidayŭh-Rejang. (See (1).)

The principal arguments of this paper are of two types—both concerned with the problem of weighing evidence in comparative linguistics. The first argument concluded that one piece of evidence, namely *l > r in Bidayŭh languages but not Rejang, is virtually weightless on phonetic grounds, i.e. because it is far too ‘natural’, unpredictable, and ubiquitous to provide useful subgrouping information; hence this evidence has been ignored. The second argument was just the opposite, contending that another piece of evidence should be weighted heavily, namely, PMP *a Raising (neutralisation) in word-final syllables except before velars. What is odd about this change is the conditioning, which (we contend) cannot be explained on phonetic grounds. Moreover, *a Raising occurred in a ‘real’ phonological system being buffeted by a pair of competing sound changes: the spread of *a Raising among unstressed vowels, and Stress Shift, which caused unstressed vowels to become stressed (and vice versa). The competition from Stress Shift blocking the spread of *a Raising resulted in the odd conditioning (‘except before velars’) of *a Raising, traces of which are very much in evidence in contemporary Rejang and Bukar-Sadong dialects, and in no other languages, as far as is known. (See (13).)

Two precedents in the literature lend some theoretical support for the form of our argument. First, Adelaar’s (1992) reconstruction of Proto Malayic demonstrates that a valid subgrouping hypothesis may be supported solely on the basis of common sound changes if there is a sufficient variety of them, in effect interpreting a rich enough *array* of changes as typologically unusual and therefore significant for subgrouping purposes (Adelaar lists eleven such changes as the basis of reconstructing Proto Malayic. In fact, as shown in (13), Proto Rejang and Proto Bukar-Sadong share an impressively rich array of common changes. At the other extreme, Blust (2006 and earlier work) presents a subgrouping hypothesis for Proto North Sarawak based almost exclusively on evidence of

languages exhibiting reflexes and drift-theoretical effects attributed to a series of rare voiced aspirates.[11]

Our approach shares features with each of these two precedents. Like the evidence for Proto North Sarawak, the evidence for Proto Sadong-Rejang is weighted heavily in favor of a single odd comparison. And like the evidence for Proto Malayic, Proto Sadong-Rejang is also supported by an interesting array of sound changes that are not at all unusual, especially in Borneo. Hypothesis (1) proposes an explanation.

Appendix: The reconstructed phonemes of Proto Rejang and Proto Bukar-Sadong

Proto Bukar-Sadong phonemes are based on the data presented in Appendix B of this paper. Proto Rejang inventories are from McGinn (2005).

Table 1: Proto Rejang phonemes

PR Consonants (23)						PR Vowels (7)			
Stops and Affricates	*p	*t	*c	*k	*ʔ	High	*i		*u
	*b	*d	*z	*g	*j [g^y]				
Fricative		*s				Mid	*e	*ə	*o
Plain Nasals	*m	*n	*ñ	*ŋ					
'Barred' Nasals	*m̄	*n̄	*ñ̄	*ŋ̄		Low	ä		*a
Liquids	*l		*r						
Semivowels	*w		*y			Diphthongs (2)	*iw	*uy	

PR *ʔ was glottal stop; PR *r was presumably a velar or uvular liquid (reflected as *h* or *ʔ* or zero in contemporary dialects); PR *ä was low, front and unrounded (reflected as /ä/ in Rawas); and the series /*m̄, *n̄, *ñ̄, *ŋ̄/ represents the 'barred nasals' (Coady and McGinn 1983). They are regular reflexes of PMP consonant sequences *-mb-, *-nd-, *-nz- and *-ŋg-, respectively.

See McGinn (2005) for extensive discussion of the evidence for Proto Rejang based on data from five contemporary dialects.

---

[11] According to Blust (2000:285), 'The 15– 20 languages of northern Sarawak form a linguistic subgroup … defined primarily by a single sound change that left typologically unusual traces in the phonology of its members, including a set of true phonemic voiced aspirates (not murmured stops) *bh, dh, gh* in Bario Kelabit, corresponding to implosive stops in Bintulu and various Lowland Kenyah dialects, and a synchronic alternation of *b* and *s* in Kiput, reflecting *bh.' As noted by Kroeger (SMJ 1998:145), '(Blust) argues that even though no other significant phonological changes have been found, the Vowel Deletion rule is so well-attested and so unlikely to have spread by borrowing that it must be regarded as outweighing all other kinds of evidence, e.g. lexical isoglosses (Blust 1974a:220).'

Table 2: Proto Bukar-Sadong phonemes

PBS Consonants (23)						PBS Vowels (7)			
Stops and Affricates	*p	*t	*c	*k	*ʔ	High	*i		*u
	*b	*d	*j	*g				*ə	
Fricative		*s			*h	Mid	*e		*o
Plain Nasals	*m	*n	*ñ	*ŋ				*ʌ	
'Barred' Nasals	*m̄	*n̄	*ñ̄	*ŋ̄		Low			*a
Liquids	(*l)	*r							
Semivowels	*w		*y			Diphthongs (3)	*ʌy	*w	*uy

Symbols have the usual phonetic values except the 'barred' nasals (for which see Scott 1964 and Court 1967a, b and 1970).

References

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

—— 2007. Review of John Lynch, ed. *Issues in Austronesian historical phonology*. Canberra: Pacific Linguistics. *Bijdragen tot de Taal-, Land- en Volkenkunde* 163/1:139–146.

Asmah Haji Omar. 1983. *The Malay peoples of Malaysia and their languages*. Kuala Lumpur: Dewan Bahasa dan Pustaka.

—— 1992. An overview of linguistic research on Sarawak. In Peter W. Martin, ed. *Shifting patterns of language use in Borneo*. BRC Proceedings Series vol. 3. Williamsburg, VA: The Borneo Research Council.

Bach, Emmon and Robert Harms. 1972. How do languages get crazy rules? In R. Stockwell and R. Macaulay, eds *Linguistic change and generative theory*, 1–21. Bloomington: Indiana University Press.

Blevins, Juliette. 2004. *Evolutionary phonology*. Cambridge: Cambridge University Press.

Bloomfield, Leonard. 1929. A note on sound change. *Language* 4:99–100.

Blust, Robert A. 1984. On the history of the Rejang vowels and diphthongs. *Bijdragen tot de Taal-, Land- en Volkenkunde* 140.4:422–450.

—— 2000. Low-vowel fronting in Northern Sarawak. *Oceanic Linguistics* 39.2:285–319.

—— 2006. The origin of the Kelabit voiced aspirates: a historical hypothesis revisited. *Oceanic Linguistics* 45.2:311–338.

—— (MS, 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 *GAVA'. Studies in Austronesian languages and cultures dedicated to Hans Kähler.* Veröffentlichungen des Seminars für Indonesische und Südseesprachen der Universität Hamburg Band 17; 437–449. Berlin: Dietrich Reimer Verlag.

Court, Christopher. 1967a. Some areal features of Měntu Land Dayak. *Oceanic Linguistics* 6:46–50.

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

1970. Nasal harmony and some Indonesian sound laws. In *Pacific Linguistic Studies in honour of Arthur Capell*, 203–217. Canberra: Pacific Linguistics.

Dahl, Otto Christian, 1951, *Malgache et Maanyan. Une comparaaison linguistique.* Olso: Egede Instituttet.

Hudson, A.B. 1970. A note on Selako: Malayic Dayak and Land Dayak languages in Western Borneo. *Sarawak Museum Journal* XVIII (36–37):301–318.

—— 1978. Linguistic relations among Bornean peoples with special reference to Sarawak: an interim report. In *Sarawak: linguistics and development problems*. Williamsburg, VA: Studies in Third World Societies No.3, 1–45.

Kroeger, Paul R. 1994. MS. The dialects of Biatah. Borneo Research Council, 3rd Biennial Meeting, 10–14 July, 1994.

—— 1998. Language classification in Sarawak: a status report. *The Sarawak Museum Journal* 74:137–173.

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

—— 2000. Where did the Rejangs come from? In Marlys Macken, ed. *Proceedings of the Tenth Annual Conference of the Southeast Asia Linguistics Society*, 247–262. Tempe: University of Arizona.

—— 2003. Raising of PMP *a in Bukar-Sadong Land Dayak and Rejang. In John Lynch, ed. *Issues in Austronesian historical phonology*, 37–64. Canberra: Pacific Linguistics.

—— 2005. What the Rawas dialect reveals about the linguistic history of Rejang. *Oceanic Linguistics* 44.1:12–64.

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

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

Tadmor, Uri. 2003. Final /a/ mutation: a borrowed feature in Western Austronesia. In John Lynch, ed. *Issues in Austronesian historical phonology*, 15–36. Canberra: Pacific Linguistics.

Topping, Donald M. 1990. A dialect survey of the Land Dayaks of Sarawak. In James T. Collins, ed. *Language and oral traditions in Borneo*, 247–274. Borneo Research Council Proceedings Series vol. 2. Williamsburg, VA: The Borneo Research Council.

Voorhoeve, P. 1955. *Critical survey of studies on the languages of Sumatra*. Martinus Nijhof: ’s Gravenhage.

Zorc, David. 2006. Review of John Lynch, ed. *Issues in Austronesian historical phonology*. Canberra: Pacific Linguistics. In *Oceanic Linguistics* 45.2:505–516.

25 *The position of Makuva among the Austronesian languages of East Timor and Southwest Maluku*

AONE VAN ENGELENHOVEN

For Bob Blust. May this be a small shoot that can be grafted somewhere on the Austronesian Tree that you have been pruning and cultivating so well.

1 Introduction[1]

Makuva is spoken in the easternmost sub-district of Tutuala in the Republic of East Timor. Speakers are concentrated in the villages of Loiquero and Porlamano, which together make up the administrative centre of Mehara municipality. Makuva is known in the literature under several names: Loikera, Lóvaia or Lóvaia Epulu and Maku'a.[2] In this paper we will refer to this language as Makuva, a term which was introduced by Hull and Branco (2002/03).

The main language in Tutuala district is a form of Fataluku, which is non-Austronesian. A third language in this region, Rusenu, used to be spoken directly on and around Ilikerekere. It was generally considered to be extinct, but two remaining elderly semi-speakers were found in January 2007. Research is being conducted to determine whether this isolect is a language of its own or a dialect of either Fataluku or Makuva. Recent literature generally acknowledges Hajek's (1995) claim that Makuva is Austronesian,[3] although there is a

1 The research underlying the present paper was done within the framework of the Fataluku Language Project (2005–08) funded by the Netherlands Organisation for Scientific Research and initially with a pilot grant from the Hans Rausing Endangered Languages Programme (2003–04). I would like to thank Juliette Huber, Justino Valentim Cailoru and the editors of this volume for their input and help.

2 See Engelenhoven and Valentim (2006).

3 For a discussion on the older perception of Makuva being non-Austronesian, refer to Engelenhoven and Valentim (2006).

Alexander Adelaar and Andrew Pawley, eds
Austronesian historical linguistics and culture history: a festschrift for Robert Blust, 425–442.
Pacific Linguistics, 2009.
© This edition vested with Pacific Linguistics.

debate on whether the language is originally Timorese (Hull 2004a) or introduced from Southwest Maluku (Hull and Branco 2002/03).

This paper assesses the phonological evidence provided in Hull and Branco (2002/03) and shows that Makuva is closely related to both the Kairui-Waimaha-Midiki-Nauti dialect chain in the districts of Manatuto, Viqueque and Baucau and the Austronesian isolects spoken on the islands directly off the coast in Southwest Maluku (Indonesia). Section 2 discusses the youngest reconstruction proposals for the Timor and Southwest Maluku languages and the position of Makuva in them.

Section 3 presents the sound system of Makuva, and the sound changes that it has undergone. These are assessed together with the sound changes of other Timorese and Southwest Malukan languages in §4 in order to demonstrate the position of Makuva in the Timoric subgroup. Section 5 concludes with a classification showing Makuva as an early offshoot of the East group of Extra-Ramelaic languages.

2 The Proto Timoric hypothesis and Proto Luangic-Kisaric

There is a long-standing surmise in the literature that suggests a close genetic relation between the Austronesian languages of Timor and Southwest Maluku.[4] In 1998, Geoffrey Hull proposes a single ancestor for the Timor languages. In the same paper, he uses lexical evidence to demonstrate the affinity between Makuva and Meher, which is spoken on the island of Kisar off the coast in Southwest Maluku. He classifies Makuva as a Malukan language imported from a nearby island rather than a Timorese language. His evidence follows in a later publication (Hull and Branco 2002–03). In his 1998 paper, Hull also dismisses Blust's (1993) proposal for a Proto Central Malayo-Polynesian subgroup (CMP) in the Austronesian language tree, using lexical data from Nusa Tenggara Timur, East Timor and Southern Maluku. Hull's classification proposes three language groups, Bima-Flores-Sumba, Timor, and South Maluku, which he labels as 'Florinic', 'Timoric' and 'Arafuric', respectively. These three groups are analysed as descendants of a single ancestor labelled 'Proto Santalic', which is closely related to the languages spoken in Southeast Sulawesi. The plausibility of Proto CMP and the genetic relation between the languages of Timor and Southeast Sulawesi are beyond the topic of the present paper, and we confine ourselves to Hull's (1998) observations on Proto Timoric, the ancestor of the Austronesian languages on Timor.

Hull explains that his research was initially based on so-called 'Holle' lists and 'Swadesh' lists. The diversity in the lexicons studied turned out to be a major problem significantly hampering the diachronic study of the languages in Timor. He therefore devised a special Timor list of about a hundred items containing nouns, verbs, adjectives and numerals that have been attested in most Timoric languages. The other items in the lexicons are identified either as direct or indirect loans from either Malay or Portuguese, or, when not identifiable, as elements from one or several 'pre-Austronesian' languages that were spoken in Timor before the introduction of the Austronesian and non-Austronesian languages Bunaq, Makasai, Makalero and Fataluku.

Hull compares the items in his 'Timoric core vocabulary list' with reconstructed Austronesian protoforms. From that comparison he distils two subgroups based on sound

[4] See Engelenhoven (2004:12–19).

changes. The Ramelaic group[5] contains three subgroups, West Ramelaic (Tukudede and Kemak), Central Ramelaic, to which belong the Mambai dialects, and East Ramelaic, to which belongs the Idaté-Isní-Lakalei dialect chain. All the languages in this group are spoken in the Ramelau mountain range in East Timor, with the exception of Lolein, an endangered language in Dili (Hull 2004a).

The Extra-Ramelaic group[6] contains four subgroups whose languages are spoken elsewhere on Timor. Helong and the Roti and Dawan dialects belong to the western subgroup and are all spoken in Indonesian West Timor and the East Timorese enclave of Oecussi. The Central subgroup contains all Tetum dialects, Bekais and Habun. Bekais, initially considered to be a Tetun dialect, was recategorised as a separate language within this group by Hull (2002/03). The position of Habun is still unclear. According to Hull (2004a) it may be an archaic Tetum dialect, which was heavily influenced by both the Idaté-Isní-Lakalei dialect chain and the Karui-Waimaha-Midiki-Naueti dialect chain. The latter chain forms the eastern subgroup within Extra-Ramelaic. The northern subgroup, finally, contains the languages spoken on Wetar Island in the Southwest Maluku Regency, the languages spoken on Atauro Island (East Timor), and Galoli, which is spoken in the Manatuto District.

Hull (1998) points out that the Ramelaic and Extra-Ramelaic groups are very similar. The Ramelaic languages all feature the retention of PAn *p as [p] and the shift of PAn *ŋ to [g]. The Extra-Ramelaic languages on the other hand tend to merge PAn *n ,*ñ and *ŋ into [n], although in Roti and the Atauro dialects, PAn *ŋ survived as a velar nasal in certain phonological contexts. The PAn labial plosives *b and *p have all become either labial or glottal fricatives (f, v/w or h) in Extra-Ramelaic. Sound shifts apparently shared by all Timoric languages are the velarisation of PAn schwa *e into back mid-open vowel [ɔ] and the 'denasalisation' (Hull 1998:117) of prenasalised PAn consonants. As a result, a clear-cut classification of Timoric languages is often hampered by the fact that many of the attested sound changes occur in both groups rather than being exclusive to one group only. For example, PAn *j, qualified by Hull (1998) as a palatal approximant, changed into [l] in the Ramelaic group and in [r] in most languages of the Extra-Ramelaic languages. However, its reflex is also a lateral in the East group of Extra-Ramelaic and in Galoli, which belongs to the North subgroup. Because these languages border directly on the regions where Ramelaic languages are spoken (Mambai and Idaté), this lateral could be attributed to Ramelaic influence. The change from PAn *j to [l] also occurred in Helong (West Subgroup) in the extreme western part of the island, because there all liquids merged in a lateral. Similarly, PAn *k became a glottal stop in most Timoric languages, except for Helong (western subgroup) and the Karui-Waimaha-Midiki-Naueti dialect chain (or East group) of Extra-Ramelaic, and Tukudede in the West group of Ramelaic. The best example, however, is PAn *R, which is generally lost in the Timoric languages. In the Mambai dialects (Central Ramelaic) and the dialects of Atauro (Northern Extra-Ramelaic), however, it is sometimes reflected as a trill [r]. In the Kairui-Waimaha-Midiki-Naueti dialect chain (Eastern Extra-Ramelaic), it survived as a post-aspiration on the initial plosive consonant within the same word, as in Waimaha *thelu* 'egg' < **(qa)teluR* (Hull 2002:7, fn.5).

---

5 Referred to as 'Group B' or Austromunic in Hull (1998) and as 'Highland languages' in Engelenhoven (2006).

6 Referred to as 'Group A' or Austrofabronic in Hull (1998) and as 'Lowland languages' in Engelenhoven (2006).

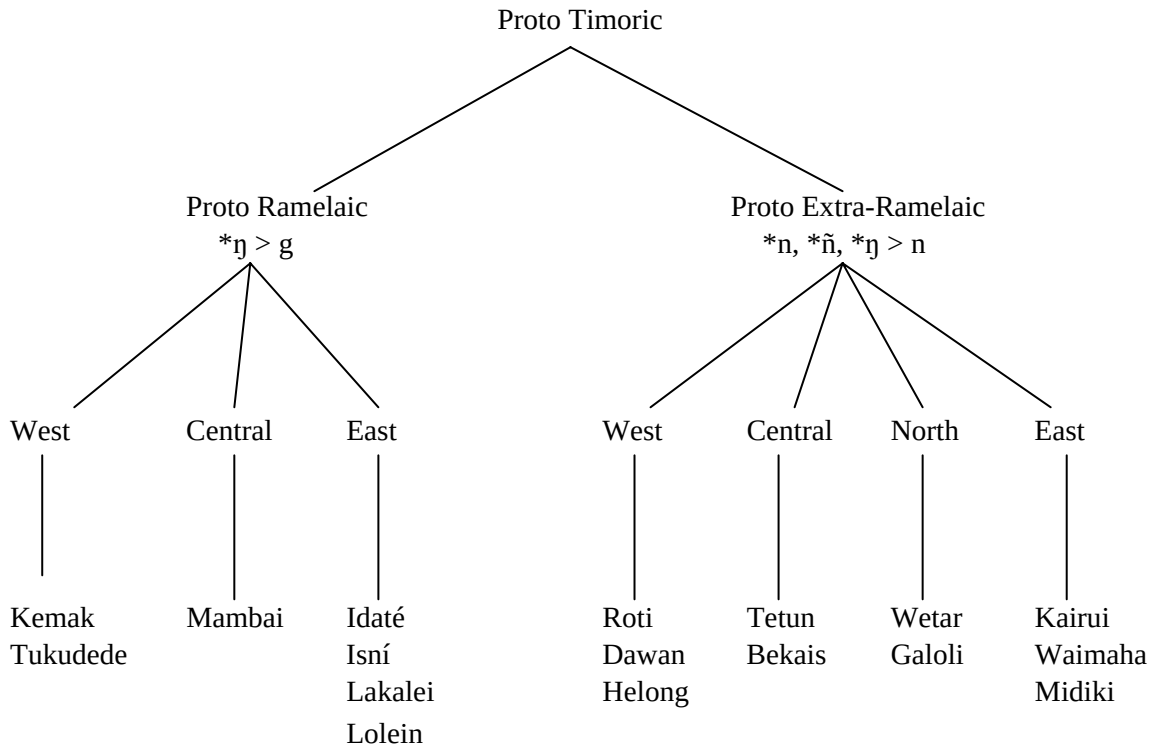


Figure 1: Proto Timoric (Hull 1998)

Elsewhere, Hull (2001) explains that the languages of Timor constitute a *Sprachbund* in which languages from both the Austronesian and non-Austronesian stocks share the same grammatical and phonological make-up. This is an additional problem for the Proto Timoric hypothesis. Hull (2004b) observed that many sound changes initially attested for his reconstruction of Proto Timoric also occurred in the non-Austronesian languages of Timor. For example, the shift of a PAn labial plosive to a labial or glottal fricative, one of the features setting off the Extra-Ramelaic languages from the Ramelaic ones, is also attested in the non-Austronesian languages. Similarly, the non-Austronesian languages display the same frequent change from /s/ to /h/, and the merger of velar plosives in a glottal stop (as observed in most Timoric languages), and a similar simplification of prenasalised stops as had happened in all daughter languages of Proto Timoric. Therefore, many sound shifts observed in Hull's (1998) paper may in fact be areal changes which are of little value for the classification of the languages involved.

Engelenhoven (2004) proposes a reconstruction for the Austronesian languages that are spoken on the island string expanding from Roma off the North-East coast of East Timor in the West, to Tepa village on Babar Island further East in Southwest Maluku Regency. The languages of Kisar[7] and Roma constitute a single Kisaric subgroup. The remaining languages together form the Luangic subgroup, of which Leti is one branch, and the other isolects spoken on Moa, Lakor, Luang, Wetan and in Tepa together form another branch. Meher, spoken on Kisar island, underwent peculiar sound changes that clearly distinguishes it from all other Luangic-Kisaric languages. The main sound change grouping it together with Roma is the loss of PAn *w which is still maintained in one way or another in the Luangic languages, as exemplified by PAn **waSiR* 'water': Rom *ori*, Meh *oiri*, Let *uèra*,

[7] Oirata on Kisar Island is excluded, as it is a non-Austronesian language closely related to Fataluku.

Moa *gèra*, Lng, Wet. *yera*. What distinguishes the Kisaric-Luangic group from the neighbouring languages to the East (in SE Maluku) and from the Timoric languages to the West, is the merger of PMP *z and *t in PLK *t.

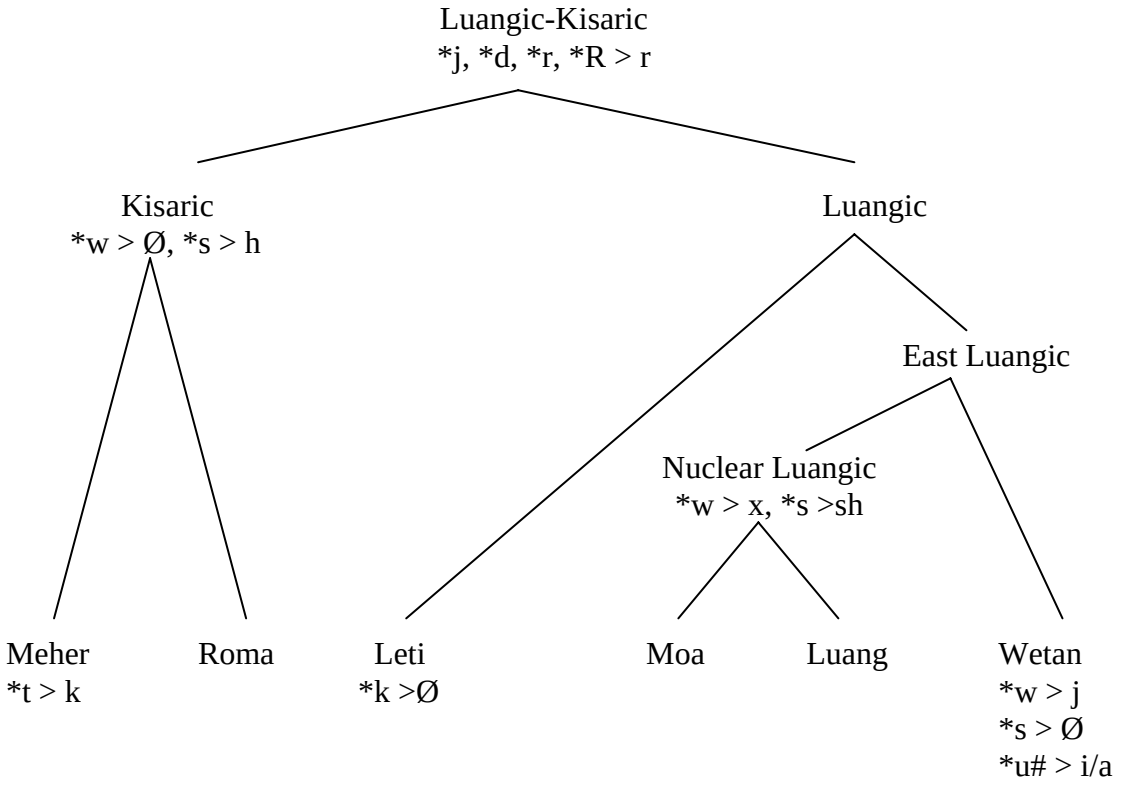


Figure 2: Proto Luangic-Kisaric (Engelenhoven 2004)

In Hull's (1998) proposal, the Austronesian languages in the Indonesian regencies of Southwest and Southeast Maluku east of East Timor belong to a subgroup initially labelled Arafuric, but later renamed Nautonic (Hull and Branco 2002/03). This group is on a par with the Bima-Sumba-Flores group (labelled Florinic in Hull (1998)) and the Timoric group. Formulated otherwise, Hull relates the languages of Timor and the regencies east of Timor through a putative ancestor to a higher node (labelled Santalic). In Hull's and Branco's (2002/03) analysis, Makuva was still categorised as a Malukan (read: non-Timorese) language imported from nearby Kisar. In his (2004a) report, Hull acknowledges that Makuva belongs to the Extra-Ramelaic group and as such is a Timorese language. He now suggests that Makuva, instead of being imported into Timor, is in fact the ancestor of Meher, and he argues that eastern Timor '[w]as the springboard for the "austronesianisation" of … the South Moluccas minus Wetar and Aru' (Hull 2004a:§8).

3 Makuva sound changes

Before discussing the sound changes of Makuva, an overview of its phoneme system is given here. Its vowel system is fairly consistent with the inventories found in the surrounding languages: /i, u, e, o, a/. There are no long vowels (unlike Fataluku, where they occur in open monosyllabic lexical morphemes, Stoel 2006). Makuva shares with the East dialect of Fataluku a variable articulation of the mid vowels /e/ [ɛ, e] and /o/ [ɔ, o].

Elicitation sessions with informants in Mehara suggest that final high back vowels are lowered to mid position, as in: /haku/ [hako] ‘stone’.

Makuva has the following consonants:

Table 1: Makuva consonant chart

	Labial	Alveo-Dental	Palatal	Velar	Glottal
Occlusive vcl	p	t	c	k	ʔ
Occlusive vcd	(b)	d	j		
Nasal	m	n			
Fricative vcl	f	s			h
Fricative vcd	β	z	ʒ		
Liquid		l, r			
Glide	w		y		

Makuva’s consonant inventory equals that of other languages in the region. However, unlike Capell’s (1972) suggestion that it lacks voiced plosives like its non-Austronesian neighbour Fataluku, ongoing research has shown that it does feature an alveolar plosive [d], albeit confined to a few words only. The voiced bilabial plosive mentioned by both Capell (1972) and Hull and Branco (2002/03) was attested only once in the word [ʔajblɔhɛ], rendered as *aiboleva* ‘wood’ in Capell, and as *ai bloheva* ‘wood, pillow’ in Hull and Branco.

Unlike what happened in Fataluku (the first language in the sub district), the glottal stop has not been completely effaced but has survived in Makuva on morpheme boundaries between vowels, for example *lo-u* [loʔu] (leg-1sgP) ‘my leg’ and *na-alrai* [naʔalraj] (3sg-read) ‘he reads’.

A sound that Makuva shares with Fataluku, but which is absent in any of the surrounding languages, is the voiceless palatal occlusive [c]. This consonant, however, has been attested only in clear Fataluku loans, for example *ci lafai* ‘gun’ (lit. thunder big).

Another similarity with Fataluku is the free variation between the palatal glide, the voiced occlusive, the voiced palatal and the voiced dental sibilant: [y~j~ʒ~z]. In Makuva, this is mainly attested in initial position, for example [yɛnɛva ~ jɛnɛva ~ ʒɛnɛva~ zɛnɛva] ‘rice’. In Fataluku, [y~ʒ~z] are free variants of intervocalic /i/, and [j~ʒ~z] are free variants of initial /j/.

East Fataluku has a preference for a bilabial glide [w], where other Fataluku dialects have a voiced bilabial fricative [β]. In Hull and Branco’s Makuva data (2002–03), this is consistently written as <v>. However, our fieldwork revealed that initial <v> is preferably realised as a bilabial fricative [β].

An exclusive feature in Makuva phonology is the gemination of intervocalic consonants. Furthermore, in intervocalic clusters consisting of occlusives and liquids, the occlusives may also be geminated, e.g. *hune ~ hunne* ‘moon’; *hake ~ hakke* ‘pig’; *tetre ~ tettre* ‘mat’. Further research may reveal the exact conditioning of this phenomenon.

Hull and Branco (2002/03) point out that the Austronesian character of the Makuva lexicon is blurred by the curious sound changes it has undergone, which are the reason why Capell (1972) did not recognise it as an Austronesian language. These sound changes are displayed below and are partly copied from Hull and Branco (2002/03). Note that no reflexes of PMP *c, *r and *g are given due to insufficient data.

PMP *p	> h /#_	**pandan* ‘Pandanus tree’ > *hene, *pusej* ‘navel’ > *+huter > hutre;*
	> Ø / V_V	**nipay* ‘snake’ > *ne, *kapuR* ‘lime’ > *+v-aur > v-aro;*
	> Ø/ _#	**ma-qudip* ‘live’ > *mori, *malip* ‘laugh’> *mali.*
PMP *b	> h / #_	**batu* ‘stone’ > *haku, *beli* ‘buy’ > *heli;*
	> Ø / V_V	**tebuh* ‘sugarcane’ > *kiu;*
	> Ø/ _#	**tutub* ‘close’ > *kuku* ‘shut’.
PMP *w	> w /#_	**wahiR* ‘water’ > *wera, *wakaR* ‘root’> *wara*
	> Ø / V_V	**buwaq* ‘fruit’ > *hue, *lawa* ‘long of time’ > *leu*
	> Ø / _#	**qalejaw* ‘sun’ > *lera, laRiw* ‘run’ > *lari, lori.*
PMP *t	> k	**telu* ‘three’ > *-kelu, *ma-putiq* ‘vomit’ > *muke, *kulit* ‘skin’ > *+ulit > ulke.*
PMP *d	> d, t / #_	**daRaq* ‘blood’ > *dar-ve, *deŋeR* ‘hear’ > *+dener > ++denre > +++dede > dete, tete;*
	> r / V_V	**si-ida* ‘3pl’ > *tira*;
	> r / _#	**likud* ‘back’ > *+liur > liru.*
PMP *s	> t	**susu* ‘breast’ > *tutu, *sakay* ‘go up’ > *tai; *asu* ‘dog’ > *ato, *tasik* ‘sea’ > *kahi; ma-panas* ‘warm > *+manat > manta,* Tetum *malus* ‘betel nut’ > *+malut > maltu.*
PMP *l	> l / #_	**lakaw* ‘walk’ > *la, *laRiw* ‘run’ > *lari, lori; *laquy* ‘bulb’ (Hull 2000) > *lau* ‘onion’;
	> l / V_V	**baliw* ‘return’ > *hali, *beli* ‘buy’ > *heli, *qulu* ‘head’ > *ulu;*
	> Ø / _VC#	**teluR* ‘egg’ > *+teuR > kiru, *bulan* ‘moon’ > *+huan > hune.*
PMP *z	> j	**jalan > +jaan > ++jana > jane, *zaqat* ‘bad’ > *jake; *quzan > +juqan > ++juan > jone.*
PMP *ñ	> n	**ñiuR* ‘coconut’> *nur-ve, *ñuka* ‘wound’ > *nua; *wañi* ‘bee’ > *van-, *miñak* ‘fat’ > *mine.*
PMP *k	> Ø	**kutu* ‘flee’ > *uku, *kahiw* ‘wood’ > *ai, *kau* ‘2sg’ > *o*; **ikan* ‘fish’ > *+ian > ++iene > yene, *takut* ‘fear’ > *+kauk > -kaku; *anak* ‘child’ > *ane-.*
PMP *ŋ	> n	**ŋajan* ‘name’ > *nade-, * ŋisi* ‘tooth’ > *nit-ve, *haŋiŋ*’wind’ > *+anin > ++anni > ane, *taliŋa ‘ear’> linu.*
PMP *j	> r	**maja* ‘dry’ > *mare* ‘yellow’, **lalej* ‘fly’ > *lare, *pusej* ‘navel’ > *+huter > hutre.*

PMP *q	> Ø	**qenay* ‘sand’ > *ini, *qatey* ‘liver’ > *ake, *qendey* ‘carry’ > *ati; *maqasuh* ‘smoke’ > *masu, ma-dequ* ‘thirsty’ > *mrou, *taqih* ‘dung’ > *ke; *ma-iRaq* ‘red’ > *meru*.
PMP *R	> r	**beRat* ‘heavy’ > $^{+}$*herak* > *herka, *quRat* ‘vein’ > $^{+}$*urak* > *urke; *damaR* ‘resin’ > $^{+}$*remar* > *remra* ‘morning’, **ikuR* ‘tail’ > $^{+}$*iur* > *iru, *teluR* ‘egg’ > $^{+}$*keur* > *kiru.*
PMP *h	> Ø	**haŋin ‘wind’* > $^{+}$*anne* > *ane, *wahiR* ‘water’ > $^{+}$*wair* > (metathesis) $^{++}$*weer* > *were* or (paragoge) $^{++}$*wer* > *were, *tebuh* ‘sugarcane’ > *keu.*

In summary, the following changes have taken place. The PMP voiceless velar and uvular stops *k and *q and the glottal fricative *h have disappeared in Makuva. All PMP nasals other than *m merged in the dental nasal [n]. In intervocalic and final position, the voiced alveolar and palato-velar stops *d and *j merged with the uvular fricative *R in the alveolar trill [r]. PMP initial *d is transcribed as <d> in Hull and Branco’s (2002/03) data. However, fieldwork revealed that in present day Makuva it is pronounced as a voiced alveolar [d] or a voiceless dental stop [t], with a preference for the latter pronunciation. For the time being we hypothesise that this devoicing is related to a similar phenomenon in the Central and East Fataluku dialects where voiced stops are no longer found.

The PMP lateral *l as onset was retained in open syllables, but was lost in closed final syllables. Also, the palatal voiced stop *z was retained as /j/ in initial position. In the only example where it is found in intervocalic position, it metathesised with initial *q: **quzan* > $^{+}$*zuqan* > *jone* ‘rain’.

In the labial series, both PMP *p and *b merged in the glottal fricative [h] in initial position and were lost elsewhere. Similarly, PMP *w was only retained in initial position.

Two specifically Makuva sound changes occurred in the dental series. PMP *t became a voiceless velar stop [k] and PMP *s became a voiceless dental stop [t].

Makuva reflexes of PMP prenasalised consonants are found in only three lexemes containing PMP *-nd-. *Hene* ‘pandanus tree’ and *marine* ‘cold’ suggest that the intervocalic nasal-consonant cluster either merged in a dental nasal, or, alternatively, in a voiced alveolar stop.

PMP **pandan* ‘pandanus tree’ > $^{+}$*henen* > *hene* , or:
PMP **pandan* ‘pandanus tree’ > $^{+}$*heden* > $^{++}$*hedene* > $^{+++}$*hedne* > *hene*

PMP **madindiŋ* ‘cold’ > $^{+}$*marinin* > *marine*, or:
PMP **madindiŋ* ‘cold’ > $^{+}$*maridin* > $^{++}$*maridini* > $^{+++}$*maridni* > *marine*.

A scenario in which the intervocalic cluster becomes a single nasal and subsequently deletes the final nasal would be more attractive and less complex than the ‘pseudo-metathesis’ scenario in Luangic proposed by Mills and Grima (1979) in which the intervocalic cluster becomes [d], an echo vowel is added, the resulting penultimate syllable’s nucleus is deleted, and finally the $^{+}$CN cluster becomes a single nasal. However, the following example, PMP **qendey* ‘carry’ > *adi, ati* supports the latter view in that first the consonant cluster is denasalised and then optionally devoiced. This is also borne out by the data in the comparison with related languages in the next paragraph.

Three other etyma reflecting historical intervocalic homorganic nasal + consonant clusters are PCEMP **mansar* 'opossum' > *mate*, PMP **empuh* 'grandparent' > *upu* 'grandchild',[8] and PMP **punti* 'banana' > *utu-ke*. It is very likely that these three items are loans from Southwest Maluku where they occur in all Luangic-Kisaric languages as *mada* 'opossum', *upu* 'grandparent/grandchild' and *udi* 'banana', respectively. (The change from *i# changed to u# in *utu-ke* 'banana' remains unexplained).

There are two clear examples of initial p reflecting an initial nasal-consonant cluster in an older stage of Makuva. This cluster ultimately derives from a prefix **ma-* followed by an initial labial consonant in PMP: **ma-peñúq* 'full' > *+(m)peñúq* > *penu*; **ma-besuR* > *+(m)besur* > *+petur* > *++peturu* > *petru* 'sated'.

The PMP penultimate vowels are neatly retained in most instances. The reflexes of ultimate syllable vowels, however, are less regular. It was pointed out above that in individual speech final back high vowels may be lowered and low vowels may be raised, so that both end up as mid vowels. It seems that when PMP high front vowels occurred in closed final syllables, they became lowered to mid-front vowels in Makuva. The high back vowel in this phonotactic context have been lowered to a back mid vowel, especially when encliticised by the nominal particle *=va*. Both realisations were accepted by informants. As such, [ɔ, o] functions as an allophone of /u/ in final position. PMP *i, however, split in two separate phonemes /i/ and /e/ which only have [i] and [ɛ, e] as their respective realisations in final position. Final PMP *a is either retained as /a/ or /u/.[9]

PMP *i	> i / #(C)_	**ina* 'mother' > *in-*, **isa* 'one' > *ite-tla*, **ikuR* 'tail' > *+iur* > *iru;* **ñiRu* 'winnowing tray' > *niru.*
	> e / _C#	**ma-diŋdiŋ* 'cold' > *+rinin* > *++rinne* > *rine*, **qulit* 'skin' > *+ulik* > *ulke*, **wahiR* 'water' > *+wair* > *++weer* > *+++wer-a* > *were.*
PMP *u	> u	**qulu* 'head' > *ulu, susu* 'breast' > *tutu* **batu* 'stone' > *haku*, **asu* 'dog' > *atu*, **kapuR* 'lime' > *+w-aur* > *w-aru.*
PMP *e	> e	**qalejaw* 'sun' > *lera*, *lere*, **telu* 'three' > *kelu*, **ma-qitem* 'black' > *+mekem* > *mekme.*
PMP *a	> a	**qabu* 'ash' > *au* 'dust', **tasik* 'sea' > *kate*, **anak* 'child' > *ana, ane*.
	> a / C_#	**lima* 'five' > *-lima*, **duha* > *-rua* 'two', **(ma-)isa* > *meta* 'alone'
	> u / C_#	**mata* 'eye'> *maku, maka*, **taliŋa* 'ear' > *linu.*

[8] In Makuva, the concept 'grandparent' is preferably indicated by the Fataluku word *calu*, or, when distinguishing the grandparent's gender, by the Indonesian terms *kakek* 'grandfather' and *nenek* 'grandmother'. Luangic-Kisaric distinguishes by generation and gender only. As such, grandparents and grandchildren are referred to by the same term, because they are one generation beyond EGO.

[9] It is imaginable that a similar scenario as has been described for the final high back vowel was also applies to final low vowels in that they were raised to back high position when enclitised by *=va*, for example: **maiRa(q)* 'red' > *meru=va.* Nominal cases with /u#/ may be explained as having a petrified 1sg possessive suffix *-ʔu*, as in **ama* 'father' > *amu* (< am-u 'father-1sgPos'). Compare with **ina* 'mother', which does allow both /a#/ and /u#/ variants: *ina, inu* (< in-u 'mother-1sgPos').

Although Makuva final syllable vowels do not always clearly reflect the original PMP vowels, PMP final diphthongs seem easier to track. PMP *uy and *aw merged in a#, whereas *iw and *ay merged in i#. PMP *ey seems to have become *e#.[10]

PMP *iw#	> i#	**laRiw* 'run' > *lari*, **kahiw* 'wood'> *ai;*
PMP *uy#	> a#	**hapuy* 'fire' > *hapuy* > *+aa+ke* > *ake*, **babuy* 'pig' > *+haa+ ke* > *hake;*
PMP *ey#	> e#	**matey* 'dead' > *make*, **qatey* 'liver' > *ake;*
PMP *ay#	> i#	**qenay* 'sand' > *ini*, **sakay* 'climb' > *tai;*
PMP *aw#	> a#	**qalejaw* 'sun' > *lera, lere*, **betaw* 'female relative' > *ma-heka, ma-heke* 'female'.

Hull and Branco (2002/03) mention metathesis as a characteristic feature in Makuva. In final closed syllables inherited from PMP, the rhyme vowel and the coda consonant change place in order to meet the preferred open syllable structure in the region (Klamer 2002:368).

						Makuva
areca nut	Tetum:	*malus*	>	*+malut*	>	*maltu*
salt	PMP:	**ma-qasin*	>	*+matin*	>	*matne* 'meat'
west	PMP:	**baRat*	>	*+harak*	>	*harke* 'wind'[11]
skin	PMP:	**kulit*	>	*+ulik*	>	*ulke*
road	PMP:	**zalan*	>	*+jaan*	>	*jana, jane*
moon	PMP:	**bulan*	>	*+huan*	>	*huna, hune*
egg	PMP:	**teluR*	>	*+keur*	>	*kiru*

The reflex for 'skin' does not follow the expected change from PMP *VlVC# to +VØVC# and then to VCV#. It is possible that this lexeme is borrowed from Meher on Kisar Island, where it is an inalienable noun and has an obligatory possessive suffix: *ulki-n* 'skin-3POS'.

4 Locating Makuva in Proto Timoric

In Hull (1998) and Hull and Branco (2002/03), Makuva was classified as a South Malukan language, which was introduced in the Tutuala sub district from Kisar. However, in his (2004a) article, Hull suggests that Makuva may well be an indigenous Timorese language because of 'its strongly Fabronic (i.e. Extra-Ramelaic,) character and its archaic nature'. In a personal communication, Geoffrey Hull suggested a close link between Makuva and the languages of the Extra-Ramelaic East group. As Hull (1998) does not provide an overview of sound changes that are exclusive to the East group, we tried to extract this from the available data in this publication and checked the result against the information in Saunders (2002/03) and Hull (2001) for Naueti, and Belo et al. (2005) for Waimaha. The following table summarises the consonant shifts in the East group and Luangic-Kisaric with reference to the PMP reflexes attested in Makuva. Because of the above-mentioned variation in Makuva vowel pronunciation and lack of space, the present discussion will be confined to consonants only.

---

[10] However, *ey became a# (Hull 1998) in the Extra-Ramelaic East group and i# in Luangic-Kisaric. Elicitation tests with informants nevertheless suggest that PMP *ey > e# is a separate phoneme rather than an allophone of /a/ or /i/.

[11] The expected allomorph [harka] was not attested during elicitations.

Table 2: Reflexes of PMP consonants in Proto Timoric (PT), Proto Extra-Ramelaic (PER), Proto East group (PEG), Proto Luangic-Kisaric (PLK), Naueti, Waimaha, Makuva, Meher and Leti[12]

PMP	PT	PER	PEG	Naueti	Waimaha	Makuva	PLK	Meher	Leti
*n	**n								
*ñ	**n	***n	$^{+}$n	n	n	n	$^{+}$n	n	n
*ŋ	** ŋ								
*m	**m	***m	$^{+}$m	m	m	m	$^{+}$m	m	m
*t	**t	***t	$^{+}$t	t	t	k	$^{+}$t	k	t
*p	**p	***f	$^{+}$h	h	h	h	Ø	Ø	Ø
*b	**β	*** β	$^{+}$ β	w-, -Ø-, -Ø	w-, -h-, -Ø	h	$^{+}$ β	β	β
*w	**w	***w	$^{+}$w	w	w	w	$^{+}$w	Ø	u-
*z	**z	***z	$^{+}$z	d	d	j	$^{+}$t	k	t
*d	**d	***d	$^{+}$d	r	r	d-, -r-, -r	$^{+}$r	r	r
*j	**Ø	***Ø	$^{+}$r	-r-, -Ø	-r-, -Ø	r	$^{+}$r	r	r
*l	**l	***l	$^{+}$l	l	l	l, Ø	$^{+}$l	l	l
*s	**s	***s	$^{+}$s	s	s	t	$^{+}$s	h	s
*k	**k	***k	$^{+}$k	k	k	Ø	ʔ	ʔ	Ø
*R	**R	***R	$^{+}$R	Ø, h	Ø, h	r	$^{+}$r	r	r
*q	**̃*	***ʔ	$^{+}$ʔ	ʔ	ʔ	Ø	Ø	Ø	Ø
*h	**h	***ʔ	$^{+}$ ʔ	ʔ	ʔ	Ø	Ø	Ø	Ø

The main feature distinguishing the East group languages from Makuva and Luangic-Kisaric is the retention of PMP *q, from which evolved a set of ejective consonants in Waimaha (Hajek and Bowden 2002; Hull 2002) and Naueti (Saunders 2002/03). Hull (1998:118) reports a '[g]eneral loss of primary [h] and secondary [h] (< *S, *H) but their occlusivisation (> [ʔ]) in Wakatobi,[13] and partly Kawaimina.[14]' Unfortunately, he does not elaborate on this observation but only provides some random examples.

PMP *q **qalejaw* 'sun' > Nau *ʔara,* Wai $^{+}$ *ʔara* > *lʔara*[15], Mak, Let *lera*, Meh *lere;*
**maqasuh* 'smoke' > Nau *ʔhasu*, Wai *nʔasu*, Mak *matu*, Meh *mahu-*.

PAn *H **CunuH* 'bake' > Nau, Wai *thunu*, Meh *kunu*, Let *tunu*, **sikuH* 'elbow' > Nau *siku*, Wai *sʔiku*, Mak *tiu*, Let *sik-siku* (expected **siu*).

---

12 Reflexes of PMP *r, *c and *g, are not be incorporated in this table due to lack of data.

13 An acronym used by Geoffrey Hull for the Tukang Besi language spoken on the Tukang Besi Islands off the coast of Southeast Sulawesi. It refers more specifically to the dialects spoken on the islands of Wanci, Kaledupa, Tomea and Binongko.

14 An acronym used by Geoffrey Hull to refer to the Karui, Waimaha, Midiki and Naueti dialects.

15 Hull's (2002) description of Waimaha suggests that the pharyngeal approximant [ʕ] and the glottal stop [ʔ] are allophones of the glottal stop phoneme: the former occurs after liquids and nasals, and the latter after stops and the voiceless sibilant.

The East group, Makuva and the Luangic-Kisaric languages all share the merger of PMP *j and *d into an alveolar trill in medial position. Whereas this sound shift also took place in initial position in the East group and Luangic-Kisaric languages, Makuva retained *d as [d] or [t]. In the East group languages, PMP *d was generally lost in final position. In Luangic-Kisaric and Makuva final *d could be retained through a metathesis from *VC# to CV# or, in the case of Meher, through the addition of an echo vowel.

PMP *d **daRaq* 'blood' > Nau,Wai *raa,* Mak *dar-ve,* Meh, Let *rara;*
**si-ida* '3pl' > Nau *sira,* Wai *sire*, Mak *tira,* Meh *hi,* Let *ira;*
**tuhud* 'knee' (Hull 1998:121) > Nau, Wai *tuu,* Mak, Meh *kuru,* Let *turu.*

PMP *j **maja* 'dry' > Nau *marén,* Mak *mare* ' yellow', Meh, Let *mar-mara* 'yellow';
**lalej* 'fly' > Nau, Wai *lale,* Mak $^{++}$*laler* > $^{+}$*laer* > *lare,* (PLK **lalar* > $^{++}$*larar* > Meh *larra,* (Proto Luangic $^{+}$*laran* >) Let *llaran.*[16]

In Makuva and Luangic-Kisaric, PMP *R also shifts to an alveolar trill. In the East group, however, final *R survives through metathesis as a post aspiration on the onset consonant of penultimate syllable. In other positions, it was lost.

PMP *R **uRat* 'vein' > Nau *uán,* Wai *uo,* Mak, Meh $^{+}$*urak* > *urke,* Let *urat;*
**qateluR* ' egg' > Nau, Wai $^{+}$*teluh* > *thelu,* Mak $^{+}$*keur* > *kiru,* (PLK **telur* > $^{+}$*terur* >) Meh *kerru,* (Proto Luangic $^{++}$*terun* >) Let *terun.*

The main distinction between Luangic-Kisaric on the one hand and the East group languages and Makuva on the other, is that Luangic-Kisaric has merged PMP *t and *z. In the East group, PMP *t was retained in initial and medial positions and lost in final position. In Makuva it shifted to a voiceless velar plosive. In final position it was preserved through metathesis with the preceding vowel. PMP *z became an alveolar plosive in the East group. Makuva stands out as the only language that retained PMP *z as a palatal occlusive.

PMP *t **telu* 'three' > Nau, Wai, Let *-telu,* Mak, Meh *-kelu*;
**batu* 'stone' > Nau, Wai *watu,* Mak *haku,* Meh *vaku*, Let *vatu;*
**takut* 'afraid' > Nau, Wai *thaku,* Mak $^{+}$*kauk* > *kaku,* Meh *kaʔuku,* Let *taut;*

PMP *z **zalan* 'road' > Nau, Wai *dala,* Mak $^{+}$*jaan* > *jane,* Meh $^{++}$ *talan* > $^{+}$*talna* > *kalla,* Let *talan*;
**quzan* 'rain' > Nau *uda,* Wai *udo*, Mak $^{++}$*ujan* > $^{+}$*juan* > *jone,* Meh *okono,* Let *utan.*

Sound shifts that are exclusive to Makuva are PMP *s > [t] and the merger of PMP *p and *b in [h] in initial position. Hull (2000) suggests that the reflex of PMP *b in Proto Timoric was unstable and therefore resulted in a glottal fricative in the Northern Extra-Ramelaic group and in a labial approximant in the Eastern Extra-Ramelaic group. In Luangic-Kisaric PMP *p was lost, whereas PMP *b became a voiced labial fricative. In final position, PMP *p was lost in all languages. It was shown above that PMP *b# was lost in Makuva. Engelenhoven (1995) reports one instance in Leti where *b# was retained as a voiced labial fricative (Let *tutuv* 'close' < PMP **tutub*); it was lost in Kisaric. No instances were found in Hull (1998), Saunders (2002/03) or Belo et al. (2005) for the East group languages in this context.

---

[16] Leti has a complex morphological and syntactic process of metathesis in order to enable phrases to end in a vowel. For the sake of convenience, Leti reflexes are given in their representation with final consonants if they are derived from PMP forms with final consonants (Let *VC# < PMP *VC#). For an elaboration on metathesis in Leti, I refer to Engelenhoven (2004:89–92).

PMP *s **susu* 'breast' > Nau *susún*,Wai *susu*, Mak *tutu* 'milk', Meh *huhun*, Let *susu;*
**asu* 'dog' (> Pr.EG ***z-asu* >) Nau, Wai *dasu*, Mak *atu*, Meh *ahu*, Let *asu;*
**ma-panas* 'hot' > Mak $^{+}$*manat* > *manta*, Meh *manha*, Let *panas.*

PMP *p **pandan* 'pandanus tree' > Nau *hede* 'screwpine', Wai *hida*, *Mak* $^{+}$*henen* > *hene*, Meh $^{++}$*edan* > $^{+}$*edna* > *esne*, Let *edan;*
**lapaR* 'hungry' > Wai *laha*, Mak $^{+}$*m-laar* > *m-lara* or $^{+}$*m-lar-a* > *m-lara*, Meh *m-lara*, Let *p-lara;*
**ma-qudip* > Nau *morín* 'new', Wai *mhori* 'new', Mak, Meh, Let *-mori.*

PMP *b **beniq* 'seed' > Nau *yinán* (y- unexplained), Wai *wine*, Mak *hin-*, Meh, Let *vini;*
**abu* 'ashes' > Nau *ou*, Wai *boo* (b- unexplained), Mak *ahu* 'dust', Meh, Let *avu.*

Makuva lost PMP *k, whereas East group languages retained it as *k* in initial and intervocalic position. In Proto Luangic-Kisaric it became a glottal stop, and in Leti it was subsequently lost. We hypothesise that in the ancestor directly preceding Makuva PMP *k changed to $^{+}$ʔ before it was effaced.

PMP *k **kita* '1plinc' > Nau *kita* '1plex', Wai *kite* '1plex', Mak, Meh *ika*, Lt *ita;*
**ihekan* 'fish' > Wai *ike*, Mak $^{+}$*ian* > $^{++}$*iene* > *yene*, Meh *iʔan*, Let *ian;*
**miñak* 'fat, grease' > Nau *mina*, Wai, Mak *mine*, Meh *mina-n*, Let $^{+}$*minaʔ* > $^{++}$*miʔan* > *mian.*

All languages kept PMP *m. From the table above we may conclude that Makuva joins the Luangic-Kisaric and East group languages in the merger of PMP *n, *ñ, *ŋ in n, which the East group languages inherited from Proto Extra-Ramelaic.

PMP *m **mata* 'eye' > Nau,Wai *mata*, Mak *maku*, Meh *maka-*, Let *mata;*
**ama* 'father' > Nau *ama-*, Wai *ama* 'male', Mak *amu*, Meh *ama-*, Let *ama;*
**ma-qitem* 'black' > Nau *itán*, Wai *ite*, Mak $^{+}$*mekem* > *mekme*, Meh *mekme*, Let *metam;*

PMP *n **nunuk* 'banyan' > Nau, Wai, Mak, Meh *nunu*, Let *nuun;*
**manuk* 'bird' > Nau, Wai, Meh *manu*, Mak *manuo*, Let *maun*;
**bulan* 'moon' > Nau *wulan*, Wai *wulo*, Mak $^{+}$*huan* > *hune*, Meh $^{+}$*volan* > *vollo*, Let *vulan;*

PMP *ñ **ñiuR* 'coconut' > Nau *nhee*, Wai *nee*, Mak *nur-ve*, Meh, Let *nura;*
**wañi* 'bee' > Nau, Wai *wani*, Mak *wan-*, Let *uani;*

PMP *ŋ **ŋisi* 'tooth' > Nau, Wai *nihi*, Mak *nit-ve*, Meh *niha-*, Let *nisa;*
**deŋeR* 'hear' > Nau, Wai *dene*, Mak $^{+}$*dener* > $^{++}$*denre* > $^{+++}$*dede* > *dete, tete*, Meh $^{+}$*dener* > $^{++}$*deren* > *derne.*

PMP *#w appears to have a reflex in all languages. Hull (1998:117) observes that in the East group languages /w/ is realised as a bilabial fricative [β]. Hull's (2002) description suggests that in the East group languages, PMP *b and *w merged in initial position, but not in medial position where the reflex of *b is [h] in Waimaha.

PMP *w **wahiR* 'water' > Nau, Wai *wai*, Mak $^{+}$*wair* > *were*, Meh *oiri*, Let *uera.*

PMP *-w- **sawah* ‘snake’ > Nau, Md, Wai *sawa*
**sawa* ‘spouse’ > Meh *ha-*, Let *so-*
**lawa* ‘spider’ > Mak *lau-lau-*

PMP *-b- **tebuh* ‘sugarcane’ > Nau, Md *teu*, Wai *tehu*, Mak *keo-*, Meh *kevu*, Let *tevu*.

Engelenhoven (1995) explains the existence of /p/, /d/ and /k/ in Luangic-Kisaric as retentions of homorganic prenasalised consonants in PMP. Hull (2000:168) mentions a comparable development in PT, where PMP *mp merged with *mb in a voiced labial plosive [b], and *Nk merged with *Ng in a voiceless velar plosive [k], in the daughter languages.[17] PMP *nt and *nd were simply denasalised as [t] and [d], respectively, but in a few cases *nt became [d]. Above, it was explained that the Makuva data we have collected thus far provide little evidence to assess whether or not this phenomenon occurred in this language.

PMP *mp **ma-peñúq* ‘full’> ***mpeñúq* > Nau *benu*, Wai *beno*, Mak, Meh, Let *penu;*
**empuh* ‘owner’ > Nau *bu-* ‘grandparent’, Wai *buu* ‘grandparent’, Mak *upu* ‘grandchild’, Meh *opu* ‘grandparent/grandchild’, Leti *upu-*’grandparent/ grandchild’;

PMP *mb **ma-besuR* ‘sated’ > ***mbesuR* > Nau, Wai *besu,* Mak $^{+}$*petur* > *petru*, Meh *pehru*, Let *pesur.*

PMP *nt **punti* ‘banana’ > Nau, Wai *hudi,* Mak *utu* (loan from Meher, u# unexplained) Meh,Let *udi.*

PMP *nd **pandan* ‘pandanus tree’ > Nau *hede* ‘screwpine’, Wai *hida,* Mak $^{+}$*heden* >$^{++}$*hedene* > $^{+++}$*hedne* > *hene*, (PLK $^{+}$*edan* >) Meh $^{++}$*edene* > $^{+}$*edne* > *esne*, Let *edan.*

The reflexes of PMP **lalej* ‘fly’ and **qateluR* ‘egg’ above suggest there was a phonotactic constraint in PLK that disallowed final closed syllables with an alveolar lateral onset and an alveolar trill coda. The Meher and Roma data below suggest that the onset lateral assimilated to the trill coda before it applied VC# > CV# metathesis. None of the Luangic languages, however, have roots with last syllables beginning and ending with an alveolar liquid. Whereas the $^{+}$rVr# sequence was allowed in Kisaric, in Luangic $^{+}$r# was again dissimilated to a dental nasal n#.

Gloss	PMP			Mak	Meh	Rom	Let
‘pigeon’	**baluj*	> $^{+}$*βalur*	> a) $^{++}$*βaur-u* >	*aru*			
			> b) $^{++}$*βarur-u* >		*βarru*	*βarur*	*βarun*
‘breadfruit’	**kuluR*	> a) $^{+}$*ulur*	> a) $^{++}$*uul-u* >	*ulu*			
		> b) $^{+}$*urur*	> b) $^{++}$*urur-u* >		*urru*	*urur*	*urun*

The reflex of PMP *-/luR# > $^{++}$-/ul# instead of expected $^{++}$-/ur# in the Makuva reflex of ‘breadfruit’ is problematic and suggests that it must be a loan, probably from Tetum (which has a corresponding *kulu*), rather than from the less expansive East group which has the same reflex as Tetum. In this scenario, Tetum or East group #k would be lost first, resulting in the Makuva form *ulu*.

[17] Hull hypothesises that there was also a process of nasalisation of plain consonants to consonant-nasal clusters (*C > **NC) in PT. Lack of time and space prevents me from a further investigation into this claim.

5 Discussion

Based on the sound correspondences PMP *R > r, *b, *p > h, *ŋ > n and *t > k, Hull and Branco (2002/03) conclude that Makuva agrees more with Meher and Luang than with Timoric languages. They consider that *R > r is rarely attested in Timoric. However, Hull (1998:118) already reported that in the North Extra-Ramelaic dialects of Wetar Island, *R was retained as r, whereas in final position it survived through metathesis in the East group (Hull 1998:118, fn.44). This means that Proto Extra-Ramelaic must have had a reflex of this PMP sound. In fact, it is the merger of *R and *j that Makuva shares with Luangic-Kisaric, alhough in the latter, the merger also includes *d.

Table 2 in §4 shows that the loss of PMP *p in Luangic-Kisaric can be explained as a sequel to its spirantisation to *f in Proto Extra-Ramelaic and consecutively to $^{+}$h in Proto East group. It still reflects the parallel shift of *b > v, which in Kairui-Waimaha-Midiki-Naueti is confined to initial position. The merger of *b and *p is exclusive to Makuva and sets this language off against Luangic-Kisaric and against Kairui-Waimaha-Midiki-Naueti.

Makuva does share the merger of *n, *ñ and *ŋ to n with Luangic-Kisaric and Kairui-Waimaha-Midiki-Naueti. All three groups classify as Extra-Ramelaic, because this merger is exactly one of the three sound changes distinguishing Extra-Ramelaic from Ramelaic, where *ŋ is retained (Hull 1998:106). Consequently, it is not evidence for a closer link between Makuva and Luangic-Kisaric.

This leaves only the sound shift *t > k. Because it also occurs in Meher, Hull and Branco (2002/03:110) propose that Makuva is '[a] colonial language [that is] implanted in the eastern tip of Timor relatively lately'. However, Blust (2005) points out that this sound shift also occurred in the isolects of East Wetar and South Babar and nearby islands further to the East in Southwest Maluku. The isolects of East Wetar, which in Hull's (1998) Proto Timoric Hypothesis belong to the North group of Extra-Ramelaic, display k for *t only in final position. In South Babar and adjacent islands, the sound shift appears to be unconditioned, although in some isolects it may have changed further to, for example, [x] (see Hein Steinhauer's contribution in this volume). Blust (2005:381) convincingly argues that *t > k is a recent innovation in Meher, postdating the introduction of loanwords in that language.

The fact *t became *k* in Meher as well as in East Wetar and South Babar suggests a language contact scenario in which Makuva induced the sound shift in Meher on nearby Kisar Island, rather than the other way around, as suggested by Hull and Branco (2002/03).

This seems borne out by the PMP *s > t shift which distinguishes Makuva from the directly surrounding Austronesian languages. It does occur in South Babar that also features *t > k. Blust (2005) explains *s > t as a drag chain effect of the *t > k. This explains the Makuva reflex *utu* from PMP **punti* 'banana' as an original morpheme, because in South Babar *k > t also effected *nt > t. In Meher, *s shifted to h in Proto Kisaric, whereas *nt already shifted to d in PLK.

The reflexes of *z and *d confirm that Makuva, Luangic-Kisaric and Kairui-Waimaha-Midiki-Naueti all belong to the same East group, while the subsequent shifts distinguish them as individual languages. Hull and Branco (2002/03) overlooked the fact that Makuva is the only language in the region that retained *z as a palatal plosive. Its merger with *t is what distinguishes Luangic-Kisaric from the others. This merger may be due to a subsequent devoicing of d < *z as it is still reflected in Kairui-Waimaha-Midiki- Naueti. As such, we hypothesise that Luangic-Kisaric and Kairui-Waimaha-Midiki-Naueti are genetically more close, and that Makuva is an early off-shoot from the Proto East group, as displayed in

Figure 3. Like *z, *d is only retained in initial position in Makuva. In medial position it merged with *R and *j in r, just as has happened in Luangic-Kisaric in all positions, whereas *d merged with *j in all positions in Kairui-Waimaha-Midiki-Naueti.

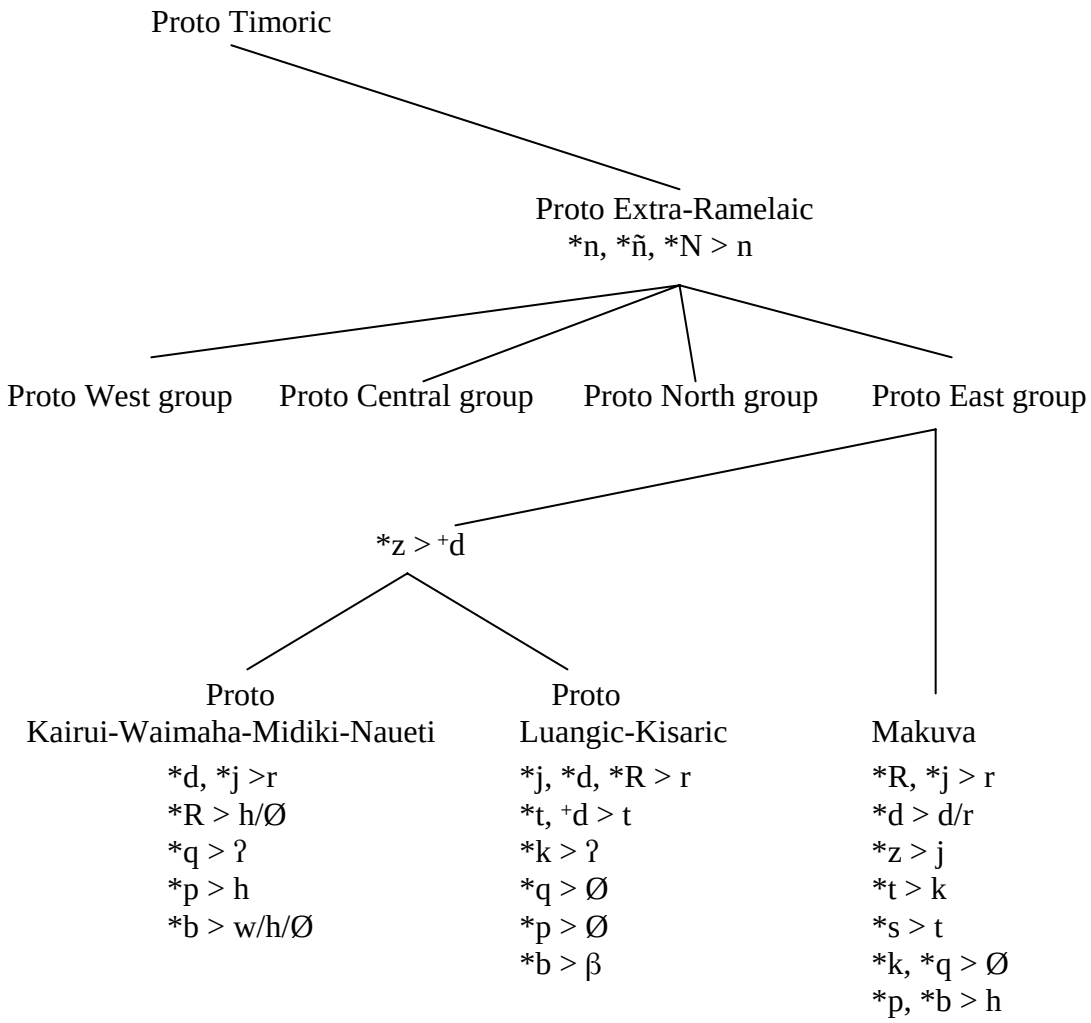


Figure 3: Makuva and other East group languages

References

Belo, Maurício, John Bowden, John Hajek and Nikolaus Himmelmann. 2005. *Glosáriu Waima´a Caisido*. Baucau: Publicações Matebian.

Blust, Robert. 2005. *t to k: an Austronesian sound change revisited. *Oceanic Linguistics* 43:365–410.

Capell, A. 1972. Portuguese Timor: two more non-Austronesian languages. In A. Capell, ed. *Linguistic papers, I: General, II: Indonesia and New Guinea*, 95–104. Oceania Linguistic Monographs 15, University of Sydney.

Engelenhoven, Aone van. 1995. Van Proto Malayo-Polynesisch naar Proto Luangisch-Kisarisch. In Connie Baak, Mary Bakker and Dick van der Meij, eds *Tales from a concave world, liber amicorum Bert Voorhoeve*, 246–264. Leiden: Leiden University, Dept. of Languages and Cultures of Southeast Asia and Oceania, Projects Division.

—— 2004. *Leti, a language of Southwest Maluku*. VKI 221. Leiden: KITLV Press.

—— 2006. 'Ita-nia nasaun oin-ida, ita-nia dalen sira oin-seluk', 'Our Nation is one, our languages are different'. Language policy in East Timor. In Paulo Castro Seixas and Aone van Engelenhoven, eds *Diversidade Cultural na Construção da Nação e do Estado em Timor-Leste*, 104–132. Porto: Publicações UFP.

Engelenhoven, Aone van and Justino Valentim Cailoru. 2006. The Makuva enigma: locating a hidden language in East Timor. Paper presented at the 2nd Conference on Austronesian Languages and Linguistics, June 2–3. Oxford: St. Catherine's College, [http://www.fataluku.com]

Grant, Anthony P. 2006. Some observations on Proto-Austronesian *t to *k*. *Oceanic Linguistics* 45(2):497–500.

Hajek, John. 1995. When is a non-Austronesian language really Austronesian? When it's Lóvaia. Unpublished paper at Melbourne University.

Hajek, John and John Bowden. 2002. A phonological oddity in the Austronesian area: ejectives in Waimoa. *Oceanic Linguistics* 41:222–224.

Hajek, John, Nikolaus Himmelmann and John Bowden. 2003. Lóvaia: an East Timorese language on the verge of extinction. *International Journal of the Sociology of Language* 160:155–167.

Himmelmann, Nikolaus and John Hajek. 2001. A report on the current sociolinguistic Situation in Lautém. *Studies in Languages and Cultures of East Timor* 4:88–97.

Hull, Geoffrey. 1998. The basic lexical affinities of Timor's Austronesian languages: a preliminary investigation. *Studies in Languages and Cultures of East Timor* 1:97–202.

—— 2000. Historical phonology of Tetum. *Studies in Languages and Cultures of East Timor* 3:158–212.

—— 2001. A morphological overview of the Timoric Sprachbund. *Studies in Languages and Cultures of East Timor* 4:98–205.

—— 2002. *Waimaha (Waima'a).* East Timor language profiles No.2. Dili: Istituto Nacional de Linguística/Universidade Nacional de Timor Lorosa'e.

—— 2002/03. Lia-Bekais nia Fatin iha Dalen-Lubun Timór. *Studies in Languages and Cultures of East Timor* 5:55–78.

—— 2004a. *The languages of East Timor. Some basic facts (revised 24.8.2004)*. retrievable at http://www.asianlang.mq.edu.au/INL/langs.ht

—— 2004b. The Papuan languages of Timor. *Studies in Languages and Cultures of East Timor* 6:23–99.

Hull, Geoffrey and Sabil José Branco. 2002/03. O enígma da língua lóvaia. *Studies in Languages and Cultures of East Timor* 5:107–134.

Klamer, Marian. 2002. Typical features of Austronesian languages in Central/Eastern Indonesia. *Oceanic Linguistics* 41:363–383.

Mills, Roger F. and John Grima. 1979. Historical developments in Lettinese. In Paz Buenaventura Naylor, ed. 273–283. *Austronesian studies. Papers from the second eastern conference on Austronesian languages*. Michigan Papers on South and Southeast Asia No.15. Ann Arbor: The University of Michigan.

Saunders, George. 2002/03. Comparative vocabulary of the Naueti dialect. *Studies in Languages and Cultures of East Timor* 5:79–106.

Steinhauer, Hein. 2009. The sounds of Southeast Babar. This volume.

Stoel, Ruben. 2006. *Fataluku as a tone language*. Paper presented at the 16th Annual Meeting of the Southeast Asian Linguistics Society, September 20–21. Jakarta: Atma Jaya University.

Part 5

Culture history and lexical reconstruction

26 *Words of Eastern Polynesia: is there lexical evidence for the origin of the East Polynesians?*

PAUL GERAGHTY

1 Eastern Polynesian[1]

The existence of an Eastern Polynesian subgroup comprising all Polynesian languages from Hawai‘i in the north to Rapanui (Easter Island) in the east and Aotearoa (New Zealand) in the south, and excluding all Polynesian languages from Pukapuka westward, has long been accepted. Hale (1846:117) was probably the first to moot such a subgroup, noting a number of features exclusive to the Eastern Polynesian languages that he studied, including the desiderative and reciprocal forms of the verb, the passive voice, and the plural of the possessive and demonstrative pronouns.[2]

Elbert (1953:154) and Haudricourt (1964:389) noted possible phonological innovations of Eastern Polynesian, then Pawley (1966:59–61) used innovations in grammatical morphemes (tense and other verbal markers, demonstratives, interrogatives etc) to lay the foundation for a more rigorous definition of the subgroup, while Green (1966:12–15) added lexicostatistical evidence. More recently, Marck (2000:131–132) has summarised the more compelling morphological innovations. Although Eastern Polynesian is a well-defined and generally accepted subgroup, there is a problem in reconstructing the lexicon, since one of its two first-order subgoups consists of only one language, Rapanui, spoken in a relatively impoverished natural environment and for which limited data is available. For

[1] It gives me great pleasure to dedicate this paper to my longtime friend and colleague Robert Blust. Since we first met at the University of Hawai‘i in the 1970s, I have found him an unfailing source of support, unstintingly generous with information and advice, and a model of dedication. Many thanks, Bob, and may you long continue to flourish and reconstruct! Many thanks also to Andrew Pawley, who suggested the topic, and gave much help and advice during the writing of this paper; and to Pila Wilson and Erik Pearthree for useful discussions.

[2] Hale (1846:118, 175) may also be said to have anticipated the modern subgrouping of Polynesian languages into Tongic and Nuclear Polynesian, in that he noted that Tongan (he was unfamiliar with Niuean) ‘differs strikingly, in several points, from the others, especially in the article, the pronouns, and the passive voice of the verb.’

Alexander Adelaar and Andrew Pawley, eds
Austronesian historical linguistics and culture history: a festschrift for Robert Blust, 445–460.
Pacific Linguistics, 2009.
© This edition vested with Pacific Linguistics.

want of a Rapanui reflex many items can only be reconstructed to Proto Central Eastern Polynesian (PCEP), the immediate ancestor of the other first-order group.

It has long been argued—or assumed—that East Polynesia was settled from Samoa (Hale 1846:119–125, 148), or somewhere in the region of Samoa and Tonga (Kirch 2000:231, 245). Apart from geographical proximity, a reason frequently cited has been that Savai'i, the largest island of Samoa, is believed to be the source of the place-name Hawaiki, commonly referred to in oral traditions as the Eastern Polynesian homeland.

An unexpected challenge to this assumption was made by Wilson (1985), who pointed out a number of shared innovations in the pronouns of East Polynesian and those of the North Central Outliers (Nukuria, Takuu, Nukumanu and Luangiua—situated north of Papua New Guinea and the Solomon Islands), and to a lesser extent Kapingamarangi and Nukuoro, in contrast to the pronominal system of Samoan, which shares no innovations exclusively with that of Eastern Polynesian.[3] Wilson suggested (1985:122–123) as a possible explanation that Eastern Polynesia was settled directly from the North Central Outliers, despite the distance of some 4000 miles, pointing out that in historic times it has been the inhabitants of small islands of scant resources, the Carolineans, the Tuamotuans, and the Tongans, who have been the most prone to sailing long distances.

Some fifteen years later, Marck (2000:xix) claimed to have found that 'Ellicean [North Central] outliers shared sporadic sound changes with Eastern Polynesian and Samoan that other Polynesian languages did not share …, a stunning bit of support for Wilson's (1985) suggestion of "Ellicean", composed of those same languages, on the basis of the pronoun prehistory.' However, the three sporadic sound changes referred to are, in my view, less than stunning, and in any case point to three distinct affiliations for Eastern Polynesian, only one of which matches that proposed by Wilson. In the first case, *fuʔaŋa 'whetstone' > *foʔaŋa, the change is shared with Kap, MFa, Nkr, and Sam.[4] In the second, *kiu 'plover, wading bird' > *kiwi, the change is shared with Oja, Nkr, Sik and Tak—that is, the North Central outliers, as proposed by Wilson. In the third, *mafo 'healed' > *mafu, the change is shared with Ren, Tik, WUv, Nkr, Sam and Tok. In particular, the inclusion of Samoan in two of these three proposed innovations is at odds with Wilson's proposal, which specifically excludes Samoan.

---

[3] I would query just one of Wilson's proposed shared innovations: 6. replacement of **ki-* initiated pronouns, e.g. **ki-taaua* 'we' inclusive dual, with forms in which **ki-* has been deleted, e.g. *taaua*. My objection is that both types of independent pronoun form can be reconstructed for PPn, given that unprefixed forms like *taaua* are also found in both Western and Eastern Fijian, e.g. Yasawa (WF) *tatou* 'first person inclusive paucal/plural' (Triffitt 2000:320), Dogotuki (EF) *mutou* 'second person paucal' (Geraghty 1983:199).

[4] Abbreviations and default sources: All Polynesian languages Biggs and Clark (n.d.) unless otherwise specified. All Fijian, PCP (Proto Central Pacific) and PEO (Proto Eastern Oceanic) from my own notes; EF unspecified communalect of Eastern Fijian, EFu East Futuna (Moyse-Faurie 1993), EP Eastern Polynesian, EUv East Uvea (Rensch 1984), Kap Kapingamarangi, Mae, Mao New Zealand Maori, MFa Mele-Fila (Clark 1998), Mqa Marquesan, Mia Mangaia, Mki Manahiki, Mva Mangareva, Niu Niue (Sperlich 1997), Nkr Nukuria, Nuk Nukuoro (Carroll and Soulik 1973), Oja Ontong Java (Luangiua), PCE Proto Central Eastern Polynesian, Pen Penrhyn, Pohnpei (Rehg and Sohl 1979), PEP Proto Eastern Polynesian, PNP Proto Nuclear Polynesian, PPn Proto Polynesian, PTa Proto Tahitic, Puk Pukapuka, Rar Rarotongan (Buse and Taringa 1995), Ren Rennell (Elbert 1975), Rot Rotuman (Inia et al. 1998), Rpn Rapanui (Salles and Pizarro 2001), Sam Samoan (Milner 1966), Sik Sikaiana, Tah Tahitian (Lemaître 1986), Tak Takū, Tik Tikopia (Firth 1985), Tok Tokelau, Ton Tongan (Churchward 1959), Tua Tuamotu (Stimson 1964), Tuv Tuvalu (Besnier 1981), WF unspecified communalect of Western Fijian, WFu West Futuna, WUv West Uvea (Hollyman 1987).

Most recently, Wilson (2008) has pointed to some lexical innovations that indicate a period of shared development between Eastern Polynesian and the Northern Outliers, including the following: PPn *kawiki ‘ghost crab’ > *kawiti (shared by EP with Kap, Tak, Sik); *watuke ‘pencil urchin’ > *fatuke (Tak); *fasi ‘split’ > *wasi (Tak, Oja, Sik); *taʔe ‘faeces’ > *tūtaʔe (Kap, Nuk, Tak); *fanaŋa ‘story’ > *wanaŋa (Kap); *ʔulupoko ‘skull’ > ‘head’ (Nuk); *laʔe ‘forehead’ > ‘headland’ (Oja).

2 Words and things

The purpose of this paper is to apply the *Wőrter und Sachen* (henceforth ‘words and things’) method of historical reconstruction to Proto Eastern Polynesian, to determine the geographical origin of the ancestral Eastern Polynesians, thus reinforcing or challenging Wilson’s hypothesis of a Northern Outlier source. I will focus on plant-names, looking for evidence as to whether the ancestral Eastern Polynesians lived on a high volcanic island, such as Samoa, in which case they may have tended to retain names of high island plants, or a low coral island or atoll, such as Pukapuka, Tokelau, Tuvalu, or any of the northern outliers, in which case they may have become unfamiliar with high island plants, and have had to reinvent names for them on arrival in the high islands of Eastern Polynesia.

The ‘words and things’ technique of historical reconstruction is based on the assumption that if a word is reconstructable for a protolanguage, then the speakers of that language must have been familiar with its referent.

In what was probably the first systematic application of this technique in the Pacific, Pawley and Green (1971) studied relevant vocabulary reconstructable to Proto Polynesian and concluded that the speakers of Proto Polynesian ‘occupied or lived near an environment where, for example, mountains, cliffs, rivers, lakes, landslides and, probably, volcanic rock were found. That is, the community lived on or near a high island or large land mass, rather than a remote atoll.’ They added: ‘the presence in PPN of many terms for plants characteristic of the Indo-Pacific tropical zone indicates that the location lay within this zone’ (Pawley and Green 1971:17). More specifically, ‘the PPN speech community were fishermen-horticulturalists, familiar with a typical tropical Indo-Pacific high island environment and also with certain objects found natively only on certain islands of this category, including the *balolo* worm, the pearl oyster, such land animals as snakes, pestiferous mosquitoes, bats, owls, rails, pigeons, parrots. [It is] highly unlikely that the homeland lay anywhere in East Polynesia, or in marginal regions of West Polynesia’ (Pawley and Green 1971:23).

In the same paper, Pawley and Green proposed a number of postulates, i.e., ways of drawing further historical inferences from the linguistic data. They suggested, for instance, that the corollory of the first tenet of ‘words and things’ is that if a protolanguage had no name for a thing, then it was probably absent from their environment. They cited (following Biggs) the example of ‘seal’, a concept for which there is no Proto Polynesian reconstruction, although some individual Polynesian languages do have words for it. This suggests that seals were not present in the Polynesian homeland. Again following Biggs, they noted that reflexes of PPn *namu ‘mosquito’ in some Eastern Polynesian languages do not mean ‘mosquito’ (e.g. ‘biting midge’ in Maori and Marquesan), and inferred that mosquitoes did not exist in these places at the time of settlement (Pawley and Green 1971:19).

More recently, there have been some striking successes of the ‘words and things’ technique in and around Polynesia. One concerns the word for ‘megapode’, a flightless bird

which buries its eggs in the sand to hatch, hence also known as the 'incubator bird'. Clark (1982:126) noted that the name for the megapode in Tonga, *malau*, is related to the names for similar birds in Vanuatu, Solomon Islands, and New Guinea. Clark argued that since we must reconstruct *malau as the word for 'megapode' in Proto Oceanic, and there must have been an unbroken transmission of this word from Proto Oceanic to Proto Polynesian for it to appear in Tongan, then it must also have been part of the lexicon of Proto Central Pacific, which is believed to have been spoken in Fiji. So 'words and things' requires that during the Lapita period, when Proto Central Pacific was spoken, megapodes must have been present in Fiji, even though there are currently none, nor have there been in recorded history. Soon after Clark's observation, the archaeologist Simon Best unearthed the remains of at least two different species of megapode in Fiji, both of which became extinct soon after initial human settlement (Clunie 1984:140)—a dramatic vindication of 'words and things'.

The 'words and things' technique has, of course, limitations. While the logic of 'if they have a word for it they must know it, so it must be there' is unassailable, there is always the possibility that the reconstruction is in some way flawed, and it is possible to reconstruct apparently ancient words that are not ancient at all. Geraghty (2004a:65–66) notes that a word for 'motor-car' can be reconstructed for Proto Micronesian (it is actually a loan from Japanese), and similarly *sāmala 'hammer' can be reconstructed for Proto Polynesian—both cases are relatively recent loanwords that have been 'etymologically borrowed' (Geraghty 2004a:77–78) among related languages. Similarly, with regard to the parent language of the family to which Pacific languages belong, Proto Austronesian, Mahdi (1994) has shown that while words for 'iron', 'gold', 'silver' and some other metals and useful plants can be reconstructed, it is highly unlikely that the speakers of Proto Austronesian had any knowledge of them—they were all introduced well after the break-up of Proto Austronesian. Another potential source of confusion is that related words can undergo parallel semantic extension: Crowley (1994:87) has pointed out that *tusi 'book' can be reconstructed for Proto Polynesian (it originally meant 'mark or adorn with colour'), to which can be added *faʔo 'nail' (originally a wooden peg used for fastening). It is important, then, that this method be applied with caution.

Bearing in mind these provisos, we now turn to plant-names, and other words relevant to the environment that can be reconstructed for Proto Eastern Polynesian. Not all are of interest. For example, we can reconstruct PEP *futu 'k large coastal tree, *Barringtonia asiatica*', but this is not particularly useful, since it does not distinguish between volcanic islands and atolls as the homeland of the Eastern Polynesians, this particular tree being found in profusion in both ecosystems. What do concern us are words in the PEP lexicon that refer to high island entities and are *not* continuations of PNP vocabulary, which would suggest that the referents had not been part of the environment of pre-PEP speakers, since they lived in an atoll environment. Conversely, if we find that names of plants that are confined to high islands and absent from atolls continue from PPn and PNP into PEP, then that would suggest that the speakers of PEP came from a high island environment.

3 Regular changes and tendencies

Curiously for such a well-defined subgroup, Eastern Polynesian shows no regular phonological innovations, other than the purely phonetic, and in any case debatable, *l > *r (Marck 2000:23–25). Elbert (1953:154) pointed out the tendency for EP languages, and also some outliers, to merge *f and *s as *h, but this strictly speaking does not constitute a unique shared innovation of PEP; and Haudricourt's (1964:389) observation that *faf-

became *vah- holds only for Proto Central Eastern Polynesian, in other words, is not valid for Rapanui.[5]

The plant-name data presented below do, however, when combined with data from other domains, suggest another phonological tendency of Proto Eastern Polynesian and its close relatives and daughter languages: for pretonic high vowels to become a low (or mid) vowel. I do not intend to explore this in detail in this paper, but the following are suggestive.[6]

*ŋiŋie 'k. coastal shrub, *Pemphis acidula'* > Puk, Tok, PEP *ŋaŋie

*(hs)ulufe 'k. fern, *Dicranopteris linearis'* > PTa *aruhe

*muti(a,e) 'grass' > PTa *matie (Pollex *mutie; *mutia is indicated by Sam, Tok *mutia* and the external evidence of EF *vūtia, mūtia* 'sea-grass')

*tufuŋa 'expert, priest' > PTa *tahuŋa, Haw *kahuna*,[7] Tua *tohūŋa* 'priest', Mao *tohuŋa*

Another possible tendency is for a final mid back vowel to be lowered:

*hoŋohoŋo 'k. nettle, *Laportea interrupta'* > PCE *oŋaoŋa (Mqa *okaoka, onaona*, Tua *oŋaoŋa*, Mao *oŋaoŋa*). (Pollex PPn *hoŋahoŋa 'nettle or other stinging plant' is incorrect, since EF *soŋa* 'sago palm' is probably not cognate, whereas Rot *usoŋo* '*Laportea interrupta'* is)

*(ka) ŋaoso 'k. shrub, *Caesalpinia*' > Mqa *kaeha, keoha* (but Rpn *ŋaoho*)

Data presented below also bear witness to a tendency for PPn words (usually, or maybe exclusively, nouns) to acquire in PEP, and its close relatives and daughter languages, a prefix consisting of a stop and a vowel, *kō- being the most common.[8] In the following comparisons the first reconstructions are all PPn.

*fai 'k large tree, leguminid' > PCE *kofai 'pod-bearing plant, *Sesbania*' (Tah *ʔofai*, Haw *ʔohai*, Tua *kohai*)[9]

*fatu 'stone' > PCE *pōfatu, *kōfatu (Mia, Mao)

*felo '*Ficus tinctoria*, k. banyan with yellow-red berries' > Haw *ʔōhelo* 'k. shrub, *Vaccinium* spp., bears yellow or red berries' (cf. PPn *felo 'yellowish, reddish')

*(fua)fua 'young mullet' > Haw *ʔōhua* 'young of certain fish'

*fue 'k shore creeping vine' > PCE *pōfue (also Puk)

*kili 'saw, file' > PEP *kōkili 'triggerfish' (also Puk)

*kisi 'k. grass, *Oxalis*' > Rar *kōkiʔi*, Mao *kōkihi* '*Tetragonia*'

*polo '*Solanum nigrum'* > PCE *kōporo

*reŋa 'turmeric' > Haw *'ōlena* '*Curcuma*'

*taʔe 'faeces' > PEP *tūtaʔe (also Kap, Nuk, Tak)

*ʔura 'crayfish' > PEP *kōʔura

---

5 Nor is it valid for an apparently older stratum of Mangareva vocabulary (Fischer 2001).

6 Three of these were noted as sporadic sound changes of Proto Tahitic by Marck (2000:134).

7 Marck (2000:134) claims that this is a loan from Tahitian, but offers no evidence, so appears to be begging the question.

8 Biggs (1994:22) notes that New Zealand's first settlers made extensive use of prefixation of *poo-* and *koo-*, which seem to have had the meaning 'pseudo-' or 'like'.

9 Ton, EUv *ʔōhai* '*Delonix*' are presumably nineteenth century loans from Hawai'ian.

Finally, a small number of forms suggest a tendency for reduplication in PPn to become deleted or reduced in PEP and/or PCE:

*kakamika '*Sigesbeckia, Ageratum*' > *kamika '*Sigesbeckia*'
*kisikisi '*Oxalis*' > *kisi
*palapala 'k. tree-fern, *Cyathea*' > *pala 'k. fern, *Marattia*'
*talatalaʔamoa '*Caesalpinia*' > *tātaraʔamoa

while two show the reverse:

*kaso 'reed, *Miscanthus* sp' > *kākaso
*kawa '*Pittosporum*' (WF *kawa*) > *ka(wa)kawa (Haw *'aʔawa*, Rar *kavakava*)

4 Results of survey

I collected and compared plant-names for most Polynesian languages, using the standard dictionaries in most cases, and compared them to plant-names of Rotuma and Fiji and some other Oceanic languages. In addition I referred to Pollex (Biggs and Clark n.d.), Göthesson (1997), and Rensch (2005), all very useful sources of information on Polynesian plants and their names. The only major language I did not study in detail was New Zealand Maori, which I judged to be less useful because of its non-tropical location, and in any case is well covered in Pollex and works such as Biggs (1991, 1994). Doubtless a detailed study of New Zealand Maori plant names, and older dictionaries and word-lists of Polynesian languages, would yield more results and cognate sets, but probably not affect the major conclusions of this study.

A large number of the plants of Western Polynesia are simply not found in Eastern Polynesia, so are irrelevant to our discussion. Nevertheless, I list below those that have a PPn reconstruction. For those reconstructions which are not found in Pollex, or differ in some way from the Pollex entry, I add some supporting data.

Table 1: Coastal plants not found in Eastern Polynesia

*lekileki	'*Xylocarpus granatum*'	
*saŋale	'*Lumnitzera littorea*'	(EF *saŋale*; Ton *haŋale* 'k. tree which like the mangrove grows in the sea', Tuv *saŋale* 'k. tree', Ren *saŋaqe*)
*sinu	'*Excoecaria agallocha*'	(EF *sinu*; MFa *sinu*, WUv *sinu*)
*tātāŋia	'*Acacia simplicifolia*'	(EF *tātāŋia*; Sam *tātāŋia*)
*toŋo	'mangrove, *Bruguiera* and *Rhizophora*'	

Table 2: Non-coastal plants not found in Eastern Polynesia

*aka	'*Pueraria lobata*'	
*alu/walu	'*Epipremnum pinnatum*'	(EF *yalu*, *walu*; Ton *alu*, Tik *valu*; Rot *rauvaru* loaned from Pn *rau-walu)
*asi	'*Syzigium* sp., not cultivated, not fragrant or edible'	(WF *yasi* 'S curvistylum', EF *yasi*; EFu *asi* '*S clusiifolium*', Sam *asi* '*S inophillodes*')

*ate	*'Wedelia'*	(also *kofekofe)
*filimoto	*'Flacourtia rukam'*	
*fiso	*'Saccharum edulis'*	
*fukafuka	*'Kleinhovia hospita'*	(Ton *fukofuka*, Sam *fuʔafuʔa*, Tik *fukafuka*)
*ŋī	'*Imperata'*	(EF *ŋī*; WUv *ŋī*, WFu *ŋi* 'k. tall thin-bladed grass or sedge', Mae *ŋī* 'k. plant')
*kalaʔapusi	*'Acalypha grandis'* [10]	
*kanume	*'Diospyros elleptica, D ferrea'*	
*kofekofe	*'Wedelia'*	(EF *kovekove*; Tik *kofekofe*; also *ate, possible convergent development from PCP *kove 'bamboo')
*laŋakali	*'Aglaia saltatorum'*	
*loŋoloŋo	*'Cycas rumphii'*	
*manaui	*'Garuga* sp.*'* [11]	
*mapa	*'Diospyros* sp.*'*	
*moli	*'Citrus* spp.*'*	
*nukanuka	*'Decaspermum vitiense'*	
*pau	'k. large hardwood tree, *Palaquium, Planchonella'*	
*pele	*'Abelmoschus manihot'* (probably a recent introduction from Fiji to Polynesia, see Geraghty 2004a:85)	
*poumuli	*'Flueggea flexuosa'*	(EF *baumuri*; Ton, Sam *poumuli*, Tik *poumuri*)
*salato	*'Laportea/Dendrocnide harveyi'*	
*sea	*'Parinarium insularum'*	
*talafalu	*'Micromelum'*	(WF *tavolali*; Sam *talafalu*; Rot *tarfai* Pn loan)
*tamanu	*'Calophyllum vitiense, C samoense* and other inland species'	
*tanetane	*'Polyscias multijuga'*	
*taputoki	*'Alectryon grandifolius'*	
*tawa	*'Pometia pinnata'*	
*tawahi	*'Rhus taitensis'*	
*usi	*'Evodia* sp.*'*	

[10] Marck (2000:64) cites Ton *kalakalaʔapusi* as an example of Pn *s 'which has yet to change into Tongan *h*'; I believe a more likely explanation is that a former *kalakalaʔapuhi has been quite recently changed to *kalakalaʔapusi* by analogy with Ton *pusi* 'cat', the tail of which the flower of this plant resembles.

[11] Note that this appears to be a direct reflex of PCP *ʔui (<PEO *ʔuRi *Spondias dulcis*), with the mala-/mana- prefix meaning 'like, false', though Tongan should be *manaʔui. The widespread Polynesian *wī *Spondias dulcis* is clearly a loan from Fijian (Geraghty 2004a:87).

A similar list compiled by Biggs (1994:23) includes a number of taxa that are not listed here. In some cases it is because they are names that were replaced in Eastern Polynesia (e.g. *aŋo '*Curcuma*' replaced by *reŋa); in others, I believe the plants and/or their names to be relatively recent introductions, so not reconstructible to PPn (*fesi '*Intsia bijuga*', *mosokoi '*Cananga odorata*', *tono '*Centella asiatica*', *wālai 'a liana', see Geraghty 2004a); and some (e.g. *makari 'k. tree') are not sufficiently defined to determine whether or not they are also found in Eastern Polynesia.

Below are two lists of names of plants that are found in both Western and Eastern Polynesia, and have relatively secure reconstructions at both levels (or at least to Proto Central Eastern Polynesian). Note that neither list is claimed to be complete. The first is of coastal plants, i.e. those that can be found on atolls; the second is non-coastal plants, i.e. those that are not found on atolls.

Table 3: Atoll plants found in both West and East Polynesia

*aloalo/walowalo '*Premna taitensis*' > *warowaro (Tah, Mva)

*fano '*Guettardia speciosa*' > *fano (cf. PPn *puapua)

*fao '*Ochrosia*' > *fao (doubtful, only reflex being Haw *hao Rauvolfia*)

*fara '*Pandanus*' > *fara

*fatai '*Cassytha filiformis*' > *tainoka

*fau '*Hibiscus tiliaceus*' > *fau

*felo '*Ficus tinctoria*, k. banyan with yellow-red berries' > Haw '*ōhelo* 'k. shrub, *Vaccinium* spp., bears yellow or red berries' (cf. PPn *mati '*Ficus tinctoria*', *felo 'yellowish, reddish')

*feta‍ʔu '*Calophyllum inophyllum*' > *tamanu[12]

*fiso‍ʔa '*Colubrina asiatica*' > *tutu (cf. PPn *tutu)

*fue '*Convolvulus*, Ipomoea pes-caprae' > *pōfue (cf. PEP *fue '*Lagenaria vulgaris*')

*futu '*Barringtonia asiatica*' > futu

*ŋasu '*Scaevola*' > *ŋasu (Pen, Rar), *naupata (Tah, Haw)

*ŋiŋie '*Pemphis*' > *ŋaŋie (also Tok, Puk; but Tua *ŋieŋie*)

*(ka)ŋa‍ʔoso '*Caesalpinia*' (Sam *‍ʔanaoso, ‍ʔāŋoso*) > *ŋa‍ʔoso (Mqa *keoha*, Rpn *ŋa‍ʔoho*) (cf. *talatala‍ʔamoa)

*kanawa, *fakanawa, *tou '*Cordia subcordata*' > *tou (probably from PPn *tou 'tapa paste', made from *Cordia* fruit)

*katafa '*Asplenium nidus*' > *katafa

*katuli 'coastal herb *Portulaca*, *Boerhavia*' > *katuri (Haw, Tah, Pen, Mki)

*kaute '*Hibiscus rosa-sinensis*' > *kaute

*kie 'k. *Pandanus* used for fine mats' > *kie 'sail'

*kofe 'bamboo' > *kofe

*kulu '*Artocarpus*' > *kuru (cf. *mei)

*lala '*Vitex*' > *rara (Rar)

[12] The replacement of PPn *feta‍ʔu '*Calophyllum inophyllum*' by PCE *tamanu (from PPn *tamanu 'inland sp of *Calophyllum*') is, to say the least, unexpected. By a strict 'words and things' interpretation, this single item suggests that Eastern Polynesia was first colonised by inland dwellers—perhaps from Samoa—who had lost knowledge of *feta‍ʔu, so called it by the name of the inland species they were familiar with. No other evidence I have come across points to this conclusion.

*mahuku ‘grass’ > *mauku
*maile ‘*Alyxia, Polypodium’* > *maire
*mati ‘*Ficus tinctoria’* > *mati (cf. *felo)
*maʔutofu ‘*Urena, Sida’* > *kulima *Sida* (Haw *ʔūlima*, Tua *kurima*)
*mei ‘breadfruit’ > *mei (Mqa, Mva) (cf. *kulu; probably introduced from Micronesia, see Geraghty 2004a:87–88)
*milo ‘*Thespesia populnea’* > *miro
*muti(a,e) ‘grass’ > *mutie
*niu ‘*Cocos nucifera’* > *niu
*nonu ‘*Morinda citrifolia’* > *nonu
*pia ‘*Tacca’* > *pia
*pipi ‘*Hernandia’* > *puka
*pua ‘*Fagraea berteriana’* > *pua (Tah, Rar)
*puapua ‘*Guettardia speciosa’* > *fano (cf. PPn *fano)
*puka ‘*Pisonia’* > *puka, *pukatea
*puka ‘*Hernandia’* > *puka
*rewa ‘*Cerbera’* > *rewa (*Pollex* *lewa in error, there being no Tongic reflex)
*tafifi ‘*Ipomoea littoralis’* (EF *sovivi*; Niu *tefifi* ‘*Ipomoea* sp.’, Sam *lautafifi*) > *tafifi ‘k. creeper’ (Tua *tāhihi* Alyxia, Tah *tāfifi* Alyxia, Mqa *tāfifi*)
*talatalaʔamoa ‘*Caesalpinia’* > *tātaraʔamoa (cf. *(ka)ŋaʔoso)
*tamole ‘*Portulaca’* > *katuri (cf. PCE *tamole ‘*Polygonum’* (Tah, Haw))
*tausun(i,u) ‘*Tournefortia argentea’* > *tausinu
*tiale ‘*Gardenia taitensis*’ > *tiare
*toa ‘*Casuarina equisetifolia’* > *toa
*tutu ‘*Colubrina asiatica’* > *tutu (cf. *fisoʔa)

Table 4: High island plants found in both West and East Polynesia

*aloŋā ‘*Pipturus aregenteus’* > *oroŋā
*aŋo ‘*Curcuma’* > *reŋa (from PPn *reŋa ‘tumeric’)
*asi ‘sandalwood’ > *asi
*falasola ‘*Pandanus* var’ > *farasora (dubious, only reflex Mqa *faʔahoʔa* ‘pineapple’)
*fai ‘k. large tree, *Leguminid*’ > PCE *kofai ‘pod-bearing plant, *Sesbania’* (Tah *ʔofai*, Haw *ʔohai*, Tua *kohai*)
*fauʔui ‘*Grewia crenata’* > *faupā (Tah, Tua)
*fenua ‘*Macaranga harveyana’* (EF *venua*) > *fenua (Tah *fenia* ‘*Omalanthus nutans’* (same family, Euphorbiaceae), Rar *ʔenua*. Note also Niu *lēhau ‘Macaranga’*; and Mqa *ferua*, Haw *lehua*, both ‘*Metrosideros collina’*.
*fanuamamala ‘*Omalanthus’* > *fenua (Tah *fenia*)
*futi ‘*Musa’* > *maika
*ŋase ‘*Geniostoma’* (EUv) > *saŋe (Rar)
*ŋase ‘k. fern’ (EF *qase* ‘*Dendrobium’*) > *ŋahe (Tua *ŋahe* ‘k. giant fern’)
*ŋatae ‘*Erythrina indica’* > *ŋatae
*hiti ‘*Grewia’* (EF *siti*; EFu *iti*) > *faupā (Tah, Tua)

*hoŋohoŋo ‘k. nettle, *Laportea interrupta*’ > *oŋaoŋa (Mqa *okaoka, onaona*, Tua *oŋaoŋa*, Mao *oŋaoŋa*)
*(hs)ulufe ‘k. fern, *Dicranopteris linearis*’ > PEP *urufe (Haw), PTa *aruhe
*ifi ‘*Inocarpus edulis*’ > *ifi
*kafika ‘*Syzygium malaccense*’ > *kafika
*kakamika ‘*Sigesbeckia, Ageratum*’ > *kamika *Sigesbeckia*
*kalaka ‘*Planchonella*’ > *karaka
*kape ‘*Alocasia*’ > *kape
*kaso ‘reed, *Miscanthus* sp.’ > *kākaso
*kawa ‘*Pittosporum*’ (WF *kawa*) > *ka(wa)kawa (Haw *ʔaʔawa*, Rar *kavakava*)
*kawa ‘*Piper methysticum*’ > *kawa
*kawakawaʔatua ‘*Piper latifolium*’ > *kawakawaʔatua
*kawasusu ‘*Tephrosia*’ > *kaususu, *sora
*kiekie ‘*Freycinetia*’ > *kiekie
*kisikisi ‘*Oxalis*’ > *kisi
*koka ‘*Bischofia javanica*’ > *koka (Rar)
*kuta ‘k sedge, *Eleocharis*’ > *kuta ‘k reed’ (Mia, Mao)
*laupata ‘*Macaranga*’ > *naupata ‘*Scaevola*’ (problematic on both phonological and semantic grounds)
*maŋele ‘*Trema*’ > *maŋele ‘k. tree’ (Mqa ‘*Alphitonia*’, Haw ‘k. tree’)
*mako ‘*Trichospermum, Melochia*’ > *mako ‘*Melochia*’ (Tah)
*manon(o,u) ‘*Tarenna sambucina*’ > *manono
*masame ‘*Glochidion ramiflorum*’ > *masame (Tah, Rar)
*maʔota ‘*Dysoxylum*’ > *maʔota (Mao)
*palapala ‘k. tree-fern, *Cyathea*’ > *mamaku, *pala ‘k. fern, *Marattia*’, cf. *poŋa
*palai ‘*Dioscorea nummularia*’ > *parai (Tah, Rar)
*pilita ‘*Dioscorea pentaphylla*’ > *pirita (Tua, Rar)
*poŋa ‘k tree-fern, *Cyathea*’ > *poŋa (Rar, Mao), cf *palapala (perhaps from PPn *poŋa ‘hole, orifice’, with reference to hollow trunk)
*polo ‘*Solanum nigrum*’ > *poro (Rar), *pōporo (Haw, Rpn), *poroporo (Haw, Tua), *kōporo (Tah, Mqa, Rap)
*siapo ‘*Broussonetia*’ > *kaute (Tah), cf. *kaute ‘*Hibiscus rosa-sinensis*’
*soaka ‘*Musa fehi*’ > *fekī, fuatū
*soi ‘*Dioscorea bulbifera*’ > *soi
*tewe ‘*Amorphophallus campanulatus*’ > *tewe
*tī ‘*Cordyline fruticosa*’ > *tī
*toi ‘*Alphitonia*’ > *toi (Tah, Rar)
*toto ‘*Euphorbia*’ > *(ka)toto (Haw, Tah)
*tuitui ‘*Aleurites moluccana*’ > *tuitui, *tutui (Haw, Tah)
*wawae ‘*Gossypium barbadense*’ > *wawai
*ʔafatea ‘*Nauclea*’ > *ʔafatea (Tah)
*ʔufi ‘*Dioscorea alata*’ > *ʔufi

The above tables leave little doubt that many names of high island plants persisted from Proto Polynesian and Proto Nuclear Polynesian into Proto Eastern Polynesian and Proto Central Eastern Polynesian. Particularly striking are *kalaka '*Planchonella'*, *kawa '*Piper methysticum*', and *koka '*Bischofia javanicus*', all relatively widespread in Eastern Polynesia, and none of which is found on any of the northern outliers that Wilson (1985) has proposed as the source for the earliest settlers of Eastern Polynesia.

A quick glance at other semantic fields suggests that results would be similar. For instance, the names of freshwater fish *hinaŋa 'whitebait' and *tuna 'freshwater eel' continue into Proto Eastern Polynesian, as do high island topographical features such as *kalā 'hard, black volcanic stone', *mato 'precipice, steep place, cliff', and *solo 'landslide'.

15 Discussion

Even though there are many etyma that indicate that PEP speakers had knowledge of exclusively high island plants and other features of the environment, this does not entirely rule out the possibility that EP was settled initially from an atoll or atolls. There are at least two scenarios, not necessarily mutually exclusive, which would allow the speakers of PEP to originate from an atoll environment, yet for words for high island phenomena to be reconstructible for PEP:

1. The atoll dwellers had knowledge of high island environments; or
2. An initial colonisation by atoll dwellers was followed by a colonisation from a high island or high islands, at such an early date as to be perceived as simultaneous linguistically, that is, before any sound change[13] or further significant population movement.

Regarding option one, that atoll dwellers, in this case in the northern outliers, at the time of the first settlement of East Polynesia, had detailed knowledge of high island environments, there is no way that we can be sure whether or not this was the case. Extrapolating from relatively recent times, it is true that Tuvaluans and Tokelauans apparently had no knowledge of Samoa around 1840 (Hale 1846:153, 5, 65), but they did know of at least some high island produce. As Hale (1846:166) noted in Vaitupu, Tuvalu: 'Yams and bananas they knew by name, but had none.' Tuamotuans in the late 18th and early 19th centuries often paid extended visits to Tahiti (Haddon and Hornell 1975:79)—though there is no indication of how well they knew the Tahitian environment.

Postulate four of Pawley and Green (1971:17) stated that 'the presence in a proto-language of a term denoting a category of objects is taken as indicating that the referents were familiar to the speakers of the language, either as part of their own immediate environment or as part of a nearby environment.' In Postulate five, Pawley and Green then defined 'nearby', taking their cue from Sharp's (1963) theory of accidental voyaging: 'deliberate two-way voyaging over distances exceeding two or three hundred miles across open sea, and using indigenous craft, is... unlikely to have occurred. We thus postulate a radius of three hundred miles around any point as the upper range of 'nearby environments'. Even by this very parsimonious estimate, inhabitants of the North Central Outliers would have had in their 'nearby environment' the high islands of the Solomons,

[13] Even after a sound change, words borrowed 'etymologically' would be linguistically invisible (Geraghty 2004a:77–78).

likewise Nukuoro, which is less than three hundred miles from Pohnpei, and Kapingamarangi, which is approximately three hundred miles from New Ireland.

Since those isolationist days ushered in by Sharp and others, the pendulum has swung the other way (Finney 1994:255–259) and there are few now who would deny that prehistoric Polynesians were adventurous long–distance voyagers, and that the 300 mile limit is an underestimate. 'They are a race of navigators, and often undertake long voyages in vessels in which our own sailors would hesitate to cross a harbour [..] not only is a constant communication kept up among the different islands of each group of Polynesia, but perilous voyages of many days between different groups are frequent' (Hale 1846:14). We know from the Ra'iatean Tupa'ia and other navigators encountered by Spanish and British explorers in the eighteenth century that Tahitians knew at least the names and approximate locations of all the islands of triangle Polynesia (except the extremities of Hawai'i , Mangareva, Rapanui and New Zealand) as well as Fiji and Rotuma (Hale 1846:124, Dening 1962:103, 135). Rotuma is about 4000 km (2400 miles) distant from Tahiti. Moreover Tupa'ia indicated that his father had even greater knowledge of the islands of the Pacific (Beaglehole 1968:157, Dening 1962:105), and we know from other sources that Polynesian navigation had been in decline since the 'little ice age' that began around 1350 AD (Nunn 2008). In Western Polynesia, Tongans told Cook of islands they knew as far as Kiribati and probably also the Solomons (Geraghty 2004b), and we can infer from linguistic and other evidence that Tongans, or other western Polynesians, travelled to and from places as far away as Vanuatu, Pohnpei (Geraghty 1994), and the Carolines. The Marquesans also have legends of voyages to and from Rarotonga to procure red feathers (Langridge and Terrell 1988:11–31). In sum, I would venture to suggest that when Polynesian voyaging was at its peak, the 'nearby environment' with which Polynesians were familiar could well have stretched to a thousand miles or even more.

The acceptance of such voyaging capabilities explains some apparent 'words and things' anomalies. For example, the Proto Central Pacific and Proto Polynesian reconstruction *lulu 'owl', has come to refer to a sea bird, usually the booby (*Sula* sp.), in Eastern Polynesian languages spoken where there are no owls (that is, all except Hawai'ian and New Zealand Maori). However, in New Zealand Maori, the referent is again the owl. While it is possible that the name for booby was transferred back to the owl, and even remotely possible that owls once existed in Central Eastern Polynesia, the most likely explanation is simply that the Eastern Polynesians who colonised New Zealand were familiar with owls, and their name, from voyaging to Western Polynesia.

Similarly, the Hawai'ian word for the tree *Myoporum, naio,* corresponds exactly to *ŋaio,* its name in the Austral Islands, Cook Islands, and New Zealand (it is only found in Eastern Polynesia). However the genus is absent not only in the Marquesas, whose languages subgroup with Hawai'ian within Eastern Polynesian, but also in the Society Islands, where part of the Hawai'ian lexicon is believed to have originated (Whistler 1995:51). The mystery of this 'words and things' conundrum dissipates when we acknowledge that the prehistoric Polynesians' world was far from confined to their own island group. It is hardly surprising that such well-travelled people should be familiar with useful plants—*Myoporum* was used as sandalwood and in house construction in Hawai'i (Degener 1973:267–268), while in Rarotonga the flowers are used to scent coconut oil—in neighbouring island groups.

The second scenario of Eastern Polynesian colonisation (not mutually exclusive with the first) that is consistent with Wilson's thesis is that an initial colonisation by atoll dwellers was followed by colonisation from a high island or high islands, at such an early

date as to be perceived of as simultaneous linguistically, that is, before any sound change or further significant population movement. In other words, that Eastern Polynesia was not colonised only once, but twice or a number of times, from different Western Polynesian sources, initially from atolls, but subsequently from high islands, and that Proto Eastern Polynesian could have been lexically enriched by later colonists from high islands of Western Polynesia.

As with the notion of limited voyaging ability, the 1960s notion that each Polynesian island or group was colonised only once has succumbed over the years to the weight of evidence from many fields (Finney 1994:263–270). As noted by Kirch (2000:244–245), recent work on language relationships, voyaging, and long-distance interaction spheres (to which could be added oral traditions) all suggest that 'rather than a single population movement into one island or archipelago of central Eastern Polynesia, which then served as a primary dispersal center … the process of expansion out of the Ancestral Polynesian homeland was more complex, involving at least three separate movements, each resulting in interaction spheres and dialect chains that persisted over significant time periods.' While I do not agree with Kirch's specific proposals, I believe that the idea of initial colonisation from northern outliers followed by a number of intrusions from elsewhere is substantially correct.[14]

References

Beaglehole, J.C., ed. 1968. *The journals of Captain James Cook on his voyages of discovery.* vol.1. *The voyage of the Endeavour 1768–1771*. Published for the Hakluyt Society. Cambridge: Cambridge University Press.

Besnier, Niko. 1981. *Tuvaluan lexicon*. Funafuti: United States Peace Corps.

Biggs, Bruce. 1991. A linguist revisits the New Zealand bush. In Andrew Pawley, ed. *Man and a half: essays in Pacific anthropology and ethnobiology in honour of Ralph Bulmer*, 67–72. Auckland: The Polynesian Society.

—— 1994. New words for a new world. In Pawley and Ross, eds 1994:21–29.

Biggs, Bruce and Ross Clark. n.d. Pollex.

Buse, Jasper and Raututi Taringa. 1995. *Cook Islands Maori dictionary*. Cook Islands Ministry of Education et al.

Carroll, Vern and Tobias Soulik. 1973. *Nukuoro lexicon*. PALI Language Texts: Polynesia. Honolulu: University Press of Hawaii.

Churchward, C. Maxwell. 1959. Tongan dictionary. Oxford: Oxford University Press.

---

[14] The name of the island Tahiti suggests that it was settled from Samoa, which lies almost due west, being composed of *ta 'the' + *fiti (> *hiti*) 'east'. Pollex does not currently list *ta as an alternative form of the definite article *te, but the evidence for it in at least Nuclear Polynesian is compelling: WFu *ta*, Haw *ka*, Mqa *ta* (Hale 1846:134–135), and loans such as Rot *tarau* 'hundred' (PPn *rau) and Pohnpei *sakau* 'kava' (PPn *kawa).

A possible explanation for the shared innovations pointed out by Wilson (1985), which I will not explore in this paper, is that Eastern Polynesia was colonised from Samoa at a time when the language spoken in Samoa subgrouped with those now spoken in the Equatorial Outliers, and that Samoan has changed dramatically since then due to both internal change and innovations spreading from other languages in the Western Polynesian interaction zone (Pawley 1996:401–403).

Clark, Ross. 1982. Proto Polynesian birds. In J. Siikala, ed. *Oceanic Studies: essays in honour of Aarne A. Koskinen*, 121–43. Helsinki: The Finnish Anthropological Society.

—— 1998. *A dictionary of the Mele language (Atara Imere), Vanuatu*. Canberra: Pacific Linguistics.

Clunie, Fergus. 1984. *Birds of the Fiji bush*. Suva: Fiji Museum.

Crowley, Terry. 1994. Proto-who drank kava? In Pawley and Ross, eds 1994:87–100.

Degener, Otto. 1973. *Plants of Hawaii National Parks illustrative of plants and customs of the South Seas.* Ann Arbor: Braun-Brumfield.

Dening, G.M. 1962. The geographical knowledge of the Polynesians and the nature of inter-island contact. In Jack Golson, ed. *Polynesian navigation*, 102–153. Wellington: A.H. and A.W. Reed, for the Polynesian Society.

Elbert, Samuel H. 1953. Internal relationships of the Polynesian languages and dialects. *Southwestern Journal of Anthropology* 9:147–173.

—— 1975. *Dictionary of the language of Rennell and Bellona*. Copenhagen: National Museum of Denmark.

Finney, Ben. 1994. *Voyage of rediscovery: a cultural odyssey through Polynesia*. Berkeley: University of California Press.

Firth, Raymond. 1985. *Tikopia–English dictionary*. Auckland and Oxford: Auckland University Press and Oxford University Press.

Fischer, Steven Roger. 2001. Mangarevan doublets : preliminary evidence for Proto-Southeast Polynesian. *Oceanic Linguistics* 40,1:112–24.

Geraghty, Paul A. 1983. *The history of the Fijian languages*. Oceanic Linguistics Special Publication 19. Honolulu: University of Hawaii Press.

—— 1994. Linguistic evidence for the Tongan Empire. In Tom Dutton and Darrell T. Tryon, eds *Language contact and change in the Austronesian world*. Trends in Linguistics Studies and Monographs 77, 223–349. Berlin: Mouton de Gruyter.

—— 2004a. Borrowed plants in Fiji and Polynesia: some linguistic evidence. In Jan Tent and Paul Geraghty, eds *Borrowing: a Pacific perspective*, 65–98. Canberra: Pacific Linguistics.

—— 2004b. Polynesian loans in the Solomon Islands. *Rongorongo Studies* 14,2:43–68

Göthesson, Lars-Åke. 1997. *Plants of the Pitcairn Islands including local names and uses.* Sydney: Centre for South Pacific Studies, University of New South Wales.

Green, Roger. 1966. Linguistic subgrouping within Polynesia: the implications for prehistoric settlement. *Journal of the Polynesian Society* 75,1:6–38.

Haddon, A.C. and James Hornell. 1975. *Canoes of Oceania*. Special Publications 27, 28, and 29. Honolulu: Bernice P Bishop Museum.

Hale, Horatio. 1846. *United States exploring expedition during the years 1838, 1839, 1840, 1841, 1842, under the command of Charles Wilkes U.S.N: ethnography and philology*. Philadelphia: Lea and Blanchard.

Haudricourt, A.G. 1964. Comment on G.W. Grace, 'Movement of the Malayo-Polynesians: 1500 B.C. to A.D. 500. The linguistic evidence.' *Current Anthropology* 5:389–390.

Hollyman, K.J. 1987. *De Muna Fagauvea — I: Dictionnaire fagauvea–français*. Te Reo monographs. Auckland: Linguistic Society of New Zealand.

Inia, Elizabeth K., Sofie Arntsen, Hans Schmidt, Jan Rensel and Alan Howard. 1998. *A new Rotuman dictionary*. Suva: Institute of Pacific Studies, University of the South Pacific.

Kirch, Patrick Vinton. 2000. *On the road of the winds: an archaeological history of thePacific Islands before European contact*. Berkeley: University of California Press.

Langridge, Marta and Jennifer Terrell. 1988. *Von den Steinen's Marquesan myths*. Canberra: Target Oceania and The Journal of Pacific History.

Lemaître, Yves. 1973. *Lexique du tahitien contemporain*. Paris: Editions de l'ORSTOM.

Mahdi, Waruno. 1994. Some Austronesian maverick protoforms with culture–historical implications — I. *Oceanic Linguistics* 33,1:167–229.

Marck, Jeff. 2000. *Topics in Polynesian language and culture history*. Canberra: Pacific Linguistics.

Milner, G.B. 1966. *Samoan dictionary*. Oxford: Oxford University Press.

Moyse-Faurie, Claire. 1993. *Dictionnaire futunien–français*. Paris: Peeters.

Nunn, Patrick D. 2008. A shock to the system: climatic disruption to Pacific Island societies around AD 1300. Paper presented at the Pacific History Association Conference, University of the South Pacific, Suva.

Pawley, Andrew. 1966. Polynesian languages: a subgrouping based on shared innovations in morphology. *Journal of the Polynesian Society* 75,1:39–64

—— 1996. On the Polynesian subgroup as a problem for Irwin's continuous settlement hypothesis. In J.M. Davidson et al., eds *Oceanic culture history: essays in honour of Roger Green*, 387–410. New Zealand Journal of Archaeology Special Publication.

Pawley, Andrew and Kaye Green. 1971. Lexical evidence for the Proto-Polynesian homeland. *Te Reo* 14:1–35.

Pawley, Andrew and Malcolm Ross, eds. 1994. *Austronesian terminologies: continuity and change*. Canberra: Pacific Linguistics.

Rehg, Kenneth L. and Damian G. Sohl. 1979. *Ponapean–English dictionary*. PALI Language Texts: Micronesia. Honolulu: University of Hawaii Press.

Rensch, Karl H. 1984. *Tikisionalio Fakauvea–Fakafalani/Dictionnaire Wallisien–Français*. Canberra: Pacific Linguistics.

—— 2005. *Plant names of Eastern Polynesia*. Canberra: Archipelago Press.

Sallés, Arturo Hernández, Nelly Ramos Pizarro, et al. 2001. *Diccionario Illustrado Rapa Nui, Español, Ingles, Francés*. Santiago: Pehuén.

Sharp, Andrew. 1963. *Ancient voyagers in Polynesia*. Wellington: Angus and Robertson.

Sperlich, Wolfgang B. 1997. *Tohi vagahau Niue: Niue language dictionary, Niuean–English with English–Niuean finderlist*. Government of Niue.

Stimson, J. Frank with the collaboration of Donald Stanley Marshall. 1964. *A dictionary of some Tuamotuan dialects of the Polynesian language*. The Hague: Martinus Nijhoff.

Triffitt, Geraldine. 2000. The dialects of the Yasawa Islands of Fiji. In Bill Palmer and Paul Geraghty, eds *SICOL: Proceedings of the Second International Conference on Oceanic Linguistics:* vol. 2*, Historical and descriptive studies,* 315–327. Canberra: Pacific Linguistics.

Whistler, W. Arthur. 1995. Folk plant nomenclature in Polynesia. *Pacific Studies* 18,4:39–59

Wilson, William H. 1985. Evidence for an outlier source for the Proto Eastern Polynesian pronominal system. *Oceanic Linguistics* 24:85–133.

—— 2008. Source of Tongic-like developments in PPN *r and *h in Eastern Polynesia. Paper presented at the Pacific Roots Symposium, University of the South Pacific, Suva, Fiji.

27 *Some clan names of the Chuukic-speaking peoples of Micronesia*

JEFF MARCK

1 Introduction

This report considers the antiquity of the clan names of the Chuukic-speaking peoples. The Chuukic-speaking islands (Figure 1) constitute the largest region of cognate matrilineal or patrilineal clan names in Oceania. We are presently confronted with a diverse Chuukic clan situation. Clan numbers are small on the atolls, usually less than ten. New clans abound in Chuuk Lagoon where there are now more than 80 clans. Saipan and other Mariana Islands have more than others due to immigration in the historic period. Twenty-seven clan names were found that occur in two or more of the Chuukic-speaking islands (counting Chuuk Lagoon as a single island) (Table 1). As few as six show evidence that allow attribution to Proto Chuukic, the language spoken around the Chuuk Lagoon[1] ca. AD 500–1000 before spreading on to the atolls of what is now Yap State, Federated States of Micronesian, and the atolls of the Republic of Palau.

The 27 clan names reconstructed with what may have been their Proto Chuukic[2] sounds are listed in Table 1 along with a guess at what may have been their meanings in English. I do not believe all 27 are as old as Proto Chuukic but it is convenient to begin by indexing each with a Proto Chuukic spelling. Table 2 then gives these same names alphabetically with their distributions according to islands for which I found regularly or irregularly agreeing forms. Table 3 then gives these same names grouped according to pattern of island distributions.

1 And perhaps the Mortlocks, Chuuk State's 'Western Islands' and Chuuk State's northern atolls.

2 The language spoken around Chuuk Lagoon some 1000 years ago or thereabouts at a time when those peoples were on the threshold of establishing permanent settlements on the atolls between Chuuk and Yap.

Alexander Adelaar and Andrew Pawley, eds
Austronesian historical linguistics and culture history: a festschrift for Robert Blust, 461–487.
Pacific Linguistics, 2009.
© This edition vested with Pacific Linguistics.

Table 1: The 27 clan names of this study given in Proto Chuukic orthography[3]

*faa-ni-mai	'Under the Breadfruit Tree'
*imwa-o	'House – ?'
*kainanga-i-liku	'Clan from Outside'
*kainanga-i-sawa	'Clan by the Pass in the Reef'
*kainanga-weneyaa	'Clan of Woleai'
*ka-sama-anga	'To Make the Outrigger'
*kaú-fanúa	'Land's Fishhook'
*luuka-(na)-fanúa	'Centre of the Island'
*mwangau-ni-faca	'Eaters of Pandanus'
*mwaoco	'Ashes'
*pike	'Sand Islet'
*pwalú	'Taro Swamp'
*raki	'Sailing Season'
*rape-fanúa	'? – Land/Island'
*talasi	'Loosen (?)'
*tapwo-ni-ppia	'Village at the Beach'
*taro	'Birthmark'
*tawu-afangi	'People of the North/Winter Tradewind'
*tawu-alai	'Tall People'
*tawu-fanaa-ciki	'Little Needlefish People'
*tawu-pelaya	'People of ?'
*tawu-pwolowasa	'People of Puluwat'
*tawu-wene	'Upright People'
*wao-ni-rae	'Forest (People)'
*wii-sakaú	'(People of) Namonuito'
*wi-TuuTuu[4]	'Opened (?)'
*(wu)waa-ni-kara	'Canoe of Sweetness (?)'

[3] Spelt 'as if' all are as old as Proto Chuukic which, I argue for many, seems unlikely.

[4] Proto Micronesian *T is reconstructed by Jackson (1983, 1984) and Bender et al. (2003a, 2003b) as a sound somehow distinct from Proto Micronesian *t. Upper case for *T is abandoned in the data and reconstruction section where upper case is used to begin proper nouns.

Table 2: Distributions of the clan names[5]

	Pur	Fai	Uli	Wol	Ifl	Far	Lam	Pul	Pol	Mur/ Uno	Chk	Mrt	Pon
*faa-ni-mai	X		X					X		X	X	X	X
*imwa-o								X			X	X	
*kainanga-i-liku					X						X		
*kainanga-i-sawa						X		X	X		X		
*kainanga-weneyaa				X			X	X			X		
*ka-sama-anga							X	X	X		X	X	
*kaú-fanúa			X	X	X		X				X		
*luuka-(na)-fanúa				X						X	X		X
*mwangau-ni-faca			X	X	X	X	X	X			X		
*mwaoco									X		X		
*pike									X		X		
*pwalú				X	X		X	X	X		X		
*raki	X						X				X		
*rape-fanúa					X						X		
*talesa										X	X		
*tapwo-ni-ppia								X	X		X	X	
*taro				X				X	X	X	X	X	
*tawu-afangi		X		X			X	X		X	X		
*tawu-alai								X			X		
*tawu-fanaa-ciki			X	X	X	X	X				X		
*tawu-pelaya									X		X		X?
*tawu-pwolowasa								X			X		
*tawu-wene			X	X	X	X	X				X	X	
*wao-ni-rae								X			X		
*wii-sakaú								X		X	X		
*wi-TuuTuu							X		X	X	X		
*(wu)waa-ni-kara											X	X	

[5] Abbreviations: Republic of Palau: Pur – Pulo Anna. Yap State FSM: Fai – Fais, Uli – Ulithi, Wol – Woleai, Ifl – Ifaluk, Far – Faraulep, Lam – Lamotrek. Chuuk State FSM: Pul – Puluwat, Pol – Pullap (Pollap), Mur – Murillo (Hall Islands – north of Chuuk Lagoon), Uno – Unon (Namonuito Islands – north-northwest of Chuuk Lagoon), Chk – Chuuk Lagoon, Mrt – Mortlock Islands (south and south-southeast of Chuuk Lagoon). Pohnpei State FSM: Pon – Pohnpei.

Table 3: Chuukic clan names grouped according to distribution

	Pur	Fai	Uli	Wol	Ifl	Far	Lam	Pul	Pol	Mur/ Uno	Chk	Mrt	Pon
*faa-ni-mai	X		X					X		X	X	X	X
*raki	X						X				X		
*tawu-afangi		X		X				X		X	X		
*taro				X			X	X	X	X	X	X	
*kaú-fanúa			X	X	X		X				X		
*luuka-(na)-fanúa				X						X	X		X
*mwangau-ni-faca			X	X	X	X	X	X			X		
*tawu-fanaa-ciki			X	X	X	X	X				X		
*tawu-wene			X	X	X	X	X				X	X	
*pwalú				X	X		X	X	X		X		
*kainanga-weneyaa				X			X	X			X		
*kainanga-i-liku					X						X		
*rape-fanúa					X						X		
*kainanga-i-sawa						X		X	X		X		
*ka-sama-anga							X	X	X		X	X	
*imwa-o								X			X	X	
*tapwo-ni-ppia								X	X		X	X	
*tawu-alai								X			X		
*tawu-pwolowasa								X			X		
*wao-ni-rae								X			X		
*wii-sakaú								X		X	X		
*mwaoco									X		X		
*pike									X		X		
*tawu-pelaya									X		X		X?
*wi-TuuTuu							X		X	X	X		
*talesa										X	X		
*(wu)waa-na-kara											X	X	

Having given the 27 clan names to be considered, I now present:

- A section on Proto Oceanic and Proto (Nuclear) Micronesian linguistic subgrouping and culture history (i.e., matrilineality).
- A section on sources.
- The main data and reconstructions section considering each of the clan names in their turn.
- Conclusions categorising the clan names according to apparent source in place and time.

2 The Proto Oceanic and Proto (Nuclear) Micronesian linguistic and matrilineal heritage

The clans of the Chuukic speakers are matrilineal as are the clans of most of Nuclear Micronesia (the geographical area of Nuclear Micronesian speech). The primary exception is found in Kiribati where there are neither matrilineal nor patrilineal clans. Pingelap atoll of Pohnpei State is also an exception in the sense that it has both matrilineal and patrilineal clans (Damas 1979, 1981; Schneider 1980), a situation referred to by anthropologists as 'double descent'. The most isolated northwesternly Marshall Island atolls are patrilineal and are the final exception. Certainly this vast region of matrilineality is due to descent from matrilineal forebears in Melanesia (cf. Hage 1998; Hage and Marck 2002, 2003; Kayser, et al. 2006; Marck MS). The matrilineality model for Ancestral Lapita society/ Proto Oceanic society, Nuclear Micronesian society's more or less immediate ancestor, is now ten years old (Hage 1998) and the evidence for it continues to grow in strength (Hage and Marck 2003; Kayser et al. 2006; Marck 2008).

A subsistence-based theory of Micronesian matrilineality has recently been developed by Petersen (2006). He suggests that a 'breadfruit revolution' of new hybrids emanating out of Kosrae and Pohnpei about a thousand years ago promoted and sustained matrilineality in much of Micronesia. I agree with Petersen's suggestion that the breadfruit revolution facilitated a more abundant life on the atolls, the spread of permanent, 'large' populations to the central and western Chuukic atolls being of the same approximate age as the breadfruit revolution. But Hage and Marck (2003) and now Kayser et al. (2006) posit that Proto Oceanic society was matrilocal as did Hage and Marck (2002) for Proto (Nuclear) Micronesian society. Petersen (2006:fn.1) ignores this evidence of ancient matrilocality and the notion that the Proto Oceanic and Proto (Nuclear) Micronesian societies were matrilineal because they were matrilocal. Rather, he forges on with an ideology-driven notion of the Micronesian situation. Hage and Marck (2003) as well as Marck (2008) attribute that residential pattern to long absences of seafaring adult males and to a common response of human societies to migration (cf. Divale [1984]) whereby part of the general process can be a shift to matrilocality. This implies that the ideology of matrilineality flowed from residence rather than vice-versa.

Figure 2 is a subgrouping of Nuclear Micronesian languages and Chuukic dialects which adapts materials from Quackenbush (1968), Jackson (1983) and Bender et al. (2003a; 2003b). It includes only the Chuukic dialects for which I have found clan names. Northern Marianas Carolinian is omitted because that community was established only during the historical period, and their clan names are now from everywhere through the Chuukic dialect chain. As that is the case, the Northern Marianas Carolinian clan names tell us nothing about distributions in prehistory. However, the Carolinian clan names are given in the data on each clan name because they add to our knowledge of vowel length and other aspects of the pronunciation of clan names.

I. Austronesian (An)
A. MalayoPolynesian (MP)
1. Central and Eastern MP (CEMP)
a. Eastern MP (EMP)
1’. Oceanic (Oc)
a’. Nuclear Micronesian and Nauruan
1’’. Nauruan (Nau)
2’’. Nuclear Micronesian (Mc)
a’’. Kosraean (Ksr)
b’’. Central Micronesian (CMc)
1’’’. Kiribati (Kir)
2’’’. Western Micronesian (WMc)
a’’’. Marshallese (Mrs)
b’’’. Pohnpeic-Chuukic (PC)
1’’’’. Pohnpeic (Pnp)
a’’’’. Pohnpeian (Pon)
b’’’’. Mokilese (Mok)
c’’’’. Pingelapese (Pin)
d’’’’. Ngatchickese (Sapwuahfik) (Ngk)
2’’’’. Chuukic (Ck)
a’’’’. Mortlocks (Mrt)
1’’’’’. Namoluk (Nam)
2’’’’’. Nama (Nama)
b’’’’. Chuuk Lagoon (Chk)
c’’’’. Northern Atolls
1’’’’’. Murillo (Mur)
2’’’’’. Unon (Uno)
d’’’’. Western Atolls
1’’’’’. Pulusuk (Hok, Hogh, Sok, Sogh) (Psk)
2’’’’’. Puluwat (Polowat)(Pul)
3’’’’’. Pullap (Pollap) (Pol)
e’’’’. Satawal (Sat)
f’’’’. Central Chuukic
1’’’’’. Lamotrek (Lam)
2’’’’’. Ifaluk (Ifa)
3’’’’’. Woleai (Wol)
4’’’’’. Faraulep (Far)
g’’’’. Western Chuukic
1’’’’’. Ulithi (Uli)
2’’’’’. Fais (Fai)
3’’’’’. Pulo Anna (Pur)

h’’’’. Unclassified
1’’’’’. Elle Carolinian (Crl)
2’’’’’. Enne Carolinian (Crn)

Figure 2: A subgrouping of Nuclear Micronesian languages[6]

[6] Adapted from Quackenbush (1968), Jackson (1983) and Bender et al. (2003a; 2003b).

The subgrouping in Figure 2 implies an early period in Nuclear Micronesian prehistory where Nauruan first became distinct from the rest of Nuclear Micronesian. Then Kosraean became distinct from Central Micronesian. Then Kiribati became distinct from Western Micronesian. Then Marshallese became distinct from Pohnpeic-Chuukic. Then Pohnpeic and Chuukic became distinct from each other. Finally, Pohnpeic became internally diverse and Chuukic became internally diverse.

The language distribution and migration interpretation that I favour is one where, but for Nauruan, the most ancient split occurred between dialects spoken on Pohnpei and Kosrae, the Pohnpeic variety of speech then dominating in the settlement of Kiribati, then the Marshalls and then Chuuk, all due to a larger population available to migrate from Pohnpei than Kosrae and, in the case of Chuukic, Pohnpei's closer proximity. The early sequence of splits: Kosraean, then Kiribatese then Marshallese from what was then left, Pohnpeic-Chuukic (Jackson 1983) is now the standard model (Bender et al. 2003a, b). The reasons for the assertion of this sequence of splits is based upon consistent but limited wisps and threads of comparative linguistic evidence in contrast to the large body of shared innovations apparent for Proto Micronesian and for Proto Pohnpeic-Chuukic which had longer periods of common development (cf. innovations reconstructed by Bender et al. 2003a, b). The protolanguages posited as intermediate between Proto Micronesian and Proto Pohnpeic-Chuukic were apparently post Proto Micronesian dialects not greatly transformed from Proto Micronesian itself. Those splits occurred rather quickly one after another, perhaps only two or four hundred years in all passing between the Kosraean split, the Kiribati split and the Marshallese split. The period of common and uniquely shared Pohnpeic-Chuukic innovations was perhaps 300–500 years and may, similar to Proto Micronesian's theoretical spread between Pohnpei and Kosrae, have occurred over an area including Chuuk as well as Pohnpei, Chuukic becoming distinct as contacts between people of the two areas became smaller relative to those in continuous residence on one island or the other. The lexicostatistical age of Chuukic internal diversity is about 1000 years (cf. Jackson 1983:Table 10).

This grand guess about Nuclear Micronesian homelands and dispersals is within the range of what archaeological evidence might presently be taken to support. Rainbird (2004) has recently summarised the archaeology of Micronesia and taken a different view. On the basis of the oldest dates that he considers reliable for the Nuclear Micronesian islands and groups, Rainbird (2004:100) suggests that the earliest Nuclear Micronesian speakers in eastern Micronesia:

> rapidly moved through the Gilberts and possibly reached the northern end of the Marshalls by 2500 to 2000 years ago. There may have been a brief hesitation here, perhaps over a couple generations, but not long enough to lose the knowledge of ceramics in these clay impoverished atolls, before the explorers located and seeded the eastern Carolines.

In any event, archaeological and linguistic interpretations agree that the greatest age of Nuclear Micronesian settlement was to the east or centre and that the development of uniquely Chuukic speech—and subsequently the emergence of that speech into the central and western Chuukic-speaking atolls—came later.

The settlement of Chuuk Lagoon is presently known to have begun by 2000 years ago or a few centuries earlier as is also the case for Kosrae and Pohnpei (Rainbird 2004:89–90). Rainbird (2004:168–179) gives the general history of Chuuk Lagoon archaeological research. King and Parker (1984) divide the history and prehistory of Chuuk into four periods:

- ca. 500 BC – AD 500 coastal sites with pottery
- ca. AD 500 – 1300 the 'Long Gap' where there is little evidence of human occupation
- ca. AD 1300 – 1900 coastal and inland sites with no pottery
- ca. 1900 – present colonial period and eventual independence

Rainbird (2004:171–172) interprets the Long Gap as being due to natural geomorphic processes which washed away the coastal sites inhabited during this period. Petersen (2004:85) notes that similar 'long gaps' occur for the same period of time in the archaeology of Kosrae and Pohnpei. Ignoring Rainbird's geomorphic explanation, Petersen attributes the end of the long gaps to the breadfruit revolution. In any event, the evidence of linguistics is clear: early Pohnpeic and Chuukic were separating languages by about the beginning of the Long Gaps and flourished on Pohnpei and Chuuk Lagoon respectively throughout the Long Gap period.

The oldest known archaeological remains on what are now the central and western Chuukic-speaking atolls presently come from two atolls nearest Yap: Fais[7] and Ulithi. Fais had settlements or other human use as early as 1900 years ago (Intoh 1996; Rainbird 2004:165–166) and Ulithi has cultural deposits as old as 1400 years (Craib 1981; Descantes 1998; Rainbird 2004:164–165). Ulithi and Fais are immediately east Yap, Ulithi being a large target about one day's sailing from Yap and Fais being a smaller target but only about 80 kilometres east of Ulithi. The archaeology of these islands connects them to Yap from the beginning. Pottery and pig, for instance, that came from Yap rather than Chuuk are found from the beginning on Fais. Thus the earliest occupations of Fais and perhaps Ulithi would not seem to be associated with Chuukic speakers.

Rainbird (2004:164) reviews just one radiocarbon date from what I here call the central Chuukic atolls—on a human bone from an excavation on Lamotrek dated to 780 years ago.

Summarising what is known for the archaeology of the apparent dispersal period of Chuukic language into the atolls, ca. AD 1000, is therefore short work: almost nothing is known for Chuuk Lagoon at the time or for the 500 years before or 300 years after and the only radiocarbon date from the central Chuukic atolls during those centuries is the Lamotrek date of 780 years ago.

I now turn to my sources and then methods in considering the history of the 27 Chuukic clan names of this study.

3 Ethnographic and dictionary sources

I first present a little background on the ethnography and sources for Chuukic's relatives to the east. Nauruans have been described as matrilocal and matrilineal. Wedgwood (1936:372) gives the names of 'the twelve' matrilineal clans of Nauru but none seem to be cognate with other Micronesian matriliclan names. Kiribati had no matrilineal or patrilineal clans in the historic period. The Marshallese were, in the main, matrilocal and matrilineal although certain very remote islands in the northwest of the chain constituted exceptions. Abo et al. (1976) (*Marshallese–English dictionary*) give 65 matrilineal Marshallese clan names but none seem to have cognates beyond the Marshalls. Kosraeans were matrilineal and the kings were crowned according to rules of matrilineal succession. *Südsee* (below) produced no listing of the Kosraean clan names and no other listings seem

[7] Fais is actually a small makatea (raised coral) island rather than an atoll.

to have been published. The Pohnpei sources employed here are Riesenberg (1968) and Rehg and Sohl (1979) which give the same 19 matrilineal clan names, Rehg and Sohl's forms being those cited below as they are in the contemporary orthography. *Südsee* (below) did not report on the Pohnpeian clan names but the Pohnpeian clans are rather few, large and stable and the same 19 probably occurred on Pohnpei during *Südsee* times.

During German Times in Micronesia (1898–1914) major ethnographic research was conducted between 1908 and 1910 and published in many volumes collectively called here the *Südsee* sources. They are the only ethnographic sources for clan names I have for many of the islands as will be mentioned presently.

The Chuuk Lagoon is matrilocal and matrilineal (cf. Goodenough [1978 (1951)]). The *Südsee* sources are very useful for Chuuk Lagoon as they give multiple localities for which clan names occur and are, therefore, a good indication of whether a clan name is or was widespread around the Lagoon. *Südsee* gives 40 or 50 Chuuk Lagoon clan names while in their *Trukese–English dictionary* Goodenough and Sugita (1980) give all the clan names found in *Südsee*, do so in the modern orthography and include a few dozen clan names not given in *Südsee*, 88 in all, a few of which may be post-*Südsee* introductions from the atolls. In general, the Chuuk Lagoon situation most resembles the Marshalls where scores of new clan names seem to have emerged locally. Nauru and Pohnpei are at the opposite extreme with, respectively, their 12 and 19 clans.

For the atolls, starting from the southeast and moving northwest and then west we begin with the Mortlock atolls (Namoluk, Nama, Lukonor and Losap) southeast of Chuuk Lagoon. These islands are all matrilocal and matrilineal. There are *Südsee* sources for the Mortlock atolls just mentioned except Losap and Marshall (1972, 2004) give more recent information on Namoluk kin and community.

For the atolls north of Chuuk Lagoon (Unon and Murillo), which are matrilocal and matrilineal, I have only *Südsee* sources.

Chuuk State's 'Western Islands' (Puluwat, Pollap and Pulusuk) are also matrilocal and matrilineal. There are *Südsee* sources for Puluwat and Pollap but neither *Südsee* nor other sources I am aware of describe the situation for Pulusuk.

Satawal is matrilocal and matrilineal. Clan names for Satawal were not published in the *Südsee* materials and I know of no other source that has done so in a comprehensive way.

West of Satawal are what I here call the 'central Chuukic atolls': Lamotrek, Elato, Olimarao, Faraulep, Ifaluk, Woleai and Eauripik. All are matrilocal and matrilineal. I do not include Satawal in this group because Satawalese is of a relatively distinct dialect type compared to those of the central Chuukic atolls and what little I can discover of the Satawal clan name situation (what I was told while living on Saipan and what I have seen on the Internet) suggests a distinctive situation as well. As seen in the list of abbreviations below (Table 4), Lamotrek, Woleai and Faraulep clan names are available from the *Südsee* sources and Ifaluk and Lamotrek clan names are available from post-war ethnographic sources. There is also the *Woleaian–English dictionary* (Sohn and Tawerilmang 1976) which gives a slightly different list of clan names than *Südsee* and is also an important source with respect to the exact pronunciation of Woleaian clan names.

The Northern Marianas Carolinians were originally, in the main, from Satawal and the central Chuukic dialect area, a later group from Chuuk State's northern atolls established a second Carolinian dialect, and post-war migrants from the atolls and Chuuk itself have, in general, assimilated to one or the other of the dialects and there is a dictionary of the two modern dialects (Jackson and Marck 1991). That dictionary is used here as a source for

knowledge of pronunciations but not for pre-European clan distributions (since the Carolinian clans arrived to Saipan and other islands of the Northern Marianas only from Spanish Times[8] onward[9]).

Then there are what I here call the 'western Chuukic atolls': Ulithi, Fais, Sorol and Ngulu in Yap State, Federated States of Micronesia and Sonsorol, Tobi, Pulo Anna and Merir in the Republic of Palau. Ulithi is still matrilineal and has at least four of the matrilineal clans (Lessa n.d.) of the central Chuukic atolls and islands further east. But the Ulithians are patrilocal (Lessa 1966:20, 22), a situation Hage and Marck (2002:152–153) would ascribe to the decline of long distance voyaging by Ulithian men. Yap is just a day away by sea but contacts with Yap islands to the east of Ulithi were maintained by central Chuukic atoll visitors to Ulithi and not by the Ulithians themselves (Lessa 1950). Lessa (n.d.) counts over 40 sibs or lineages but only four 'clans' (*hailang* < PCk **kainanga* 'clan, lineage)'. The four clan names are found on other islands (**kau-fanua*, **mwangau-ni-faca*, **tau-fanaa-ciki*, **tau-wene*), but only one of the sibs or lineages (**faa-ni-mai*).

We know of matrilineal clans on Fais and Pulo Anna from the *Südsee* sources. Two of the Pulo Anna clans and just one of the Fais clans have the same name as a clan from an island to the east.

The abbreviations used in the data materials for the atolls and those of Pohnpei and Chuuk Lagoon are given in the following figure.

Table 4: Sources and abbreviations

ChkDic	Chuuk Lagoon dictionary (Goodenough and Sugita 1980)
ChkGeo	Chuuk Lagoon geographical distribution according to ChkSrc
ChkSrc	Chuuk Lagoon source (Krämer 1932)
CrlDic	Elle Dialect Northern Marianas Carolinian (Jackson and Marck 1991)
CrnDic	Enne Dialect Northern Marianas Carolinian (Jackson and Marck 1991)
FaiSrc	Fais/Feis source (Krämer 1937) (western Chuukic atolls)
FarSrc	Faraulep source (Damm 1938) (central Chuukic atolls)
IfaSrc	Ifaluk source (Burrows and Spiro 1957) (central Chuukic atolls)
LamoSrc	Lamotrek source II (Alkire 1965) (central Chuukic atolls)
LamSrc	Lamotrek source (Krämer 1937) (central Chuukic atolls)
LukSrc	Lukonor source (Krämer 1935) (Mortlocks)
McEty	'Micronesian etymology'
MurSrc	Murillo source (Krämer 1935) (Chuuk State's northern atolls)
NamaSrc	Nama source (Krämer 1935) (Mortlocks)
NamoSrc	Namoluk source II (Marshall 1972; 2004) (Mortlocks)
NamSrc	Namoluk source (Krämer 1935) (Mortlocks)
OLMc	*Oceanic Linguistics* Proto Micronesian (Bender et al. 2003a; 2003b)
PCk	Proto Chuukic
PolSrc	Pullap/Pollap (Tamatam) source (Krämer 1935) (Chuuk State's 'Western Islands')
PonDic	Pohnpeian dictionary (Rehg and Sohl 1979)

[8] 'Spanish Times' – in Micronesian parlance, the period from Magellan's first landing on Guam (1521) to the end of the period of Spanish administration in 1898.

[9] There was central Carolinian voyaging to the Marianas before Spanish Times but here we consider when there were first settlements (with women and children) in the Northern Marianas that continued up to the present.

PulSrc	Puluwat/Polowat source (Damm 1935) (Chuuk State's 'Western Islands')
PurSrc	Pulo Anna source (Eilers 1935) (western Chuukic atolls)
UliSrc	Ulithi source (Lessa n.d.) (western Chuukic atolls)
UnoSrc	Unon source (Krämer 1935) (Chuuk State's northern atolls)
WolDic	Woleai dictionary (Sohn and Tawerilmang 1976) (central Chuukic atolls)
WolSrc	Woleai source (Krämer 1937) (central Chuukic atolls)

4 Data and reconstructions

Each of the 27 clan names reconstructed is given a separate subsection below. For purposes of presentation, the ancestral names are presented in their Proto Chuukic spellings, even if I believe the name may be of lesser age. The reconstruction is followed, where possible, by information on what morphemes the name is composed of and what those morphemes meant as common nouns, possessives or other parts of speech. Then the evidence is given from each locality for which a source has been identified. For example, the 'Land's Fishhook' discussion begins with:

***Kau-Fanúa 'Land's Fishhook'**

McEty OLMc PMc *kaú 'fishhook', *fanúa 'land'

ChkDic Kofénú
LamoSrc Hofalu
IflSrc Kovalú
WolSrc Gófalu
WolDic Géefalúwa

I now turn to the 27 reconstructions.

***Faa-ni-Mai 'Under the Breadfruit Tree'**

McEty OLMc *faa- 'under' + *ni 'of' + *mai 'breadfruit (tree)'

PonDic Dipwinpaanmei

ChkSrc Fánimei
ChkDic Fáánimey
ChkGeo Iras, Metitiu, Sabuk

NamSrc Fánimei
NamoSrc Fáánimey

MurSrc Fänemei
UnoSrc Fengimi (Ham.); Fánimei (Kram.)

PulSrc Fänimëi (dia.)[10]

UliSrc Fal Le Mei
PurSrc Féremäü (dia.)

CrlDic Fáálimáy

[10] '(dia.)' – indicates that the *Südsee* source used complex diacritics not included in the present report.

This clan name gives the appearance of being at least as old as Proto Chuukic. Otherwise we can observe that the clan name is not found in the central Chuukic atolls while all the other clan names of the western Chuukic atolls which have any outside cognates have at least one cognate in the central Chuukic atolls.

***Imwa-O 'House(?) of ___?'**

McEty OLMc PMc *imwa 'house'; no PMc word '**o' is presently reconstructed

ChkSrc Ivö / Imö (dia)
ChkSrc Imö oder Moánimö (dia)
ChkGeo Vidjap, Tunuk, Moen, Lúkula sam. Tol. Besitzer des Tóloman-Berges (Tóloman mountains)
ChkDic Imwó

PulSrc Umüa

Südsee notes that this is considered a recent clan on Puluwat as does Marshall (2004:37) in reference to Namoluk (and *Südsee* does not give this clan name for Namoluk or any atoll at all other than Puluwat, so I omit Namoluk data above). This clan name is therefore not well supported as a possible Proto Chuukic clan name.

***Kainanga-i-Liku 'Clan from Outside'**

McEty PMc OLMc *kainanga- 'clan' + PCk *-i- 'of (locative)' + PMC *-liku 'outside'

ChkDic Eyinangeyinúk

IflSrc Kailangailúk

As there are only two of the islands for which sources give this clan name there is the question of whether this simple name, 'clan from outside', developed in both places independently. But both reflect the PMc locative *i which is little used in the languages today and descends mainly from ancient constructions. So possibly the Chuuk and Ifaluk names are in fact related. Since this clan is not reported for Chuuk Lagoon in *Südsee* I am left wondering if this was first an Ifaluk clan name that came with an immigrant woman or women to Chuuk Lagoon between *Südsee* times and the time of the ChkDic work during the United Nations Trust Territory period.

***Kainanga-i-Sawa 'Clan by the Pass in the Reef'**

McEty PMc *kainanga 'clan' + *i (locative) + sawa 'pass in the reef'

ChkSrc Alegëitau (Elingeidau)
ChkGeo Iras, Metitiu
ChkDic Enengeyitaw

PolSrc Elangeitaf (dia.) (sic. ?)
PulSrc Ellangaidau

FarSrc Ellan'aidjau (dia.)

CrlDic Alengeitaw

The Pollap final consonant is unexpected. Unless confirmed by other historic or contemporary sources it seems a possible a typesetting error where ‘u’ or ‘w’ would be expected. Once again we see the **kainanga-i-* construction with good evidence of antiquity (the Chuuk Lagoon, Puluwat, Pollap and the much more distant Faraulep) consistent with (but not necessarily as old as) Proto Chuukic. Comparison to the previous clan name and the clan name that follows results in the observation that something unusual has happened to the *kai- of what I here posit as the first word of the name. If the most ancient pronunciation of this clan name was actually *kainanga-i-sawa, it would seem that before the name had spread beyond Chuuk Lagoon, the pronunciation had become *alanga-i-sawa. Perhaps it was, in fact, always pronounced *alanga-i-sawa and it doesn’t have anything to do with *kainanga ‘clan, lineage’. But Bender et al. (2003a, b) reconstruct no Proto Micronesian, Proto Chuukic or other word of the form **alanga and I presently interpret this as an idiosyncratic development out of the *kainanga-i- pattern.

***Kainanga-ni-Weneyaa ‘Clan of Woleai’**

McEty PMc OLMc *kainanga ‘clan’ + *ni (attributive) + Weneyaa ‘Woleai’

ChkDic Eyinangen Wéneya (eyinanga + ni + wéneya)

PulSrc Wonëi (‘von Oleai’)

LamSrc Gailang ali Oléa
LamoSrc Gailangúwoleai
WolSrc Gailang ali Voléa
WolDic Gailengaliweleyaa (gailangi + li + weleyaa)

This clan name is not mentioned for Chuuk Lagoon in *Südsee* and the *Südsee* Puluwat source notes the clan as coming from Woleai, which we would expect considering what the name means. So here we have excellent evidence that a clan originating in the central Chuukic atolls came to be present in Puluwat and the Chuuk Lagoon. Naming a (chiefly?) clan after its island is common around the Marshalls.

***Kasa-Maanga or Ka-Sama-anga**

McEty PMc OLMc unknown

NamoSrc Katamak

ChkSrc Kétevang/Ketemang/Kétemang
ChkGeo Moen

PulSrc Gadamän (dia.)
PolSrc Katamáng

LamoSrc Hatamang
WolDic Gatemaanga ‘name of a clan in Lamotrek and Satawal’

CrlDic Kkatamaam

This form is distributed in what seem to be regularly agreeing forms from Chuuk to Woleai and there are also the Namoluk and Carolinian forms with their irregular final consonant agreements. None of the anthropological sources indicate vowel length but both WolDic and CrlDic indicate that the third vowel was anciently long. *Südsee* gives the clan as occurring on Moen but the name is not mentioned in ChkDic.

Due to its antiquity or the origin of the name in uncommon words, it is not obvious what everyday words may have been the source of this clan name. There is no OLMc reconstruction of **kasa, **kasama, **sama(a)nga etc. PMc and PCk *sama meant 'outrigger' and *ka- was the causative. There was a PMc nominaliser *-anga which is seen very rarely but otherwise occurs in the important, ancient word *kai-n-anga 'matrilineal clan or lineage'. The only suggestion I have as to what this word may have once been is 'people who make / facilitate the outrigger' or something of the sort. The causative, *ka-, would be seen to have made a verb out of *sama 'outrigger' and then *-anga would have been added as a new nominal meaning developed.

Because the clan seems absent from the central and western Chuukic atolls other than Lamotrek and not well distributed around the Chuuk Lagoon, there is the appearance that the clan may have had its origin in Chuuk State's Western Islands or Satawal.

***Kaú-Fanúa 'Land's Fishhook'**

McEty OLMc PMc *kaú 'fishhook', *fanúa 'land'

ChkDic Kofénú

LamoSrc Hofalu
IflSrc Kovalú
WolSrc Gófalu
WolDic Gééfalúwa

UliSrc Hofalu

CrlDic Ghéfalúw

This clan name is apparently composed of the ancient forms *kaú 'fishhook' and *fanúa 'land'. WolDic notes that aside from meaning 'fishhook', Woleaian *géé* also means 'formal system of political ties which exist between islands and which is symbolised by exchanges of food and goods'—the *sawei* system (Lessa 1950) or aspects of it, apparently.

The first vowel seems to have shortened everywhere but Wol which does not have short '*é*'.

Such shortening is rather rare in Lagoon Chuukese but ChkDic shows a short vowel. So one is left wondering if the clan name was borrowed by Chuukese from the west. As the Chuukese dictionary contains the form but the *Südsee* sources do not, borrowing since *Südsee* times seems a possibility. It is not, after all, otherwise reported from the Mortlocks, Chuuk Lagoon or Chuuk State's Western Islands. Like *Kainanga-ni-Weneyaa, I take this clan name to be one that originated in what are now Yap State atolls.

***Luuka-(ni)-Fanua 'Centre of the Land'**

McEty OLMc PMc *luuka 'center, middle', *ni 'of', *fanúa 'island, land'

PonDic Dipwiluuk

ChkDic Sowunuuk
ChkDic Nuukan
ChkDic Nuukanap
ChkDic Nuukenfénú

MurSrc Lugonofánu, Lugefanú (Ham.)

WolDic Luugofalúwa

Note that both MurSrc and WolDic have the unusual change where the first *-a- became *-o-* due to the vowel(s) *before* it rather than *after* it, adjustment to *following* vowels being the dominant pattern, when change occurs, in Nuclear Micronesian languages.

PonDic *Dipwiluuk* contains the *luuka 'center, middle' word. But one might expect such a common meaning to be adopted into a clan name at different times in different places and not always to have come from just one ancestral clan name. Comparison of the fourth ChkDic and MurSrc raises the question of whether *-na- existed at an ancestral level because Woleaian does not have it. This is perhaps another reason to wonder if the name developed independently at different times in different places.

No form of the clan name occurs in the Chuuk Lagoon *Südsee* sources. On the whole, this seems a candidate for independent developments.

***Mwangau-ni-Faca 'Eaters of Pandanus'**

McEty OLMc *mwangau 'eat' - *ni 'of, at' - *faca 'pandanus'

ChkSrc Mongülufadj 'von Lámotrek oder Póloat stammend'
('of Lamotrek or Puluwat origin')
ChkGeo Tol
ChkDic Méngúnúfach

PulSrc Mongölifäl (dia.)
PolSrc Mangaulifadj 'von Póloat stammend' ('of Puluwat origin')

LamoSrc Móngalïfach
IflSrc Mangaulevár
WolSrc Mongaulifadj
WolDic Mengaulifasha
FarSrc Mon'ólifer (dia.)

UliSrc Mongolfach

CrlDic Mwongoulufasch, Mwongounufarh

In its *Méngúnúfach* 'clan name' entry ChkDic mentions *méngú* 'pandanus leaf' and references the form to *OLMc*[11] PMc *mangu 'pandanus leaf'. But certainly the name was borrowed into the Puluwat area out of the central Chuukic atolls and came to be pronounced differently/incorrectly in the Chuuk lagoon area when the clan became known there.

The suggestion of borrowing from west to east is required not only because the 'fullest' form is found to the west, but because sources (in data listing above) to the east mention knowledge of the clan coming from the west and because the clan is very rare around the lagoon and not reported at all in the Mortlocks, Unon or Murillo. So this is another clan name that we posit to have originated in the atolls of what is now Yap State.

***Mwaoco 'Ashes'**

McEty OLMc PCk *maoco 'coals, ashes'

ChkSrc Módj/Mot
ChkGeo Iras, Vinepis, Lugan
ChkDic Mwóóch

[11] Which was under development at the time of ChkDic's publication.

ChkDic Sowumwóóch

PolSrc Modj

CrlDic Kkemwóósch
CrlDic Mwóósch

The sound correspondences between Chk, Pol and Crl are regular but since the clan is not distributed much beyond Chuuk Lagoon, this seems a possible recent extention of the clan out of Chuuk Lagoon to Pollap or vice versa.

***Pike 'Sand Islet'**

McEty OLMc PMc *pike 'sand islet'

ChkDic Piik

UnoSrc Pig (Ham.); Bik (Kram.)

PolSrc Pïk

CrlDic Piigh

The final vowel is known from the 'sand islet' reconstruction. West-southwest of Unon and west-northwest of Pullap is the islet know as Piik or Pikelot. Possibly this clan name originated there at a time when the islet was inhabited (it has not been inhabited through the historic period due to its small size).

***Pwalú 'Taro Swamp'**

McEty: OLMc PCMc *pwalú 'taro swamp, soil as found in taro swamp'

ChkSrc Boën (dia.)
ChkGeo Sabuk
ChkDic Pwéén

PulSrc Böl
PolSrc Boál, Böl

LamSrc Boll
IflSrc Bwél
WolDic Béélú

CrlDic Bweel

The vowels of the CrlDic form are unexpected. One would expect 'éé' but there is also *bweel* 'taro swamp' so the irregularity occurs in both the proper and common nouns.

With cognate/regularly corresponding forms from Chuuk Lagoon, Chuuk State's Western Islands and the central Chuukic atolls, this form can probably be attributed to Proto Chuukic, but we should remain cautious in light of the limited ChkGeo distribution.

***Raki 'Sailing Season'**

McEty OLMc PMc *raki 'year, sailing season'

ChkSrc Rak / Rek / Rëg
ChkGeo Sabuk, Tol
ChkDic Ráák

LamSrc Rak
LamoSrc Rakh

PurSrc Rékiri Saralé(sa) (dia.)

CrlDic Ráágh

With regularly agreeing forms in the Chuuk Lagoon and Lamotrek and a possible cognate in Pulo Anna, this form seems a possible Proto Chuukic candidate.

***Rape-Fanúa '? Land/Island'**

McEty OLMc PMc *fanua 'land, island'

ChkDic Ropéfénú
IflSrc Rapevelú

This clan name is composed of PMc *fanua 'land' added to a second word, **rope or **rape, for which no ancient meaning is presently reconstructed. The IflSrc mentions that *rape* is an Ifaluk word meaning 'chief' but not 'high(est) chief'.

The clan is reported only for Ifaluk and for Chuuk Lagoon in ChkDic but not from *Südsee* times. Perhaps this distribution involved an Ifaluk woman who married into Chuuk lagoon some time after the *Südsee* materials were collected.

***Talasi 'Loosen'**

McEty OLMc PMc *talasi 'loosen something'

ChkDic Sanet

MurSrc Salet; Zelet (Ham.)

As this clan name is found only in Murillo and, more recently, in the Chuuk Lagoon, it would seem possibly to be a Murillo clan that came to Chuuk after *Südsee* times.

***Tapwo-ni-Ppia 'Village at the Beach'**

McEty OLMc PMc *tapwo 'district, village'—*ni 'of'—*pipia 'sand, sand beach'

ChkSrc Sabúnubi
ChkGeo Moén, Mëtitiu, Iras, Lëaua, Vídjap, Vinepis
ChkDic Sópwunupi

NamSrc Sabúnubi
NamoSrc Sópwunupi; plus four Südsee spellings not entered.
NamaSrc Thau a bunn

PolSrc Haubúnubit - source notes Chuuk Lagoon Sabúnubi
PulSrc Haubónibi (dia.) ('Ausgestorben ist die ainang Haubónibi'—('died out')

CrlDic Sabwoloppi
CrnDic Habwonoppi

Due to its wide distribution around the Chuuk Lagoon (ChkGeo) and its presence in the Mortlocks and Chuuk State's Western Islands, there is the appearance that this clan name is old and well established. Perhaps it was a Proto Chuukic clan that did not participate in the settlement of the central and western Chuukic atolls. It is also possible that it was part

of the dispersal to Satawal and further west but died out in those areas as the PulSrc says it had on Puluwat (by *Südsee* times).

The final consonant of the PolSrc may be a *Südsee* typo. A suspicious consonant ends the previously listed clan name as well.

Here again a vowel has changed due to the vowel *before* it rather than the vowel *after* it (NamSrc, NamoSrc, ChkDic, ChkSrc, PolSrc, CrlDic, CrnDic). Data from all relevant witnesses other than PulSrc suggest an intermediate *sapwo-nu-ppi (where *-ni- is expected rather than *-nu-).

***Taro 'Birthmark'**

McEty ChkDic *sóór*, *sóro-*, PulDic[12] *hór* (sic – one would expect *hóór*),
CrlDic *sóór* 'birthmark'

ChkDic Sóór
ChkSrc Sorr
ChkGeo Moén, Lëaua, Vídjap, Sãbuk, Peliëséle, (Pélia Kub), Pada

NamSrc Sorr
NamoSrc Sór
NamaSrc Sorr

MurSrc Sorr; Tsor (Sor) (Ham.)

PolSrc Sorr
PulSrc Sär (auch Här)

LamSrc Sorr
LamoSrc Saur
WolSrc Sor

CrlDic Sóór

This is a very widely distributed and apparently ancient clan name within Chuukic. While WolSrc lists this as a clan name WolDic does not. Possibly the clan died out between the times of the two works.

***Tawu-Afangi 'People of the North / Winter Tradewind'**

McEty OLMc PMc *tawu 'master, expert', PMc *tau 'people',
PMc *afangi 'north, winter tradewind season'

ChkSrc Sáuefang
ChkGeo Moen, Fefan, Udot, Lekutanufidj, Sabbou
ChkDic Sowuwefeng

MurSrc Sáuefang (dia.); Zauupang, Zóufang (Ham.)

PulSrc Sauúfan (dia.)

WolDic Sauwefangi 'name of a clan in Outer Islands (including Woleai)'

FaiSrc Matalíefang (*mata 'face' + *ni 'of' + *afangi 'north, winter tradewind')

CrlDic Sóufáng

---

12 Elbert (1972).

ChkDic notes that this is the ‘local name for the Sowupwonowót clan formerly on Romónum Island’ but *Südsee* gives a wider distribution and a status independent of any other clan name. It is curious that the name only occurs on Woleai and not any of the other central or western Chuukic atolls (except perhaps on Fais where an imperfect agreement is found). It is also curious that WolDic relates that it is a clan of the Outer Islands (of Yap State) and Woleai itself when *Südsee* did not list this as a clan of Woleai and neither *Südsee* nor later ethnographic sources mention the clan in any of their lists of central and western Chuukic atoll clan names.

***Tawu-Alai ‘Tall People’**

McEty OLMc PMc *tawu ‘master, expert’, PMc *tau ‘people’, *ala(i,ú) ‘tall, long’

ChkDic Sowuyáney
ChkSrc Sáuanei, Souëlei
ChkGeo Meititiu, Vidjap, Pellia, auch Udot, Pollap

PolSrc Soualei, auch Souëlei
PulSrc Sauwälei (dia.)

CrlDic Sóuwaley

While this form is well distributed around Chuuk Lagoon, it is otherwise reported only for the Western Islands (Pol, Pul) and seems not to have participated in the dispersals to the central and western Chuukic atolls.

***Tawu-Fanaa-Ciki ‘Small Needlefish People’**

McEty OLMc PMc *tawu ‘master, expert’, PMc *tau ‘people’,
PCMc *fanaa ‘a kind of needlefish’, PMc *ciki ‘small, little’

ChkDic Sowufanachik

LamSrc Saufaládjik
LamoSrc Saufalacheg
IflSrc Sauvelárik
FarSrc Saufaläs’ik (dia.).
WolDic Saufelaashigi
WolSrc Saufaládjik

UliSrc Soflachikh

CrlDic Sóufelááschigh
CrnDic Sóufenaarhig

There are two reasons to believe that this clan name originated in the central or central and western Chuukic atolls and was borrowed into Chuuk Lagoon, presumably with the arrival of a central or western Chuukic atoll woman after Südsee times. The first involves the sounds of the Chuukic clan name. If *tawu-fanaa-ciki was a name from around the Chuuk Lagoon which the Lagoon Chuukese always had or had from an early time, we would expect the modern Chuuk Lagoon pronunciation to be **sowufánááchik but instead we find ‘-a-’ instead of ‘-á-’ and the second one is short rather than long. So the name seems borrowing into Chuuk Lagoon speech. The second reason borrowing from the central or western Chuukic atolls is suggested is distributional: the name is found only there except in Chuuk Lagoon (ChkDic) but was not reported from Chuuk Lagoon during

Südsee times and has not been reported from Chuuk State's Western Islands. Finally, the origin of the name is said by Burrows (1957) to mean 'People of Falárik', Falárik being a village of Ifaluk. So possibly the clan name originated on Ifaluk or the village took the name of the clan of some of the people who lived there.

***Tawu-Pelaya 'People of - ?'**

McEty OLMc PMC *tawu 'people', PWMc *palia 'side'

(PonDic Sounpelienpil)

ChkDic Sowupinay

PolSrc Soubelëi, Hóupílei

CrlDic Soupeley

This is a clan name with a very limited distribution, problems with the agreement of sounds between the forms identified and no obvious source of the second word in the compound. It resembles a Pohnpeian clan name but only vaguely—the vowels beyond those coming from **tawu-* do not actually agree. It is reported in ChkDic but not in *Südsee*. The CrlDic form most resembles the first PolSrc form while the ChkDic form most resembles the second PolSrc form. On balance, there are no good grounds for attributing any great antiquity to this clan name.

***Tawu-Pwolowasa 'People of Puluwat'**

ChkDic Sowupwonowót

PulSrc Saupölowat

WolDic Saubwulowata 'name of a legendary chief who lived on an island in the eastern part of the Caroline Islands'

CrlDic Sóubwolowat

This clan was not reported for Chuuk Lagoon from *Südsee* times and is found only on Puluwat and in the Chuuk Lagoon Its name suggests that it originated on Puluwat and later, perhaps since the *Südsee* fieldwork, came to be established in Chuuk Lagoon.

***Tawu-Wene 'Upright People'**

McEty OLMc PMc *tawu 'master, expert', PMc *tau 'people', PCk *wene 'straight, upright, honest, true'

ChkDic Sowuwen, Sowuwén

NamoSrc Souwon

LamSrc Sauel
LamoSrc Sauwel
FarSrc Sauwöl
IflSrc Sauwél
WolSrc Sauel
WolDic Sauwele

UliSrc Sawol

CrlDic Sóuwel

Proto Chuukic *n rather than *l is reconstructed on the basis of NamoSrc. None of the other languages above still distinguish the *l and *n sounds of Proto Chuukic. But the clan is not mentioned by *Südsee* for Namoluk, any of the other Mortlocks or Chuuk Lagoon so it is possible that the clan came, since *Südsee* times, to Chuuk then Namoluk from, ultimately, the central or western Chuukic atolls. So this is another candidate for a clan name that originated in the central and western Chuukic atolls and came to Chuuk Lagoon after *Südsee* times.

If that were the case and the clan subsequently became established on Namoluk, we lose our reason for suggesting that Proto Chuukic *n rather than *l was involved since only *n* would have been present in post- *Südsee* Lagoon Chuukic speech and it would have been introduced to Namoluk with an *n* even if the earliest pronunciation was with an *l*. In any event, there is no early or Proto Chuukic *wele reconstructed while we do have a secure reconstruction in Proto Chuukic *wene ‘straight, upright, true’.

***Wao-ni-Rae ‘On the Branch’**

McEty OLMc PCk *wao- ‘on’, *-ni ‘of’, *rae ‘branch, stick’

ChkDic Wóónirá

PulSrc Wonirä

CrlDic Weliirá

This clan name is not mentioned in the *Südsee* sources for Chuuk Lagoon and does not extend beyond Chuuk except for on Puluwat. The ChkDic and PulSrc spellings suggest PCk *wawo-ni-rae which would mean something like ‘on (a) branch (of a tree)’. The CrlDic spelling shows an unexpectedly short first vowel and an unexpectedly long second vowel. The clan name shows no evidence of having been present in the central or western Chuukic atolls.

***Wii-sakaú ‘(People of) Namonuito’**

McEty OLMc PMc *sakau ‘reef, reef island’, ChkDic Nómwun Wiité ‘Namonuito Atoll (excluding Ulul Island)’

ChkSrc Vida
ChkGeo Sabuk, Tunuk, Mëtitiu, Iras, Moen, Vidjap, Lëseráda
ChkDic Wiitéé

MurSrc Uida (dia.); Uiza (Ham.)

PulSrc Wuïdo

CrlDic Wiité

This clan seems never to have become established in the central and western Chuukic atolls. The clan name is the same as the adjectival part of the name of the island of Namonuito atoll, is reported as a clan name by *Südsee* for Puluwat and numerous localities in Chuuk Lagoon, but is not reported at *Südsee* as a clan name on Namonuito itself. Perhaps *wii-sakaú had some special meaning in early Chuukic that came to be a clan name around Chuuk Lagoon and an atoll name to its north. No special meaning for PCk *wii- is reconstructed in the OLMc works. In Lagoon Chuukese *wii-* refers to being on the top of things. We see something like it again at the beginning of the next clan name below.

***Wi-Tuutuu**

McEty OLMc PCk *Tuu ‘to be open(ed)’; ChkDic *wi-* ‘pulled up, extracted’, *wiiwi-* ‘uprooted’, *wi-* ‘situation, locus, manner’, *wii$_1$-* ‘replaced, succeeded, exchanged’, *wii$_2$-* ‘arrange, plan’, *wii$_3$-*, *wi-* ‘(at) the top (of)’; WolDic *wiiwii* ‘to pull, uproot, extract’; CrlDic *wi-* ‘to be fat’, *wi* ‘to uproot, remove or place’

ChkSrc Uisüsu
ChkGeo Iras, Metitiu, Tunuk, Leaua, Sabuk, Lúkula, Tol
ChkDic Wisuusu

MurSrc Uisusu; Uitsutsu (Ham.)

PolSrc Uisúsu

LamSrc Uisúsu

CrlDic Wisusu

The *T sound of PCk was rather rare and the agreements for it are consistently regular through the languages that have this clan name so it seems likely that the main word in this clan name had to do with things that are ‘open(ed)’. But there is no *wi(i)- reconstructed for PCk in OLMc and the forms from ChkDic, PulDic, WolDic and CrlDic that I show above give us no one meaning to combine with PCk*TuuTuu (which may have meant, most precisely, ‘[be] opening’) that would result in a meaning which strikes me as having to do with people and their situations (‘[People] Under the Breadfruit Tree’, ‘Village by the Beach’, ‘People of the Northern/Winter Tradewind’ and so on).

The distribution suggests a clan spread widely in the Chuuk Lagoon clan that did not spread widely into the central and western Chuukic atolls. It is a bit of a curiosity that it occurs on Lamotrek but is reported from none of the other central or western Chuukic atolls. LamoSrc makes no mention of the clan being small or recent.

***(Wu)waa-ni-kara ‘Sweet Canoe (?), Burnt Canoe (?)’**

McEty OLMc PMc *waxa ‘canoe’, *-ni ‘of’, *kara ‘burnt’, PCk *waa ‘canoe’, *kara$_1$ ‘burnt’, *kara$_2$ ‘sweet’; ChkDic *wu$_1$-* ‘male person’, *wu$_3$-* ‘pierce’, *wu$_4$-* ‘throw water on something’, *wu$_5$-* ‘fish trap (in compounds)

ChkSrc Uenakar, Valigar
ChkGeo Peliëséle
ChkDic Wuwáánikar

NamSrc Venegar
NamoSrc Wáánikar
NamaSrc Uannigar

CrlDic Wwaleghár

This distribution suggests a clan originating in Chuuk Lagoon or the Mortlocks that never extended into the central and western Chuukic atolls. The meanings suggested, ‘sweet canoe’ or ‘burnt canoe’, are just the most obvious guesses. We have, in both the ChkDic and CrlDic spellings, evidence that there was an additional syllable at the beginning of the word—something like ‘*wu-’—so perhaps the original meaning didn’t have to do with canoes at all.

5 Distributions, categorisation and conclusions

The first group of clan names I shall discuss are those which are candidates for Proto Chuukic age:

*Faa-ni-Mai ‘Under the Breadfruit Tree’
*Kainanga-i-Sawa ‘Clan by the Pass in the Reef’
*Ka-Sama-nga ‘To Make the Outrigger’
*Pwalú ‘Taro Swamp’
*Raki ‘Sailing Season’
*Taro ‘Birthmark’
*Tawu-Afangi ‘People of the North / Winter Tradewind’
*Wi-Tuutuu ‘Opened (?)’

There are no phonological or distributional reasons to suggest that these are anything but Proto Chuukic clan names although the rarity of the final two in the central and western atolls leaves open the possibility that they arrived later and in small numbers to the few atolls involved. But we cannot distinguish this possibility from that which would have them amongst the earliest names but less widely distributed than the others nor from a possibility that those clan names were once spread more widely in the centre and west, only to have died out in one or more localities.

The second group I would nominate as old but perhaps not as old as Proto Chuukic or perhaps Proto Chuukic clans which never spread west beyond Chuuk State’s Western Islands:

*Tapwo-ni-Ppia ‘Village at the Beach’ (Chuuk Lagoon, Namoluk, Puluwat, Pollap)
*Tawu-Alai ‘Tall People’ (Chuuk Lagoon, Puluwat, Pollap)
*Tawu-Pelaya ‘People of ?’ (Chuuk Lagoon, Pollap, (?) Pohnpei)

A third group of eight clan names seem to have had their origins in the central and western atolls. These distributions suggest that old clans from the Proto Chuukic times dispersal to the centre and the west sometimes took new names in the centre and west. We can imagine that they might have done this upon arriving to a new place or due to some other event in their history. We can imagine that they might have done this when a clan became large and one or more of its lines came to be called by distinct names, eventually becoming clans in their own right. The clan names that I assign to this group are:

*Kainanga-ni-Weneyaa ‘Clan of Woleai’ (Woleai origin)
*Kainanga-i-Liku ‘Clan from Outside’ (Ifaluk origin)
*Kaú-Fanúa ‘Land’s Fishhook’ (central Chuukic atoll origin)
*Luuka-(ni)-Fanúa ‘Centre of the Island’ (Murillo, Woleai)
*Mwangau-ni-Faca ‘Eaters of Pandanus’ (central Chuukic atoll origin)
*Rape-Fanúa ‘? Land/Island’ (Ifaluk origin)
*Tawu-Fanaa-Ciki ‘Little Needle Fish People / People of Fanaa-cik’ (Falarik ‘Little Needlefish’ island origin [central Chuukic atolls])
*Tawu-Wene ‘Upright People’ (central and western Chuukic atoll origin)

Then there are five clan names that may have their origins on atolls other than those I here call the western and central Chuukic speaking atolls:

*Pike '(People) of Pike Island' (Unon, Pollap)
*Talasi 'Loosen (?)' (Murillo origin)
*Tawu-Pwolowasa 'People of Puluwat' (Puluwat origin)
*Wao-ni-Rae 'On the Branch' (Puluwat origin)
*Wii-Sakaú '(People of) Namonuito' (Namonuito origin)

There is at least one clan said by the sources to be a recent introduction to certain atolls from Chuuk Lagoon and two more with similar sorts of distributions:

*Imwa-O 'House-?' (Chuuk Lagoon origin, recent arrival to Namoluk and Puluwat)
*Mwaoco 'Ashes' (Chuuk Lagoon origin, recent (?) arrival to Pollap)
*(Wu)-waa-ni-Kara 'Sweet Canoe (?), Burnt Canoe (?)' (Chuuk Lagoon, Nama and Namoluk)

So this study ends with about ten clan names that give the appearance of being oldest and perhaps as old as Proto Chuukic, a number similar to Nauru's 12 clans and not much fewer than the 19 presently populating Pohnpei (taking into consideration that modern Pohnpei has had about 1000 years longer than Proto Chuukic for new clans to arrive from elsewhere or to emerge by a splitting of an existing clan). After the settlement of the central and western Chuukic-speaking atolls, eight new clan names seem to have emerged in the centre and west. Some new names, we must imagine, emerged by changes in clan names and some new names, we must imagine, came from some clans splitting into two over time. Other of our 27 clan names seem never to have been present in the central and western Chuukic atolls. The geographical proximity of these Chuuk State atoll clans to Chuuk Lagoon leaves us, in most instances, with a very uncertain picture of their antiquity.

The approximately ten clan names which seem most clearly to be candidates for Proto Chuukic age show that clan names can and do survive for a thousand years or more. But comparison to Pohnpeian suggests clan names have more difficulty surviving periods approaching two thousand years. But for two or three exceptions, the original clan names shared between the Pohnpei and Chuuk populations have died out or changed in Pohnpei, Chuuk or both.

The eight clan names which appear to have arisen in the central Chuukic atolls suggest there may have been change upon immigration from what is now Chuuk State but we cannot distinguish this, by any method but oral histories, from cases of bifurcation or, for instance, that one of more have their origin in an entirely different population such as Yap. Indeed, we cannot distinguish central Chuukic atoll origin clan names from Proto Chuukic clan names that died out in Chuuk Lagoon or simply came, there, to be called by a different name.

Of the eight clan names with an apparent central/western Chuukic atoll origin, one has a name suggestive of a sparse atoll environment. Does 'Eaters of Pandanus' come from a time before the breadfruit revolution? Is it the chiefly clan on Lamotrek because it is the oldest clan that hasn't died out. In fact it was near extinction at the time of Alkire's (1965) research and its highest ranking female was chief of the island as she had no brothers or mother's-sister's-sons to take the position. Were the Eaters of Pandanus part of an early central Chuukic atoll population that were maintaining way-stations on the route to Yap

more than they were a stable population in the sense we have seen in the historic period, a lifeway which had changed for the better after the breadfruit revolution?

I found nothing to shed light upon the '*Achaw*' period (Goodenough 1986). None of the 27 clans names are clearly intrusive to the Chuukic language area and the *Achaw* clan names that I know from Saipan are apparently from Satawal and/or Pulusuk and do not show up generally around the Lagoon or elsewhere.

References

Abo, Takaji, Byron W. Bender, Alfred Capelle and Tony Debrum. 1976. *Marshallese–English dictionary*. Honolulu: University of Hawaii Press.

Alkire, William H. 1965. *Lamotrek Atoll: inter-island economic ties*. Prospect Hts. IL: Waveland Press.

Bender, Byron, Ward Goodenough, Frederick Jackson, Jeff Marck, Kenneth Rehg, Ho-min Sohn, Stephen Trussel and Judith Wang. 2003a. Proto Micronesian reconstructions: 1. *Oceanic Linguistics* 42(1):1–110.

—— 2003b. Proto Micronesian reconstructions: 2. *Oceanic Linguistics* 42(2):271–358.

Burrows, Edwin Grant and Melford E. Spiro. 1957. *An atoll culture: ethnography of Ifaluk in the Central Carolines*. New Haven CT: Human Relations Area Files Press.

Craib, John L. 1981. Settlement on Ulithi Atoll, Western Caroline Islands. *Asian Perspectives* 24:47–55.

Damas, David. 1979. Double descent in the eastern Carolinas. *Journal of the Polynesian Society* 88:177–198.

—— 1981. The Keinek of Pingelap and patrilineal descent. *Journal of the Polynesian Society* 90(1):117–122.

Damm, Hans. 1935. Inseln um Truk: Polowat, Hok und Satowal. vol. II, *Ethnographie: B.Mikronesien*, Band 6, 2. Halbband. Hamburg: Friederichsen, De Gruyter and Co.

—— 1938. Zentralkarolinen: Ifaluk, Aurepik, Faraulip, Sorol, Mogemog. vol. II, *Ethnographie: B.Mikronesien*, Band 10, 2. Halbband. Hamburg: Friederichsen, De Gruyter and Co.

Descantes, Christophe. 1998. Integrating archaeology and ethnohistory: the development of exchange between Yap and Ulithi, Western Caroline Islands. PhD, University of Oregon.

Divale, William. 1984. *Matrilocal residence in pre-literate society*. Ann Arbor: UMI Research Press.

Eilers, Anneliese. 1935. Westkarolinen: Songosor, Pur, Merir. vol. II, *Ethnographie: B.Mikronesien*, Band 9, 1. Halbband. Hamburg: Friederichsen, De Gruyter and Co.

Elbert, Samuel H. 1972. *Puluwat dictionary*. Canberra: Pacific Linguistics.

Goodenough, Ward H. 1978 (1951). *Property, kin and community on Truk*. Hamden, CT: Archon Books.

—— 1986. Sky world and this world: the place of Kachaw in Micronesian cosmology. *American Anthropologist* 88:551–568.

Goodenough, Ward, and Hiroshi Sugita. 1980. *Trukese-English dictionary*. Philadelphia: American Philosophical Society.

Hage, Per. 1998. Was Proto-Oceanic society matrilineal? *Journal of the Polynesian Society* 107(1):365–379.

Hage, Per and Jeff Marck. 2002. Proto-Micronesian kin terms, descent groups, and overnight voyaging. *Oceanic Linguistics* 41(1):159–170.

—— 2003. Matrilineality and the Melanesian origin of Polynesian Y-chromosomes. *Current Anthropology* 44(S):S121–S127.

Intoh, Michiko. 1996. Multi-regional contacts of prehistoric Fais islanders in Micronesia. *Bulletin of the Indo-Pacific Prehistory Association* 15:111–117.

Jackson, Frederick H. 1983. On the internal and external relationships of the Trukic languages of Micronesia. PhD, University of Hawaii.

—— 1984. Reflexes of Proto-Oceanic in the Trukic languages in Micronesia. In Byron W. Bender, ed. *Studies in Micronesian Linguistics*, 259–280.

Jackson, Frederick H. and Jeffrey C. Marck. 1991. *Carolinian–English dictionary*. Honolulu: University of Hawai'i Press.

Kayser, Manfred, Silke Brauer, Richard Cordaux, Amanda Casto, Oscar Lao, Lev A. Zhivotovsky, Claire Moyse-Faurie, Robb B. Rutledge, Wulf Schiefenhoevel, David Gil, Alice A. Lin, Peter A. Underhill, Peter J. Oefner, Ronald J. Trent, and Mark Stoneking. 2006. Melanesian and Asian origins of Polynesians: mtDNA and Y-chromosome gradients across the Pacific. *Molecular Biology and Evolution* 23:2234–2244.

King, Thomas F., and Patricia L. Parker. 1984. *Pisekin Nóómw Nóón Tonaachaw*. Carbondale: Southern Illinois University at Carbondale Center for Archaeological Investigations.

Krämer, Augustin. 1932. Truk. vol. II, *Ethnographie: B. Mikronesien*, Band 5. Hamburg: Friederichsen, De Gruyter and Co.

—— 1935. Inseln um Truk: Lukunór-Inseln und Námoluk; Lósap und Nama; Lëmárafat, Namonuito oder Onóun; Pollap-Támatam. vol. II, *Ethnographie: B.Mikronesien*, Band 6, 1. Halbband. Hamburg: Friederichsen, De Gruyter and Co.

—— 1937. Zentralkarolinen: Lamotrek-Gruppe, Oleai, Feis. vol. II, *Ethnographie: B.Mikronesien*, Band 10, 1. Halbband. Hamburg: Friederichsen, De Gruyter and Co.

Lessa, William A. 1950. Ulithi and the outer native world. *American Anthropologist* 52(1):27–52.

—— 1966. *Ulithi: a Micronesian design for living*. New York: Holt, Reinhart and Winston.

—— n.d. Papers of William Armand Lessa (field notes). Washington DC: National Anthropological Archives, Smithsonian Institution.

Marck, Jeff. 2000. *Topics in Polynesian language and culture history*. Canberra: Pacific Linguistics.

—— 2008. Proto Oceanic society was matrilineal. *Journal of the Polynesian Society* 117(4):345–382.

Marshall, Mac. 1972. *The structure of solidarity and alliance on Namoluk Atoll*. University of Washington.

—— 2004. *Namoluk beyond the Reef*. Boulder: Westview Press.

Petersen, Glenn. 2006. Micronesia's breadfruit revolution and the evolution of a culture area. *Archaeology in Oceania* 41(2):82–92.

Quackenbush, Edward M. 1968. From Sonsorol to Truk: a dialect chain. PhD thesis, University of Michigan.

Rainbird, Paul. 2004. *Prehistory of Micronesia*. Cambridge: Cambridge University Press.

Rehg, Kenneth L. and Damian G. Sohl. 1979. *Ponapean–English dictionary*. Honolulu: University Press of Hawaii.

Riesenberg, Saul. 1968. *The native polity of Ponape*. Washington DC: Smithsonian Institution Press.

Schneider, David M. 1980. Is there really double descent in Pingelap? On Damas 1979 (vol. 88:177–198). *Journal of the Polynesian Society* 89(4):525–528.

Sohn, Ho-Min, and Anthony F. Tawerilmang. 1976. *Woleaian–English dictionary*. Honolulu: University of Hawai'i Press.

Wedgwood, Camilla H. 1936. *Report on research work in Nauru Island, Central Pacific*: Part 1. Oceania 6(4):359–391.

28 *Kinship terms in Bungku-Tolaki languages: inheritance, innovation and borrowing*

DAVID MEAD

If I have seen a little further it is by standing on the shoulders of giants.
Isaac Newton

I am indebted to Robert Blust on at least two counts. When I was a graduate student, Bob oversaw my initiation into the world of Austronesian historical linguistics. His guidance and personal attention—conducted through a long email correspondence concerning the reconstruction of Proto Bungku-Tolaki (Mead 1998)—started me down a road which has led to the present paper. Second, there is the communal debt which we all owe to Bob for his many insightful contributions regarding the prehistory of Austronesian languages. Not only has Bob been a prolific writer, he also sets the standard for what lucid and well-argued writing should read like.

The Bungku-Tolaki languages are spoken on the mainland of southeastern Sulawesi, Indonesia, and on a few immediately offshore islands (Sneddon 1983; Mead 1999). This paper has two aims: (a) to argue for the kinship terms which can be reconstructed for their common ancestor language, Proto Bungku-Tolaki (PBT), and (b) to account for present-day forms which do *not* continue the reconstructed kinship term, in other words to determine whence such terms were borrowed, or how they were innovated. This paper is not a comparison of kinship systems per se. Nonetheless, anyone with a knowledge of western Malayo-Polynesian kinship systems will find the terms and distinctions discussed herein comfortably familiar.

Following Mead (1998, 1999), fifteen Bungku-Tolaki languages can be recognised. Genetically these languages fall into an eastern and a western branch. Eastern Bungku-Tolaki comprises the following eight languages: Moronene, Kulisusu, Koroni, Taloki, Wawonii, Bungku, Bahonsuai and Mori Bawah. Western Bungku-Tolaki comprises seven languages: Mori Atas, Padoe, Tomadino, Waru, Tolaki, Rahambuu and Kodeoha. Overlaying this genetic classification, we can recognise at least four zones of contact (see Map 1).

Alexander Adelaar and Andrew Pawley, eds
Austronesian historical linguistics and culture history: a festschrift for Robert Blust, 489–507.
Pacific Linguistics, 2009.
© This edition vested with Pacific Linguistics.

Map 1: Bungku-Tolaki languages and other languages of Sulawesi mentioned in this study

- Tomadino, Bahonsuai and Koroni are small languages located in a Bungku-speaking area, and often exhibit convergence with Bungku. Tomadino and Bahonsuai speakers originate from the Mori area, while Koroni speakers originate from northern Buton Island (near the Kulisusu language area).
- Mori Bawah, Padoe and Mori Atas—often together considered one language 'Mori'—have had a long period of contact. This zone of convergence also included the smaller Tomadino and Bahonsuai isolects before their removal to the Bungku area.

- Moronene and Tolaki, particularly the Mekongga dialect of Tolaki, exhibit convergence.
- Wolio and Muna have mutually influenced each other, and both have also had a profound influence on Kulisusu. While geographically contiguous, these three languages belong to different language stocks (Wotu-Wolio, Muna-Buton and Bungku-Tolaki). Wolio is the language of the former sultanate at Baubau, and served as a lingua franca for Muna and Buton Islands.

Finally, South Sulawesi languages, particularly Bugis, have had an influence on a number of Bungku-Tolaki languages. In recent years speakers of some (if not all) Bungku-Tolaki languages have begun borrowing, or at least using from time to time, modern Indonesian kinship terms. Developments in this area, however, lie beyond the scope of this paper especially as they are mostly undocumented in my sources.

To date, the kinship terminologies of most Bungku-Tolaki languages remain underdescribed. Only two published descriptions of kinship terminologies exist—for Mori Bawah, found in Kruyt (1924:76–81), and for Tolaki, found in Tarimana (1989: 108–116)—and I could not have written this paper without the reliable descriptions of Tolaki, Bungku and Moronene kinship terminology provided to me by Scott Youngman and David Andersen. Kulisusu terms are from my own field research, which in recent years has also extended into Mori Bawah. A Holle list exists for Wawonii (Stokhof 1985:113–125), while Padoe kinship terms have been gleaned from the small draft dictionary by Lara, Larobu, et al. (1991). Information concerning kinship terms in other Bungku-Tolaki languages comes solely from 200-item word lists collected for basic survey purposes (Mead 1995, 1999:101–178).[1]

1 Person

The presumed Proto Bungku-Tolaki word for ‘person’ was *tau, directly reflecting PMP *tau, though at present reflexes are encountered only in Waru, Kodeoha and Tolaki *toono*.[2] The form *toono* has a frozen third person singular possessive pronoun *-no*.[3]

The term for ‘person’ in all other Bungku-Tolaki languages is *mia*. Despite its wider distribution, *mia* must be regarded as a borrowed term in Padoe and Mori Atas—if inherited, we would expect to find ***mie* with vowel raising (Mead 1998:105)—and therefore on internal grounds cannot be reconstructed at a level higher than Proto Eastern Bungku-Tolaki.

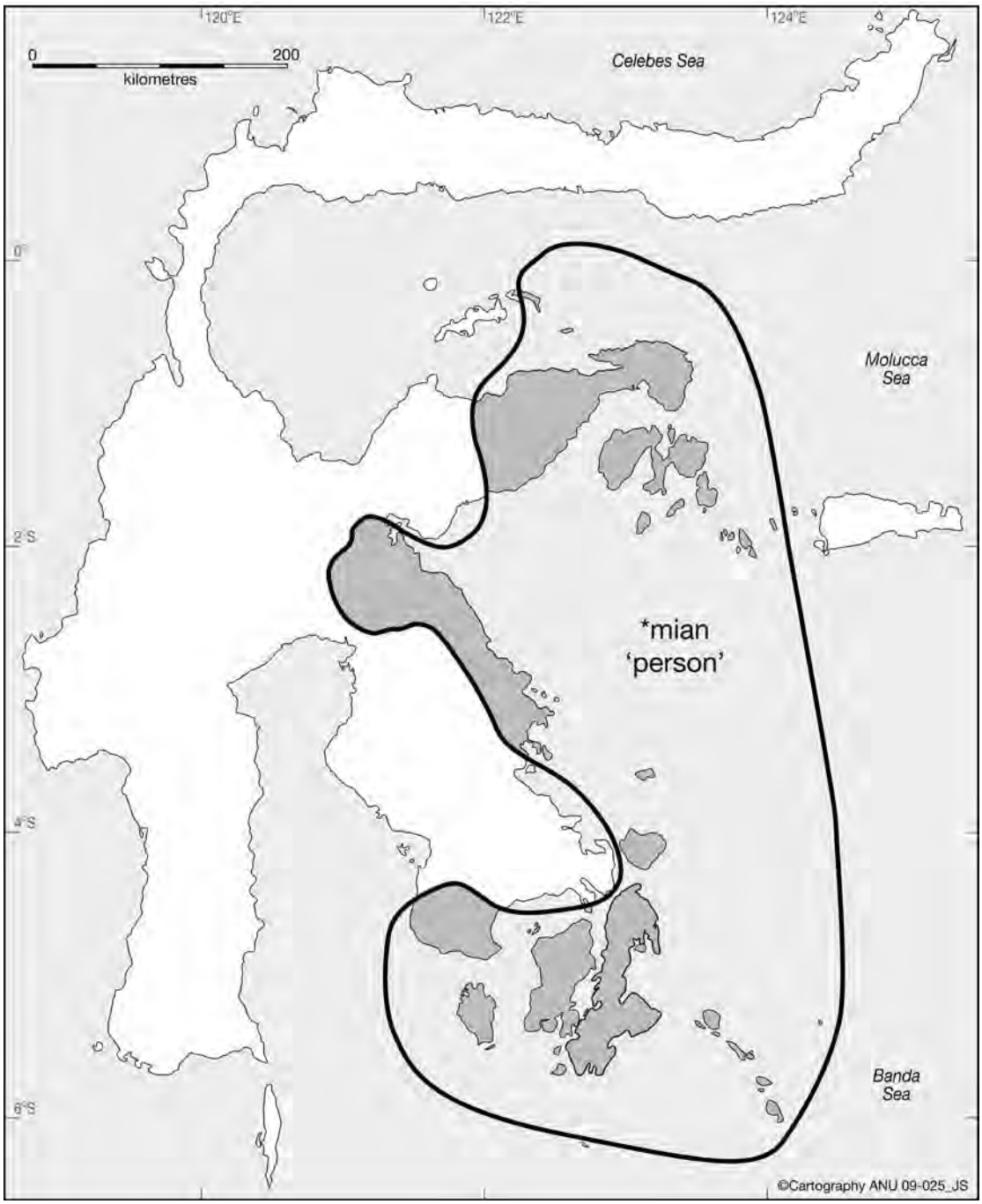
---

1 Unless otherwise noted, sources of data for other languages mentioned in this paper are: Kamaru and Southern Muna (René van den Berg pers. comm.), Muna (van den Berg 1996), Tukang Besi (Donohue 1999, 2000), Wolio (Anceaux 1987), Balantak and Andio (Robert Busenitz pers. comm.), Banggai (van den Bergh 1953), Saluan, Batui and Bobongko (author’s field notes), Makasar (Tim Friberg pers. comm.), Bugis, Konjo and Da’a (respective wordlists in Tryon 1995), Moma, Tado and Uma (Michael Martens pers. comm.), Pamona (Adriani 1928) and Tomini-Tolitoli languages (Himmelmann 2001). I owe much to Michael Martens whose insights and broad knowledge of Sulawesi languages led to many improvements in this paper.

2 A reduced form *to* meaning ‘member of a certain (ethnic) group’ occurs throughout the Bungku-Tolaki area in contexts such as *to Wuna* ‘Muna person(s), the Muna people’, *to Kolensusu* ‘Kulisusu person(s), the Kulisusu people’, etc.

3 In this respect, Tolaki *toono* is entirely parallel to Pamona *taunya* ‘person’.

Nonetheless, reflexes of a term *mian ‘person’ have a curious distribution across the eastern seaboard of Sulawesi, as they are also found in all Saluan-Banggai languages (Bobongko, Saluan, Andio, Balantak and Banggai *mian*) and Muna-Buton languages (Muna, etc. *mie*, others *mia*). See Map 2. Even Wolio and Kamaru have *mia* ‘person’, though the presumption must be that here, too, the term was borrowed, since other Wotu-Wolio languages have *ito* ‘person’ (Laidig and Maingak 1999:66).



Map 2: Distribution of reflexes of *MIAN ‘person’

Esser (1933:351) postulated that *mian ‘person’ developed via apheresis (the dropping of the initial vowel of a word) from the derived form *um-ian ‘inhabitant, dweller’ (cf. PMP *ian ‘dwell, reside, live in a place’). I follow Esser in this hypothesis, and further suggest that the term must have spread amongst the inhabitants of eastern Sulawesi at a very early date, though presumably after the people who eventually came to speak Western Bungku-Tolaki languages had migrated away from the coast and into the interior.

PMP *qaRta ‘outsiders, alien people’ is reflected as PBT *ata ‘slave’ (all Bungku-Tolaki languages *ata* ‘slave’). See Blust (1972) for a discussion of this semantic development, which is common to other Austronesian languages.

2 Man, woman, male, female

Table 1 shows present-day Bungku-Tolaki terms for ‘male’ and ‘female’ for which I have information.

Table 1: Bungku-Tolaki terms for ‘male’ and ‘female’

	+ human		– human	
	‘man, male’	‘woman, female’	‘male’	‘female’
Moronene	***tama***	***tina***	***tama***	***tina***
Kulisusu	***tama***	***cina***	***tama***	***cina***
Taloki	***tama***	***tina***		
Koroni	***tama***	***tina***		
Wawonii	***tama***	***tina***		
Bungku	***tama***	***tina***	***tama***	***tina***
Bahonsuai	***tama-tama***	***tina-tina***		
Mori Bawah	***tama***	*beine*	*laki*	*wai*
Mori Atas	***tuama***	*irowai*		
Padoe	***tama***	*irowai*	*laki*[4]	*waino*
Tomadino	***tuama***	*beine*		
Waru	*analaki*	*(ana) irowai*		
Tolaki	*langgai*	***tina***, *more*	***tama***	***tina***
Rahambuu	*analaki*	*irowai*		
Kodeoha	*kinalohi*	*pinihupi*		

PBT *tama and *tina are discussed in the following section. Despite the widespread occurrence of reflexes of these forms, shown in bold in Table 1, there is evidence for reconstructing two other terms, *laki ‘male’ (< PMP *laki ‘male (probably originally of animals)’) and *wai ‘female’ (< PMP *bahi ‘woman, female’). As *tama and *tina originally meant ‘father’ and ‘mother’, respectively, the general pattern must be that these terms have been spreading into the semantic domain of ‘male’ and ‘female’ at the expense of reflexes of *laki and *wai. In some languages, reflexes of *laki and *wai are to be found only in historical compounds, where the original meaning has become opaque to present-day speakers.

[4] Padoe *laki* specifically means ‘male (of carabao or cattle), bull’, cf. *lase* ‘boar’, and *manu lengke* ‘rooster’. A general term for ‘male (animal)’ in Padoe is unknown to me.

Evidence for PBT *laki ‘male’ includes Moronene *tama laki* ‘warrior, mighty man’, Kulisusu *laki* ‘head or chief, usually referring to the leader of a village’, *laki lima* ‘middle finger’ (lit. ‘chief of the hand’), Mori Bawah *laki* ‘male (of animals)’, Padoe *laki* ‘male (of bovines)’, Tolaki *tamalaki* ‘brave, purposeful and trusted male leader’, *tombalaki* ‘condition where the husband controls the kitchen, including the use and distribution of food; one of three recognised grounds for a wife to request divorce’ (cf. *mondotomba* ‘control the distribution of food at a party so that it is not wasted’) and Waru and Rahambuu *analaki* ‘man, male’.

Evidence for PBT *wai ‘female’ includes Moronene *waipode* ‘young woman’, Mori Bawah *wai* ‘female (of animals)’, Padoe *wai* ‘woman, girl’ (term of endearment), Padoe, Mori Atas, Waru and Rahambuu *irowai* ‘female’, and Tolaki (Mekongga dialect) *waipode* ‘young girl, unmarried woman’ and *uewai* ‘species of rattan’.[5]

The source of the Kodeoha replacement forms *kinalohi* ‘man, male’ and *pinihupi* ‘woman, female’ and Tolaki (Mekongga dialect) *more* ‘woman, female’ are unknown to me. Mori Bawah and Tomadino *beine* ‘woman, female’ is ultimately from PMP *ba-b[in]ahi, but limited distribution in Bungku-Tolaki suggests borrowing.[6]

Cognates of the Tolaki form *langgai* ‘man, male’ are found elsewhere in the Bungku-Tolaki area, cf. Bahonsuai, Mori Bawah, Padoe, Mori Atas, and Tomadino *langkai*, but in these languages this form has the meaning ‘big, large’ (cf. also Mori Bawah *mompelangkai* ‘honor, esteem’). Cognates are found widely across Sulawesi with an unmistakable association with maleness, for which Mills (1981:68) reconstructed *laŋkai ‘male; revered (?)’. Mills cited Bugis *lakkai* ‘husband’, Ledo *langgai* ‘male’, Pamona *langkai* ‘man, male’ and Wana (dialect of Pamona) *rongo talangkai* ‘husband’, to which we may add Banggai *langkai*, *langgai* ‘man’, Balantak, Andio, Saluan *langkai’* ‘husband’, Bobongko *langkai’* ‘old man; old (of persons)’, Dampelas *langkai* ‘old man’, other Tomini-Tolitoli languages *langkai, langgai* ‘man, male’.

In many of these same languages there is a parallel form for a respected female, as evidenced by Andio *bengkele*, Saluan *bengkele’*, *bingkele’* and Batui *bingkele’* ‘wife’, Bobongko *bengkele’* ‘old woman’, Tado and Moma *bangkele* ‘wife, woman’, Totoli *bakele* ‘grandmother’, Ampibabo-Lauje *bengkel* ‘old woman’, and Dampelas, Pendau, and Lauje *bengkel*, Tajio *benggel* ‘woman, female’. In Bungku-Tolaki languages, I know of only Kulisusu *bangkele*, said to be an archaic term for ‘wife’ (also with reduplication *bangke-bangkele* ‘a very old woman’) along with Mori Bawah *bakele* ‘old woman’.

The role which borrowing played in the distribution of these forms across Sulawesi has yet to be determined, though potentially it was significant. At any rate, the forms were

---

5 For the etymology of Tolaki *uewai*, cf. Kulisusu *ue ncina* and Wolio *lauro ɓawine* ‘species of rattan’, but literally ‘woman’s rattan, female rattan’.

A root *wai ‘female’ possibly also lies at the root of PBT *wainto, reflected in Moronene *wainto* ‘actually, although’, sometimes used as an expression of mild surprise ‘I didn’t know that before’, Mori Bawah *wainto* a word used as an expression of pity or lament, ‘too bad, poor thing’ (Esser 1933:221), Padoe *wainto* ‘a pity!’, and Tolaki *waindo* ‘it’s clear to me now; sure enough it’s true’, originally *wai-nto ‘our (incl.) female, our (incl.) dear’.

6 With South Sulawesi as the most likely source, cf. Bugis *baine* ‘wife’, Makasar *baine* ‘woman, female, wife’, Konjo *bahine* ‘woman, female, wife’. Whilst other, nearby languages of Sulawesi reflect PMP *ba-b[in]ahi, phonological developments make them less likely as loan sources, compare among others Uma *tobine* ‘woman, wife’ and Banggai *boine* Saluan *boune*, Batui *boine*, Andio *bobine* and Balantak *wiwine* (< Proto Saluan-Banggai *bobine), Muna *roɓine* Tukang Besi *wowine* and Wolio *ɓawine*, all meaning ‘woman, female’.

possibly bimorphemic in origin: *la + ŋkai and *be + (ŋ)kele,[7] since in Tolaki we find *La* 'particle used preceding a man's name' and *We* 'particle used preceding a woman's name' (in Kulisusu, Muna, Wolio and in the Tukang Besi Islands respectively *La* and *Wa*).[8] One can also cite forms without *la* and *be*, cf. Pamona *ngkai*, Saluan and Andio *kai'*, Balantak *kakai'* and Banggai *kakai*, all meaning 'grandfather', most Tomini-Tolitoli languages *kai* 'grandfather' (vocative); and Saluan, Batui and Andio *kele'* 'grandmother' and Totoli *kele* 'grandmother' (vocative). Alternatively, of course, such forms could have developed secondarily through clipping (loss of the initial or final part of a word).[9]

Finally, on scant evidence we can also reconstruct PBT *bokeo[ ] 'respected male person', based on Tolaki *bokeo* 'honorific term used for a regent (Indonesian *bupati*)' and Kulisusu *ɓokeo*, said to be an archaic term for 'husband' (also with reduplication: *ɓoke-ɓokeo* 'a very old man'). In Tolaki *bokeo* also means 'crocodile', a semantic extension found in other parts of Austronesia and which is to be ascribed to 'totemic ideas that were present in Proto Malayo-Polynesian social organization' (Blust In progress). Cognates are unknown to me elsewhere, unless one includes Muna *ɓoke* 'prominent person, VIP'.

3 Father, mother

Proto Bungku-Tolaki terms for 'father' and 'mother' straightforwardly reflect the Proto Malayo-Polynesian 't-form paradigm' reconstructed by Blust (1979:221) and shown Table 2.

Table 2: PMP and PBT terms for 'father' and 'mother'

	'father'	'mother'
vocative	*ama	*ina
referential	*t-ama	*t-ina

Three patterns emerge in the daughter languages.

(a) In eastern Bungku-Tolaki languages (apart from Mori Bawah and the Bungku area, see below), this system was maintained, cf. Moronene *ama* (v), *tama* (r), *ina* (v), *tina* (r) and Kulisusu *ama* (v), *tama* (r), *ina* (v), *cina* (r).

(b) In western Bungku-Tolaki languages and in Mori Bawah, *t-forms became disused in the meaning 'father' and 'mother', consequently these languages have only one form used in both reference and address. Cf. Bahonsuai, Mori Bawah and all western Bungku-Tolaki languages *ama* 'father', Bahonsuai *ina*, Mori Bawah, Mori Atas, Padoe, Tomadino *ine*, Waru, Tolaki, Rahambuu and Kodeoha *ina* 'mother'.

---

[7] Adriani (1928:s.v.) adopted a similar etymology for Pamona *langkai*, stating elsewhere (Adriani 1900:312) that in Pamona *ngkai* probably originally meant 'penis', though it is unclear on what grounds he arrived at this hypothesis.

[8] These 'title' particles are possibly to be related to the aforementioned PMP *laki and *bahi, but as far as I know pairing of a 'masculine' particle *la* with a 'feminine' particle *wa*/*we* is currently limited to southeastern Sulawesi. In Bugis the particle *la* used with masculine names is paired with *i* used with feminine names. In Pamona *la* is used with both masculine and feminine names, and is preceded by *i* (e.g. *i Laboti-boti*, the name for 'Monkey' in a folktale).

[9] In Balaesang, Dampelas, Dondo and Lauje *beke* 'grandmother' (vocative), it appears instead that the final syllable was clipped.

(c) The Bungku area underwent unique developments. Standard Bungku employs the following terms: *apu* or *tu'a* 'father' (r), *tama* 'father' (v); *indo* or *ina* 'mother' (r), and *indo* 'mother' (v). Dialectally *bae* or *bai* is also encountered for 'mother', and *ua* for 'father'. Ultimate PMP sources are *apu 'grandfather', *tuqah 'old', *indu-q 'mother' and *bahi 'woman, female' (for *ua*, see §6), but the Bungku forms *indo* and *bai*/*bae* may not be directly inherited. Further investigation is needed.

4 Child

The term for 'child' found universally throughout the Bungku-Tolaki area is *ana*, for which we reconstruct PBT *anaQ, from PMP *anak. A final consonant *-Q must be reconstructed for the PBT form based on the higher level reconstruction. A trace of this consonant is also to be seen in the glottal stop which occurs in Moronene *ana'u* 'your (sg.) child' (Mead 1998:127). Regarding the basis for reconstruction PBT *-Q (and also below in some forms, *-N), see Mead (1998:71ff.).

5 Sibling

As shown in Table 3, all present-day Bungku-Tolaki languages distinguish between elder and younger sibling.

Table 3: Bungku-Tolaki terms for elder and younger sibling

	'elder sibling'	'younger sibling'
Moronene	*tukaka*	*tuai*
Kulisusu	*kaaka*	***andi***
Koroni	*kaaka*	***andi***
Taloki	*kaka*	***andi***
Wawonii	*tukaka*	*tuwai*
Bungku	*tukaka*	*tuai*
Bahonsuai	*aka*	***andi***
Mori Bawah	*aka*, *kaka*	*uai*, ***andi*** [10]
Mori Atas	*kaka*	***andi***
Padoe	*kaka*	*hai*
Tomadino	*kaka*	***andi***
Waru	*kaka*	*hai*
Tolaki	*kaaka*	*hai*
Rahambuu	*kaka*	*hai*
Kodeoha	*kaka*	*hai*

The term *andi* 'younger sibling' (highlighted in Table 3) is widespread in the Bungku-Tolaki area. Nonetheless I would argue that this distribution is to be attributed to borrowing, for the following reasons. First, probable sources can be readily identified, namely Wolio *andi*, Makasar *andi* or Bugis *anri*.[11] Second and more importantly, once *andi* forms are removed from consideration, all other present-day terms can be traced back

[10] Kruyt (1924:79) does not mention the forms *kaka* or *andi*, which may therefore be recent introductions.
[11] Compare Proto South Sulawesi *a(n)di 'younger sibling' (Mills 1975:616).

to one of four Proto Bungku-Tolaki sibling terms which form a natural paradigm, namely *tukaka 'elder sibling (referential)', *kaka 'elder sibling (vocative)', *tuai 'younger sibling (referential)', and *hai 'younger sibling (vocative)'.[12] Blust (1979) gives a full account of the referential prefix *tu- of *tukaka and the *t- of *tuai (see also above §3). It remains only to note that the old referential/vocative distinction appears to have been everywhere lost—at least insofar as we have data available to us—and where it has been renewed, it has been reformed along other lines. Compare for example Kulisusu *kaaka* 'elder sibling (referential)', *kaka* 'elder sibling (vocative, rarely used)', *andi* 'younger sibling (referential)', *ade* 'younger sibling (vocative, rarely used)'. The Mori Bawah forms *aka* and *uai* have their origin in a vocative paradigm formed by consonant subtraction (Blust 1979:227).

Parallel- or cross-sibling terms are unknown in any Bungku-Tolaki language. Terms which mean 'sibling' (without respect to relative age or relative sex) also exist in some Bungku-Tolaki languages, but are of secondary origin. Compare Moronene *petila* 'sibling' (stem *tila* 'divide, apportion'), Kulisusu *sahinaa* 'sibling; more generally, anyone with whom one shares a common ancestor' (lit. 'one place of being born', root *hina* 'be born, exist'), Bungku *petutuai* 'sibling' (stem *tuai* 'younger sibling'), Mori Bawah *pepaekompo* '(full) sibling', nominalised form of the verb *mepaekompo* 'having at least one parent in common, though more particularly both parents' (< *pae* 'pull', *kompo* 'belly'), and Tolaki *paekombo, kotukombo, serekombo* 'full sibling' (cf. *mokotu* 'severed' *mosere* 'cut s.t. across', *kombo* 'intestines, guts').

6 Grandparent, grandchild

Bungku-Tolaki forms for 'grandparent' and 'grandchild' (in most cases the same term is used reciprocally) are given in Table 4.

Table 4: Bungku-Tolaki terms for grandparent and grandchild

	'grandparent'	'grandchild'
Moronene	*mbue*	*mbue*
Kulisusu	*apua*	*apua*
Taloki	***waowa***	***waowa***
Koroni	***ua***	***ua-ua***
Wawonii	***uwa***	***uwa***
Bungku	***ua***	***ua***
Bahonsuai	***ue***	***ue***
Mori Bawah	***ue***	*(ana)* ***ue***
Mori Atas	***ue***	*ana* ***ue***
Padoe	***ue***	*(ana)* ***ue***
Tomadino	***ue****kai*, ***ue****kele* [13]	***ua***
Waru	***ue***	***ue***
Tolaki	*pue, mbue*	*pue, mbue*
Rahambuu	*peepe*	*peepe*
Kodeoha	*nene*	*peepe*

[12] PBT *hai is a regular development from pre-PBT *wai from *uai, ultimately from PMP *huaji 'younger sibling'.

[13] Specifically, *uekai* 'grandfather' and *uekele* 'grandmother'. Regarding *kai* and *kele*, see §2.

Based on the bolded forms in Table 4, I reconstruct PBT *uaN ‘grandparent, grandchild’. The final nasal is reconstructed on the premise that *uaN was a term of address formed by subtraction from earlier *puaN, which in turn was possibly an aphetic form (a form which has dropped the initial vowel) of *apuaN or *opuaN, cf. Southern Muna *awua* ‘grandparent, grandchild’, Wolio *opua ~ pua* ‘grandparent, grandchild’, the ubiquitous South Sulawesi title of respect *puang* ‘sir, master, lord’, and Malay *uan*, *puan* and *tuan*.[14] As far as I know, however, there is no incontrovertible evidence from within Bungku-Tolaki languages for the supposed intermediary stem *puaN.

It is tempting to derive Moronene *mbue* and Tolaki *pue*, *mbue* within this same framework. However, whilst raising of *a > *e* following a high vowel was regular in Mori Atas and Padoe (Mead 1998:105–106), it would be irregular for Tolaki or Moronene. Possibly these forms are connected to *pue* as found in Da’a *pue* ‘grandparent’, Pamona *bue* (dialectally *pue*) ‘grandmother’, and Uma *pue’* ‘lord, master’. Whether the connection should be regarded as through inheritance or borrowing, however, remains to be determined. Indeed, various forms meaning ‘owner’, ‘lord’, ‘ancestor’, ‘grandparent’, etc. and containing the syllable …*pu*… are found across Sulawesi, and would require a separate study to fully explicate.

The etymology of Kodeoha and Rahambuu *peepe* is unknown. Kodeoha *nene* is from Malay *nenek*.

7 Aunt, uncle, nephew, niece

The terms for ‘aunt’ (parent’s sister) and ‘uncle’ (parent’s brother)[15] which can be reconstructed for Proto Bungku-Tolaki are *naina and *maama. As throughout the Bungku-Tolaki area, there is no distinction whether this person is a sibling of one’s father or one’s mother. The Proto Bungku-Tolaki terms in turn must originate via apheresis from the reduplicated forms *ina-ina and *ama-ama. Compare Wawonii, Mori Bawah, Padoe and Tolaki *maama* ‘uncle’, Wawonii, Mori Bawah and Tolaki *naina*, Padoe (with regular vowel raising) *neine* ‘aunt’. Even Kulisusu, which has innovated replacement terms for ‘aunt’ and ‘uncle’, still retains *maoma* and *naina* as polite terms of address for older men and women.

Whilst we lack information on the lesser described languages, it seems safe to reconstruct PBT *laki-anaQ for ‘nephew, niece’ (*not* distinguished according to sex) based on Moronene, Wawonii, Bungku, Mori Bawah and Tolaki *laki’ana* (no other terms known).

Bungku has innovated terms for ‘uncle’ and ‘aunt’ by analogy with the term for ‘niece, nephew’, thus we find throughout the Bungku area *laki’ama* ‘uncle’ and *laki’ina* ‘aunt’, including in the Koroni and Tomadino languages. The terms *lakiama* and *lakiina* were also recorded in Kodeoha on the western coast, and *lakiama* and *lakine* in Mori Atas in the interior.

Moronene has replaced the inherited terms for ‘aunt’ and ‘uncle’ with *tina’ate* and *ma’e* (from earlier *tama’ate*), formed from the respective terms for ‘father’ and ‘mother’ plus the diminutive suffix *–’Vte* (where *V* reduplicates the final vowel of the stem to which it attaches) (Andersen 1999:16).

---

[14] These are not necessarily from the same source. Presumably Muna *awua* and Wolio *opua* are from PMP *apu ‘grandfather’ plus *-an or *-en, but Blust (in progress) reconstructs PMP *puaŋ ‘title of respect’ for Makasar, Sa’dan Toraja, etc. *puang*.

[15] In general we lack information concerning terms for the spouse of one’s blood-aunt or blood-uncle, but see §12.

Kulisusu has an innovative set of terms for 'uncle', 'aunt' and 'nephew/niece', respectively *pinolitama*, *pinolicina* and *pinoli'ana*. These terms resemble the Wolio terms *pinoama*, *pinoina* and *pinoana*,[16] but the syllable *-li-* in the Kulisusu forms is a point of discrepancy. Perhaps *pinoli-* seen in the Kulisusu forms originated from shortening of earlier *pino-laki-.

8 Cousin

The term for 'cousin' which is reconstructed for Proto Bungku-Tolaki is *topisaN, where *to- is a reduced form of PMP *tau 'person', and *pisaN is derived from PMP *pisan 'united, joined'. PBT *topisaN is supported internally by Moronene, Kulisusu and Wawonii *topisa*, and by the wider external evidence, including Muna *ntopisa ~ pisa*, Tukang Besi South *sapi* (from metathesis), Tagalog *pinsan* and Balinese *misan* 'cousin' (Tryon 1995:210). Another form for 'cousin' is *poteha*, found in Bungku, Mori Bawah and Tolaki. I regard *poteha* forms to represent an irregular development, including metathesis, from PBT *topisaN, and I reconstruct *poteha[ ] only at the level of Proto Western Bungku-Tolaki.[17]

So far as we have data, Bungku-Tolaki languages are able to distinguish degree of laterality. Usually the basic term for cousin is followed by words which mean 'once', 'twice', etc. (in Tolaki, 'one time', 'two times', etc.) in a pattern similar to English. See Table 5. In Kulisusu (as well as Muna and Wolio, also shown), the pattern is to attach *to-* directly to the iterative numeral.[18]

Table 5: Degree of laterality in cousin terms in selected languages

	'first cousin' (PSbC)	'second cousin' (PPSbCC)	'third cousin' (PPPSbCCC)
Moronene	*topisa ponoha*	*topisa pendua*	*topisa pentolu*
Bungku	*poteha ponoha*	*topisa pendua*	*topisa pentolu*
Mori Bawah	*poteha mpohona*	*poteha mpendua*	*poteha mpentolu*
Tolaki	*poteha monggo'aso*	*poteha monggoruo*	*poteha monggotolu*
Kulisusu	*topisa*	*topendua*	*topentolu*
Muna	*pisa* (or: *ntopisa*)	*ndua* (or: *topendua*)	*ntolu* (or: *topentolu*)
Wolio	*tolida*	*topendua*	*topentalu*

[16] The Wolio terms literally mean 'the one who is called father', 'the one who is called mother', and 'the one who is called child'. Parallel forms are found in parts of Central Sulawesi, compare among others Da'a *pinutina* 'aunt' and *pinuana* 'niece, nephew'.

[17] An alternative hypothesis would be to ascribe the innovative form *poteha[ ] to the level of Proto-Bungku-Tolaki, under the assumption that Wawonii, Kulisusu and Moronene *topisa* reflect subsequent borrowing rather than inheritance—presumably under the influence of one or more Muna-Buton languages. Another small piece of evidence is that in a separate development, PMP *pisan 'united, joined' also yielded PBT *pihaN 'all together, all at one time, once', in which *s > *h. Hopefully improved knowledge of cousin terms in the lesser known Bungku-Tolaki and Muna-Buton languages will allow us to better weight arguments one way or the other.

[18] Kulisusu also has *topempaa* 'fourth cousin' (Wolio *topeapa*), and *topelima* 'fifth cousin', but this represents somewhat of a practical limit of the system. How many degrees are distinguished in other Bungku-Tolaki languages is unknown. Kulisusu *topisa* also serves as a general term for 'cousin' without respect to degree of laterality.

9 Spouse

For ‘spouse’ the term *wali is reconstructed, from PMP *baliw ‘dual division, moiety’ (replacing reflexes of PMP *qasawa ‘spouse’). The reference to ‘spouse’ was specialised from a more general meaning of ‘friend, companion’, which meaning can also be attributed to PBT *wali.[19] Compare the distribution of reflexes of *wali (bolded forms) in Table 6.

Table 6: Present-day Bungku-Tolaki terms for ‘spouse’ and ‘companion’

	‘spouse’	‘companion’
Moronene	*samotu’a*, *sampora*	*salako*, *simbau*, *tokia*
Kulisusu	*mia raha*	***wali***
Koroni	*teba*	*kunsi*
Wawonii	***wali***	*pewali*
Bungku	*teba*	***fali***
Bungku (Torete d)	***wali***	*bangkona*
Bungku (Tulambatu d)	*ciwa*	*banggona*
Bungku (Waia d)	*tifa*	*petifai*
Bahonsuai	*sombori*	*fali-fali*
Mori Bawah	*kombia*	***wali***, *pewali*
Mori Atas	***wali***	*pokowali*
Padoe	*sombori*	***wali***
Tomadino	*inalo*	***fali***
Waru	***wali***	*banggona*
Tolaki (Wiwirano d)	*rapu*	***wali***
Tolaki (Asera d)	*pekombu*	*wali-wali*
Tolaki (Konawe d)	***wali***	*banggona*
Tolaki (Mekongga d)	***wali***, *sangginaa*	***wali***
Tolaki (Laiwui d)	***wali***	*bagona*
Rahambuu	*pehoko*	***wali***
Kodeoha	*sagina*	***wali***

As Table 6 also indicates, these meanings have usually undergone lexical differentiation. Either the term for ‘spouse’ was replaced, or a derived form of *wali came to be used in the meaning ‘friend, companion’. (Less frequently, the term for ‘companion’ was replaced.)

In not a few cases the replacement term for ‘spouse’ has been derived from a term related to the hearth or kitchen, compare Mori Bawah *kombia* which means both ‘spouse’ and ‘the part of the floor in the house which is nearest to the hearth’ (Kruyt 1924:80). Similarly both Bahonsuai and Padoe have *sombori* ‘spouse’, a term found back in Mori Bawah *sombori* with the meaning ‘kitchen’.[20] Finally, Moronene has *rapi* ‘family unit’ and

[19] Similarly, reflexes of Proto South Sulawesi *bali ‘side (friend, companion)’ have come to mean ‘spouse’ in some languages, compare for example Seko Padang *hali* ‘in balance’, *haliang* ‘spouse, partner, one side (team) in a sports event, counterpart’ (Laskowske 2007:158). This point of similarity between South Sulawesi and Bungku-Tolaki languages is probably to be attributed to drift (parallel development).

[20] In Padoe and Mori Bawah, *sombori* can also mean ‘household, family’.

Wiwirano (a dialect of Tolaki) has *rapu* 'spouse',[21] both of which ultimately derive from PMP *dapuR 'hearth'. In all three cases, it is possible that the semantic shift was mediated by a verb form which meant 'to have, be supplied with a hearth or kitchen' but which by extension came to mean 'be married' (in that newlyweds would establish their own cooking place), with the term for 'spouse' then developing via backformation. Compare Mori Bawah *mekombia*, Padoe *mesombori*, and Tolaki *merapu*, all meaning 'be married'.

The Tomadino term *inalo* is derived from the transitive stem *alo* 'take' (< PMP *ala), and literally means 'that which (the one who) is taken (in marriage)'.

Other replacement terms for 'spouse' are also encountered in the Bungku-Tolaki area, of which the most widespread is *teba*, found in Bungku (dialectally also *ciwa*, *tifa*) and Koroni. The origin of this term is unknown. In Mekongga (western dialect of Tolaki), *sangginaa* is derived from *saN-* 'one' and *kinaa* 'rice, food'.[22] Asera *pekombu* 'spouse' is probably derived from *kombu* 'kind of small-fruited areca', in reference to the chewing of betel nut during marriage negotiations (compare Malay *pinang* 'areca', *meminang* 'propose, ask for in marriage'). The Kulisusu term *mia raha* 'spouse' is a compound derived from *mia* 'person' + *raha* 'house'. Moronene *samotu'a* 'spouse' is derived from *saN-* 'one' + *motu'a* 'old'. Moronene *sampora* 'spouse' also means 'fiancé', which must be regarded as the original meaning.[23]

Terms which specifically mean 'husband' or 'wife' (differentiated according to sex) are not reconstructed. Where known to exist, these terms are almost invariably the corresponding terms for 'man' and 'woman', such as Mori Bawah *tamano* 'her husband' (lit. 'her man'), *beineno* 'his wife' (lit. 'his woman'), Tolaki *langgaino*, *moreno*, etc. (see §2). In Kulisusu, speakers employ the polite circumlocutions *tamano anano* 'her husband' (lit. 'the father of her children') and *anahakono* 'his wife' (from *ana* 'child' + plural marker *hako*). Similarly, in Moronene, *tinano ana'ate* can be used for 'wife'.

10 Parent-in-law, child-in-law

So far as we have data, only Kulisusu has contrastive terms for 'parent-in-law' (*poniana*) and 'child-in-law' (*ana monia*). However, since these terms continue forms which are reconstructed for Proto Celebic (respectively *panianan and *manian, see Mead 2003:134) and since there is negligible possibility of borrowing, we must perforce reconstruct PBT *ponianaN 'parent-in-law' and *moniaN 'child-in-law'.

Other languages employ a single, reciprocal term. Moronene, Bungku and Mori Bawah have *poni* meaning both 'parent-in-law' and 'child-in-law' (where *poni* is clipped from *ponianaN). Wawonii and Tolaki have *baisa*, also a reciprocal term. Although *baisa* is ultimately derivable from PMP *baisan 'kinship tie between the parents of a married couple', a consideration of sound change suggests that it was not directly inherited. Bugis *baiseng* [baˈisɨŋ] 'co-parent-in-law' is the most likely loan source, assuming of course semantic shift in the borrowing process.

---

[21] Compare also Tolaki (Konawe dialect) *sarapu* 'fiancé', from *saN-* + *rapu*, but in the present language regarded as monomorphemic.

[22] In the eastern Konawe dialect of Tolaki *sangginaa* has vulgar sexual connotations. In Moronene *sangkinaa* means 'slave'.

[23] Compare also Kulisusu *sampora*, Padoe, Bungku, Mori Bawah *sambora* 'fiancé' < PBT *sampora[ ] from *saN- 'one' + *pora[ ] (stem unknown).

In fact a term for ‘co-parent-in-law’ cannot be reconstructed for Proto Bungku-Tolaki. In Bungku and Mori Bawah, the respective parents of a married couple can also call each other *poni* (in Moronene, *asa mponi*), while Tolaki and Kulisusu simply lack any term for the co-parent-in-law relationship.

11 Sibling-in-law

Bungku-Tolaki terms for ‘sibling-in-law’[24] such as they are known to me are set out in Table 7 (information on Wawonii and Padoe terms is likely incomplete).

Table 7: Bungku-Tolaki terms for ‘sibling-in-law’

Moronene	*bela, ela*	male’s brother-in-law
	andea	female’s sister-in-law
	oleo	sibling-in-law of opposite sex
	asa	spouse’s sibling’s spouse
Kulisusu	*ɗawo*	spouse’s sibling (also sibling’s spouse?)
	era	spouse’s sibling’s spouse
Wawonii	*dawo*	sibling-in-law
Bungku[25]	*peleleanga*	spouse’s sibling or sibling’s spouse
	asa mpo’ala	spouse’s sibling’s spouse
Mori Bawah	*oleo*	spouse’s sibling or sibling’s spouse
	ansa	spouse’s sibling’s spouse
Padoe	*oleo*	brother-in-law, sister-in-law
Tolaki	*ela*	male’s brother-in-law
	bea	female’s sister-in-law
	hine	sibling-in-law of opposite sex
	asa	spouse’s sibling’s spouse

The Bungku form *asa mpo’ala* ‘spouse’s sibling’s spouse’ (sometimes shortened to *sampo’ala*) literally means ‘one taking’, in reference to two people taking their respective spouses from one set of siblings. I consider it likely, though far from conclusive, that Tolaki and Moronene *asa* and Mori Bawah *ansa* derive historically through clipping of this or some similar construction.

Kulisusu *era* ‘spouse’s sibling’s spouse’ derives ultimately from PMP *idas ‘affine of ego’s generation’, but limited distribution suggests it could be borrowed. Like Kulisusu, Wolio also has *ɗawo* and *era*, but these are not distinguished by Anceaux (1987:s.v.) who glosses *both* terms simply as ‘brother-in-law, sister-in-law’.[26]

Next to Wawonii *dawo* and Kulisusu and Wolio *ɗawo*, the Southern Muna term for ‘spouse’s sibling’ and ‘sibling’s spouse’ (reciprocal term) is also *ɗawo*. Because these languages are geographically contiguous but belong to three different language groups, one

[24] Specifically in this paper: male’s brother-in-law = WB, ZH of male ego; female’s sister-in-law = HZ, BW of female ego; and sibling-in-law of opposite sex = WZ, BW of male ego, HB, ZH of female ego.

[25] Specifically, standard Bungku. Sibling-in-law terms in the Torete dialect of Bungku are identical to—and doubtless borrowed from—Tolaki.

[26] Anceaux also incorrectly assigns *era* as a borrowing from Muna, which has only *tamba* ‘spouse’s sibling, sibling’s spouse’ (Southern Muna: *ɗawo*) and *kamodu* ‘spouse’s sibling’s spouse’.

must expect borrowing to have played a role with this term as well, though the donor language or ultimate source have not been identified.

The Moronene and Tolaki term *ela* 'male's brother-in-law' (in Moronene also *bela*) have cognates in Kulisusu *ɓela* 'male friend of another male' and in two Mori Bawah forms, *bela, ela*, which originally meant 'friend' (Esser 1927:115, 118), but which are now used primarily as interjections of annoyance, disapproval, or disbelief.[27] Whatever the Bungku-Tolaki internal evidence, outside witnesses clearly support 'friend' as the original meaning, compare among others Muna *ɓela* 'friend (term of friendly address)', Tukang Besi *ɓela* 'dear, beloved; spouse', Balantak *bela'* 'friend', Pendau *bela* 'friend' (according to some, 'girlfriend') (Phil Quick pers. comm.), and even Proto South Sulawesi and Proto Ambonese *bela 'friend' (Mills 1975; Stresemann 1927, cited in Wurm and Wilson 1975:86). However, if Mills is correct to connect Sulawesi forms meaning 'friend' with Old Javanese *bela* 'to die together with a person' and Malay *bela* 'sacrifice' (Mills 1981:66), then the forms must be of relatively recent introduction—that is, they could not be reconstructed at the level of Proto South Sulawesi or earlier—because the Javanese and Malay forms are borrowed from Sanskrit *velā* 'time, hour of death' (Gonda 1952:370).

Similarly, cognates of the Moronene term *andea* 'female's sister-in-law' are to be found in Kulisusu *andea* '(close) female friend of another female' and Wolio *andea* 'friend, comrade, companion, mate, fellow' (I am unaware of cognates outside of these two languages). And even Tolaki *bea* 'female's sister-in-law' stands next to *sabea*, a Tolaki word meaning 'close friend' (< *saN-* + *bea*). Given these connections, reconstructing same-sex sibling-in-law terms for Proto Bungku-Tolaki would be ill-advised. The only term which can be reconstructed on present evidence is thus PBT *oleo[ ] 'sibling-in-law', by which we mean a sibling of one's spouse or the spouse of one's sibling.

12 Stepparent, stepchild

PBT *awo indicated a relationship in a blended family. Where a reflex is known, it is invariably placed after the kinship term which it modifies, as in Kulisusu *tama awo* 'stepfather', *cina awo* 'stepmother' and *ana awo* 'stepchild', and mutatis mutandis for Moronene, Wawonii, Bungku, Mori Bawah and Tolaki (in Bungku, *awo* by itself can mean 'co-wife', Indonesian *madu*).[28] Compare elsewhere in Sulawesi: Wolio *amo awo* 'stepfather', *ina awo* 'stepmother' and *ana awo* 'stepchild', Balantak *silaabo*, Moma, Lindu, Napu *awo*, Uma *awo'* 'stepparent/stepchild', and Pamona *awo* 'stepchild'. From this internal evidence we can reconstruct Proto Celebic *abo 'step(parent/child)',[29] but this may be too simplistic. In three widely separated areas we find that Torete (dialect of Bungku) *awo*, Kabaena (dialect of Moronene) *awo* and Uma *awo'* also mean 'spouse of one's uncle or aunt' (PSbSp) and, reciprocally, 'the niece/nephew of one's spouse' (SpSbC). An alternative hypothesis is that the original reference of *abo was 'an affine of the first ascending generation', which later came to be used reciprocally, and/or used to signify relationships in a blended family (through remarriage of the parent).

---

[27] As in: *Nahi ta hina hapa kada, bela* 'Nothing is going to happen, friend!' (where the speaker is annoyed by the timidity of another) (Esser 1927:159).

[28] Padoe has *poro'uma* 'stepfather', *poro'ine* 'stepmother'. The prefix *poro-* is from South Sulawesi influence, cf. Bugis *poro-ambo'* 'stepfather' and *poro-indo'* 'stepmother'.

[29] Konjo *ana' aho* 'stepchild', also 'child of a second-wife marriage' (and *anrong aho* and *amma aho* with analogous meanings) (Timothy Friberg pers. comm.) suggests that the reconstruction should be assigned to an even greater time depth.

13 Miscellaneous

The Proto Bungku-Tolaki term for 'orphan' can be reconstructed as *elu, from PMP *qilu 'orphan'. Compare Moronene, Bungku and Tolaki *ana elu* and Mori Bawah and Padoe *elu-elu.*

A Proto Bungku-Tolaki term for 'widow' can be reconstructed as *walu, from PMP *balu 'widow(er)', on the strength of Moronene *walu* 'widow, divorcee', Mori Bawah *walu* 'widow, widower' and Tolaki *walu* 'widow young enough to marry again'. Mori Bawah also has *ta'ubalu* (with stem *ta'u* 'old') and *tuwua*[30] 'widow(er)', but how these terms differ from *walu* is not presently understood. Other terms are known to exist in the Bungku-Tolaki area, but require further investigation.[31]

14 Conclusions

The kinship terms which I reconstruct for Proto Bungku-Tolaki are summarised in Table 8. Some related terms, discussed in §1, §2 and §13, include PBT *tau 'person' (next to Proto Eastern Bungku-Tolaki *mian 'person'), *laki 'male', *wai 'female', *elu 'orphan' and *balu 'widow(er)'. Following Fox (1994:30), we can ask two questions about the terms of Table 8. First, is this list complete? Second, have we correctly determined the reference of each term?

Table 8: Reconstructed Proto Bungku-Tolaki kinship terms

Consanguinal		
Direct	*uan	PP, CC
	*ama (r) *t-ama (v)	F
	*ina (r) *t-ina (v)	M
	*kaka (r) *tukaka (v)	eSb
	*hai (r) *tuai (v)	ySb
	*anaq	C
Collateral	*maama	PB
	*naina	PM
	*topisan	PSbC
	*laki-anaq	SbC
Affinal		
	*wali	Sp
	*ponianan	SpP
	*monian	CSp
	*oleo[ ]	SpSb, SbSp
	? *asa mpoqala	SpSbSp
	*awo	PSbSp, SpSbC
	(*tama awo, etc.)	(MSp, FSp, PSpC, SpC)

30 (?) Derived from the stem *tuwu* 'live'.

31 Moronene and Tolaki have *lakiwa* 'widower' (with stem *laki* (?), see §2). Compare also Kulisusu *lako ntama* 'widower', *lako ncina* 'widow'.

In regard to the first question, we need to be clear at the outset that all protolanguages are idealised constructs which are subject to revision, particularly as new or better data becomes available. Furthermore, even when we use the best best data and best linguistic and anthropological theory, there will always be limits. For example, in order to argue for completeness, we must perforce assume that every kinship term in the protolanguage has been continued by at least one present-day language. Furthermore, even though I have used the best *available* data, in many cases our knowledge of present-day kinship systems is sadly lacking. In particular, the following strike me as three areas worthy of more detailed investigation: (a) the original reference of *la-ŋkai and *be-(ŋ)kele (§2) and how these terms spread and entered the kinship systems of various Celebic languages; (b) how terms for 'grandparent' interact with and have been affected by other terms of respect used in the wider social context (§6); and (c) sibling-in-law terms (§11).

On the other hand, insofar as we do know about present-day Bungku-Tolaki kinship systems, there appears to be an overarching similarity between systems, such that even where terms differ formally, nonetheless the same number of terms are used to make the same kinds of distinctions. Assuming that this sameness is a reflection of the proto-kinship system—that is, it is the product of inheritance, rather than simply being a kind of areal feature—then this similarity helps us make the claim that not only is the system complete, but that we have also correctly identified the reference of each term.

A further question to ask would be what these reconstructions might reveal about social organisation in the ancestral society. On this point I will say only that in these terms themselves there is no evidence for the descent groups or the asymmetrical marriage alliances which Blust (1980, 1994) attributes to early Malayo-Polynesian society. Note particularly the *lack* of evidence for cross- or parallel-sibling terms, or for any differentiation of parental siblings according to whether they are on the father's or the mother's side. Indeed, in that collateral terms are invariably *derived* forms (see Table 8), we might even say that the reconstructions point to an earlier time when collaterality was unmarked—the Hawaiian type, such as Murdoch (1949) proposed for Malayo-Polynesian. However, given that we lack the needed ethnographic studies to understand social organisation even in present-day Bungku-Tolaki societies, speculations about their ancestral social organisation would be premature.

The present paper, even as far as it goes, then, is in some ways just scratching at the surface. Nonetheless, the Bungku-Tolaki languages constitute—borrowing a phrase from Joseph Greenberg (1978:83)—a 'laboratory situation' for investigating linguistic and semantic change. I hope in a small way to have added to the body of knowledge about what kind of developments—derivations and semantic extensions—are not only likely and common, but also possible in kinship systems.

References

Adriani, N. 1900. De talen der To Boengkoe en To Mori. *Mededeelingen van wege het* Nederlandsche *Zendelinggenootschap* 44:249–318.

—— 1928. *Bare'e-Nederlandsche woordenboek met Nederlandsche-Bare'e register*. Leiden: E.J. Brill.

Anceaux, J.C. 1987. *Wolio dictionary (Wolio–English–Indonesian)*. Dordrecht: Foris.

Andersen, T. David. 1999. Moronene phonology. *Studies in Sulawesi linguistics, part VI* (*NUSA*: Linguistic Studies of Indonesian and Other Languages in Indonesia, 46), Wyn D. Laidig, ed. 1–45. Jakarta: Badan Penyelegarra Seri NUSA, Universitas Katolik Indonesia Atma Jaya.

Berg, René van den. 1996. *Muna–English dictionary*. KITLV Press.

Bergh, J.D. van den. 1953. *Spraakkunst van het Banggais*. The Hague: Martinus Nijhoff.

Blust, Robert. 1972. Note on PAN *qa(R)(CtT)a 'outsiders, alien people'. *Oceanic Linguistics* 11:166–171.

—— 1979. Proto-Western Malayo-Polynesian vocatives. *Bijdragen tot de Taal-, Land- en Volkenkunde* 135:205–251.

—— 1980. Early Austronesian social organization: the evidence of language. *Current Anthropology* 21:205–247.

—— 1994. Austronesian sibling terms and culture history. In Pawley and Ross, eds, 31–72.

—— In progress. Austronesian comparative dictionary. Computer files, Department of Linguistics, University of Hawai'i.

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.

Esser, S.J. 1927. *Klank- en vormleer van het Morisch, part 1*. (Verhandelingen van het Bataviaasch Genootschap van Kunsten en Wetenschappen, 67/3). Leiden: Vros.

—— 1933. *Klank- en vormleer van het Morisch, part 2*. (Verhandelingen van het Bataviaasch Genootschap van Kunsten en Wetenschappen, 67/4). Bandoeng: Nix.

Fox, James J. 1994. Who's who in ego's generation: Probing the semantics of Malayo-Polynesian kinship classification. In A.K. Pawley and M.D. Ross, eds 127–139.

Gonda, J. 1952. *Sanskrit in Indonesia*. Nagpur: International Academy of Indian Culture.

Greenberg, Joseph H. 1978. Diachrony, synchrony, and language universals. In Joseph H. Greenberg, ed. *Universals of human language,* vol. 1: *Method and theory*, 61–91. Stanford: Stanford University Press.

Himmelmann, Nikolaus P. (compiler.) 2001. *Sourcebook on Tomini-Tolitoli languages: general information and word lists*. Canberra: Pacific Linguistics.

Kruyt, J. 1924. De Moriërs van Tinompo (oostelijk Midden-Celebes). *Bijdragen tot de Taal-, Land-* en *Volkenkunde van Nederlandsch-Indië* 80:33–217. 's-Gravenhage: Martinus Nijhoff.

Laidig, Wyn D. and Sahabu Dg. Maingak. 1999. Barang-barang phonology: a preliminary description. In *Studies in Sulawesi linguistics, part VI* (NUSA Linguistics Studies of Indonesian and Other Languages in Indonesia, vol. 46), Wyn D. Laidig, ed. 46–83. Jakarta: Badan Penyelenggara Seri NUSA, Universitas Katolik Indonesia Atma Jaya.

Lara, Tahe, Sipantu Larobu, Paula Vuorinen and Marjo Karhunen. 1991. Pau mpu'uno: Pau Padoe—Indonesia—Inggirisi (Basic Vocabulary: Padoe—Indonesian—English). Unpublished typescript, 134 pp.

Laskowske, Tom. 2007. The Seko languages of South Sulawesi: a reconstruction. *Studies in Philippine Languages and Cultures* 15:115–210.

Mead, David (compiler). 1995. *The Bungku-Tolaki languages: wordlists*. 2 volumes. Kendari: Universitas Haluoleo.

Mead, David. 1998. Proto–Bungku-Tolaki: reconstruction of its phonology and aspects of its morphosyntax. PhD dissertation, Rice University, Houston, Texas.

—— 1999. *The Bungku-Tolaki languages of south-eastern Sulawesi, Indonesia*. Canberra: Pacific Linguistics.

—— 2003. Evidence for a Celebic supergroup. In John Lynch, ed. *Issues in Austronesian historical phonology*, 115–141. Canberra: Pacific Linguistics.

Mills, Roger F. 1975. Proto South Sulawesi and proto Austronesian phonology. PhD thesis, University of Michigan.

—— 1981. Additional addenda. In *Historical linguistics in Indonesia,* part 1 (NUSA Linguistic Studies in Indonesian and Languages in Indonesia, 10), Robert A. Blust, ed. 59–81. Jakarta: Badan Penyelenggara Seri NUSA, Universitas Katolik Indonesia Atma Jaya.

Murdoch, George Peter. 1959. *Social structure*. New York: Macmillan.

Pawley, A.K. and M.D. Ross, eds. 1994. *Austronesian terminologies: continuity and change*. Canberra: Pacific Linguistics.

Sneddon, J.N. (compiler). 1983. Southern 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 44. Canberra: Pacific Linguistics.

Stokhof, W.A.L., ed. 1985. *Holle lists: vocabularies in languages of Indonesia,* vol. 7/4: *South-east Sulawesi and neighboring islands, west and north-east Sulawesi*. Canberra: Pacific Linguistics.

Stresemann, Erwin. 1927. Die *Lauterscheinungen in den ambonischen Sprachen*. (Beiheft zur Zeitschrift für Eigeborenensprachen, 10.) Berlin: D. Reimer.

Tarimana, Abdurrauf. 1989. *Kebudayaan Tolaki*. (Seri Etnografi Indonesia, 3.) Jakarta: Balai Pustaka.

Tryon, Darrell T., ed. 1995. *Comparative Austronesian dictionary: an introduction to Austronesian studies*. (Trends in Linguistics, Documentation 10.) Berlin: Mouton de Gruyter.

Wurm, S.A. and B. Wilson. 1975. *English finderlist of reconstructions in Austronesian languages* (post-Brandstetter). Canberra: Pacific Linguistics.

29 *Cat's cradle: a disappointing field for lexical reconstruction*

MEREDITH OSMOND

This paper was prompted by a footnote in Kirch and Green's 2001 volume, *Hawaiki, ancestral Polynesia*. The footnote reads (p.301):

> There is a substantial ethnographic literature on the subject of string figures, which are ubiquitous in Polynesia (e.g. Handy 1925; Hornell 1927; Firth and Maude 1970). One aim of this early twentieth-century ethnographic concern with string figures was comparative analysis to infer culture-history (Hornell 1927:6–9). Given the rich lexical data associated with the figures themselves, a renewed analysis informed by a modern phylogenetic approach might well be worth the effort. Blust (pers. comm., 1999) also informs us that such string figures have a very deep history among Austronesian speakers, and that the PPn term **fai* probably derives from PAn **paRiS* ('stingray', term applied to the constellation Scorpio, Southern Cross, or other astronomical features).

String figure games, collectively known as cat's cradle, were played in traditional societies across the world long before recorded time. They have been recorded on every inhabited continent. In this respect I am mindful of Bob Blust's forays into other cultural universals such as the taboos on pointing at rainbows and the thunder complex. The ubiquity of string figures leads to the assumption that they must have been invented independently in a number of places. They are, however, readily learnt and highly borrowable. They are attractive (both in the sense of attracting curiosity, and of visual appeal), they are played frequently by children, and so are within the capabilities of anyone with normal dexterity, they require no equipment apart from a length of string, and they can be transmitted with minimal language, so language barriers do not impede borrowing.

Early last century, ethnographers like Rivers and Haddon (1902), Jenness (1920), Handy (1925), Hornell (1927), Wedgwood (1932–33), Rosser (1932) collected patterns of string games in various parts of the Pacific, recognising that the same patterns recurred in different parts of the world and believing that there was something of their past history to be gained by comparison of these patterns. Handy, for instance, noting identical patterns that occurred not only between the Marquesas, Tahiti, and the Caroline Islands, but also in far flung locations such as the Philippines and Queensland, suggested that 'patterns handed

Alexander Adelaar and Andrew Pawley, eds
Austronesian historical linguistics and culture history: a festschrift for Robert Blust, 509–514.
Pacific Linguistics, 2009.
© This edition vested with Pacific Linguistics.

down from generation to generation may very often offer sound data for determining past cultural relationships' (1925:8).

The purpose of this paper is to examine the nature and scope of the lexical data for string games within Oceanic languages to see whether there are cognate sets that support reconstructions to POc or other interstages. An immediate obstacle to identifying POc level reconstructions is that while we have good language-specific descriptions from the Central Pacific, and a smaller amount from the Solomons and Vanuatu, our sources from Papua New Guinea are meagre, with Jenness's Bwaidoga terms (Jenness 1920) our only useful published data. Of the other New Guinea ethnographers listed in the references, Wedgwood (Maude and Wedgwood 1967) and Rosser (Rosser and Hornell 1932) name and describe patterns, but only loosely define region of collection, do not specify language, or distinguish Austronesian from non-Austronesian. Although Noble (1979) describes and names 140 patterns from South East Papua and the Highlands, all bar a couple are from non-Austronesian-speaking areas. Dictionaries and some of the more extensive wordlists add a number of additional terms, but here no descriptions of the patterns are included.

In the language-specific descriptions, a generic term is typically given for the activity, one that is frequently both noun and verb. All descriptions contain names and descriptions of dozens of patterns. A few of them include terms for particular moves or sets of moves frequently used. Several also record the chants that typically accompanied the making of particular patterns.

1 Generic term

To cognates of the numerous Polynesian terms supporting the PPn reconstruction **fai*, can be added Bauan Fijian *vei* and possibly, and more significantly, a Motu term, *hari(kau)*, all generic terms. The meaning of the bracketed form *-kau* is unclear but it is included in several net-related terms in Motu, an association of meaning which is echoed in the Tahitian and Hawaiian terms below. Although we need cognates from other subgroups to strengthen the reconstruction, PAn **paRiS* 'stingray' is, as suggested by Blust, a plausible antecedent for the generic term for cat's cradle at POc level. A stingray is roughly diamond-shaped. Probably the most common patterns created in cat's cradle also contain one or more diamond shapes. Reflexes of PAn **paRiS* at times refer to both a stingray and to a constellation seen as stingray-shaped, usually the Southern Cross or Scorpio. It may be relevant that both fish shapes and constellation shapes are well represented in the names of particular Oceanic cat's cradle patterns.

No non-Oceanic reflexes of PAn **paRiS* with the meaning 'cat's cradle' have been located.

Table 1: POc **paRi* 'generic term for cat's cradle' (also 'stingray')

PT:	Motu	*hari(kau)*	'cat's cradle' (*kau* meaning unclear, but included in a number of net-related terms)
Fij:	Bauan	*vei (saŋa)*	'general term for cat's cradle when using both hands and feet' (*saŋa* 'crotch, fork')
		vei (ciu)	'cat's cradle with hands alone' (*ciu* 'carved, cut to a shape')

Table 2: PPn **fai* 'cat's cradle, string games; play at cat's cradle'

Pn:	Tongan	*fai*	'cat's cradle'
Pn:	Pukapukan	*wai-wai*	'cat's cradle; to make string figures'
Pn:	Tuamotuan	*fai*	'string games, cat's cradle'
Pn:	Tahitian	*fai*	'name of a game played by children; string game, cat's cradle' (Same word as used for meshes of sorcerer's net – Handy:6)
Pn:	Maori	*whai*	'string game, cat's cradle; play at cat's cradle'
Pn:	Hawaiian	*hei*	'cat's cradle; to make string figures' (Also 'net, snare; to ensnare, entangle')

Generic terms for the activity have been collected from all major Oceanic subgroups with the exception of the Admiralties, and that may well be an accidental gap in wordlists. They include: NNG: Lukep Pono *barau*, Manam *kinarua*, Gedaged *mol-mol*; PT: Bwaidoga *giwala*, Dobu *ʔabi*, Kiriwina *ninikula*; MM: Tolai *weweuk*; SES: Gela *hoŋgo*, Tolo *tinabe*, Arosi *raʔuna*; NCV: Ambae *lelegaro*; Mic: Kiribati *wau*; Pn: Tikopia *siko-siko*, Marquesan *pehe*. No cognate sets for generic terms other than **paRi* have been noted except in the Southeast Solomons where three Malaitan languages, 'Are'are, Sa'a and Ulawa all use the term *isu-isu* 'play at cat's cradle'. *isu* also means 'to count, to read' the shared meaning evidently carrying the sense of 'relating in some sort of sequence'. De Coppet, who has collected string figures from 'Are'are writes that there are stories one can tell in doing the figures (p.xvii, preface to Maude 1978). Evidently, successive patterns are associated with particular parts of the story. To move through the stages of a particular pattern is in effect to tell its story. This attribution of an ulterior purpose to the playing of cat's cradle may be the last traces of a function formerly widespread but now all but lost. It is mentioned by Noble in his description of patterns known in the Managalas-speaking (non-Austronesian) regions of South East Papua. He writes that string games 'appear to have served a purpose as a repository of traditional knowledge with regard to hunting, gardening, building; a kind of primitive picture folk lore. Warnings are implicit in the snake bite, the tree that falls down, and the danger of letting the dog get into the bush hen's nest to break the eggs' (Noble 1979: foreword). He believes further that this associated meaning is usually the first aspect of a game to be lost.

2 Names of patterns

Although dozens of terms have been listed for particular patterns, they are a dubious source for reconstruction because attachment of term to pattern is so subjective. Patterns are seen as resembling objects familiar to the community, and thus named by the local term for the object. In the Oceanic world, motifs like crab, fishing net, bowl, the setting sun and so on recur. But a pattern seen as a food bowl in Tikopia may be labelled a well in Lifu and a basket in Bwaidoga. Conversely, although pattern names such as the local term for the Pleiades occur in various widespread languages including Bwaidoga (*yavunuga*), the Marquesas (*mataʔiki*), and the Gilbert Islands (*nei auti*), the patterns so labelled may differ substantially. Named patterns like 'Ten Men' (Caroline Islands), 'Navel of Maui'

(Tuamotus), 'Woman showing her backside' ('Are'are) or 'Man crazed by betelnut' (Bwaidoga) are more likely to be local inventions, referring to local identities, historical events or legends. The name may be little more than a rough clue as to shape, and possibly relevant only within a community. If it is the shape of the pattern itself rather than its label that is the clue to its traceable cultural history, this is surely a matter for semiotics rather than linguistics. A possible answer for linguists lies in the third kind of lexical data, terms given for the various moves involved in the creation of multiple shapes.

3 Starting positions and other moves

What seized ethnographers' interest was the evident recurrence not just of completed patterns but also particular moves, or sets of moves. They looked at traditional starting positions, opening gambits and recurring moves. Although there may be no theoretical limit to the number of patterns possible, in practical terms games tend to start with a few basic patterns, and proceed using a variety of familiar moves. These would evolve through trial and error, the more attractive, or the simpler the operation for a pleasing result, the more likely to be retained and passed on. Davidson (1941), in an examination of string games among Australian aborigines, noticed that people in different regions sometimes favoured different ways of reaching the same pattern or the same stage in a pattern, and suggested that an expert might be able to identify where a game was played by examining its moves.

Comparison of patterns was facilitated by Rivers and Haddon who, in 1902, published a method of recording string figures and tricks. They compiled a small standardised lexicon of English terms. Included were terms like 'Position 1' and 'Opening A', to which others have been added—'Navaho thumbs', 'Caroline extension', 'Murray Opening', and so on. 'Position 1' and 'Opening A' occur in every collection we have of Oceanic string games, albeit not in every pattern recorded. Unfortunately we have minimal local terms for these. Handy (1925:9,10) lists Tahitian names for 'Position 1' (*e fai*) and 'Opening A' (*e tui*) together with half a dozen Tahitian terms for other moves. They are *taviri* 'any twisting movement either of digits or strings', *e pana* 'to pick up a string or loop on the dorsal side of a finger', *e iti* 'to lift the proximal thumb loop with the mouth and drop it between the thumb and index finger', *e tuʔu* 'to exchange loops between fingers or withdrawing a finger from the figure', *iriti* 'to remove a loop from one digit to another', *taume* 'to pull a string down on the palmar side of a finger', *taamu* 'to wrap a string round a finger'. Jenness recorded two movements that recurred with great frequency in Bwaidoga, *nauwa* and *luatataga*, although neither are opening moves. The Tikopia utilise two movements, called *tao* and *ta*, in nearly half their figures 'thus providing a characteristic technique pattern not hitherto found to be predominant in any other locality' (Firth and Maude 1970:9).

Other than these, no terms for moves or sets of moves have been located in the Oceanic literature. Moreover, none of the moves named by Rivers and Haddon are restricted to the Oceanic region. Honor Maude, recognised early as the authority on Oceanic string figures, writes that 'Position 1' and 'Opening A' 'are the most common openings in almost any part of the world' (Maude and Wedgwood 1967:203). 'Navaho thumbs', widespread in Oceanic patterns, is so named because the movement is a familiar one among the string games of the Navaho Indians. The 'Caroline Extension', widespread in Oceania except for peripheral regions of Polynesia (Marquesas, Tahiti, Tuamotus, Hawaii) is also well represented in Australian patterns where it is called 'Pindiki' (Davidson 1941). In their

account of the pastime in the Gilbert Islands, Maude and Maude write that repeated questioning failed to elicit any native terms for the commoner movements involved in construction of the figures (1958:3).

4 Chants

A limited examination of the chants which accompanied the making of a considerable number of the string figures as a source of lexical data has also proved fruitless. Collectively these accompaniments are known as *vinvina* in Kiriwina, *patautau* in Tahitian and *haʔa-naunau* in Marquesan. Although a number of early ethnographers recorded particular chants and in some cases attempted translations, many are described by their own speakers as untranslatable, perhaps because they have been borrowed from a language unknown to the player, have become distorted through transmission over time, or are simply meaningless jingles of the fol de rol variety. Handy (1925:10) suggests that 'so hazy is the native memory regarding the ancient legends and tales whose events and characters are referred to in a fragmentary way in these sing-song jingles, that few of them could be explained'.

5 Conclusion

Overall then, very little has been gained by an attempt to use the terminology of cat's cradle as a tool in tracing culture history. On distributional grounds we can be confident that speakers of Proto Oceanic played cat's cradle, but we can say little more than that. The largest obstacle in reconstructing higher-level terms is undoubtedly the dearth of recorded terms to do with the activity. And the data with perhaps the greatest potential for reconstructing culture history, terms for standardised or most frequently used moves, may turn out to be little more than an artefact to facilitate ethnographers' descriptions of pattern construction. If terms for opening gambits or frequent moves were ever in widespread use among the players, they have disappeared almost without trace.

Other considerations that weigh heavily against the regular transmission of old terms from generation to generation include the proclivity of the activity for borrowing, mentioned earlier, and the game's potential for creativity. Most of the descriptions available to us mention that string games also provide an outlet for creativity, sometimes with overtones of ridicule or humour. As an example, string games in Tikopia constituted a living art in which new figures were invented or traditional ones modified. Firth (1970:5) writes that 'One of the most striking features of Tikopia string figures is the overt recognition of inventiveness in them ... As in their songs so in their string figure they have recognised individual creativity by the attribution to specific persons of responsibility for particular examples'. The Nauruans in recent times have been encouraged to reinvigorate an almost forgotten art, and have added many previously unknown variations to those recalled by older people (Maude 2001). Variations to the basic figures are frequently mentioned. Creating new figures is an admired activity.

It is tantalising to believe that there may have been substantial folklore bound up in the nature of the patterns, the names of the patterns, the stories behind them, and the chants that accompany them, but such knowledge is at this stage beyond recording. The playing of string games is a disappearing art, submerged under the increasing exposure to western-style activities and sources of entertainment. Most of the traditional terms and associated

rituals were disappearing or were already lost when ethnographers first became interested in their collection early last century. Any re-emergence of the activity as in Nauru will include new patterns and associations with non-traditional objects. So although it seems entirely possible that string games and frequently used moves were known to Proto Oceanic speakers, it is unlikely that comparative linguistics can ever offer lexical proof beyond the somewhat tentative reconstruction of POc **paRi* as the generic term for string games and the activity of playing them.

References

Davidson, D.S. 1941. Aboriginal Australian string figures. *Proceedings of the American Philosophical Society* 84(6):763–901.

de Coppet, Christa. 1978. Preface to Honor Maude, *Solomon Islands string figures.* xvii-xix.

Emory, K.P. and Honor Maude. 1979. *String figures of the Tuamotus*. Canberra: The Home Press.

Firth, Raymond and Honor Maude. 1970. *Tikopia string figures*. London: Royal Anthropological Institute of Great Britain and Ireland.

Haddon, K. 1930. *Artists in string*. London: Methuen and Co. Ltd.

Handy, Willowdean Chatterton. 1925. String figures from the Marquesas and Society Islands. *Bernice P. Bishop Museum Bulletin* 18:1–92.

Hornell, James. 1927. String figures from Fiji and Western Polynesia. *Bernice P. Bishop Museum Bulletin* 39:1–88.

Jenness, D. 1920. Papuan cat's cradle. *Journal of the Royal Anthropological Institute* 50:299–326.

Kirch, Patrick and Roger C. Green. 2001. *Hawaiki, Ancestral Polynesia*. Cambridge: Cambridge University Press.

Maude, Honor. 1978. *Solomon Islands string figures*. Canberra: The Home Press.

—— 1984. *String figures from New Caledonia and the Loyalty Islands*. Canberra: The Home Press.

—— 2001 (revised first edition 1971 from fieldnotes made 1937–38). *The string figures of Nauru Island*. The University of the South Pacific Centre in Nauru and Institute of Pacific Studies.

Maude, H.C. and H.E. Maude. 1958. *String figures from the Gilbert Islands*. Wellington: The Polynesian Society.

Maude, Honor and Camilla Wedgwood. 1967. String figures from Northern New Guinea. *Oceania* 37:202–229.

Noble, P.D. 1979. *String figures of Papua New Guinea*. Boroko: Institute of Papua New Guinea Studies.

Rivers, W.H.R. and A.C. Haddon. 1902. A method of recording string figures and tricks. In *Man* 1–2:146–153.

Rosser, W.E. and J. Hornell. 1932. String figures from British New Guinea. *Journal of the Royal Anthropological Institute* 62:39–50.

30 *The role of the Solomon Islands in the first settlement of Remote Oceania: bringing linguistic evidence to an archaeological debate*

ANDREW PAWLEY

1 Introduction

This paper looks at some problematic aspects of the history of human settlement of the Solomon Islands over the last three millennia.[1] The initial spread of Oceanic languages into Remote Oceania[2] can be strongly associated with the movement into the Reefs/Santa Cruz group and Vanuatu, at about 3200–3100 BP, of bearers of the archaeological culture known as Lapita. Lapita is first attested in the Bismarck archipelago and on geographic grounds one would expect the islands in the main Solomons group (extending from Bougainville to Makira) to have been stepping stones for the Lapita expansion eastwards into Remote Oceania. Thus, archaeologists have been puzzled as to why no early Lapita archaeological sites been found in the main Solomons group, and why almost no pottery-bearing sites of any kind have been found in the southeastern part of the group. Does this mean that the main Solomons group was bypassed in the initial Lapita colonisation of Remote Oceania, as was suggested by Sheppard and Walter (2006), or is the archaeological record too fragmentary to allow any firm conclusions to be drawn?

[1] I am delighted to contribute to a volume honouring Bob Blust's distinguished and diverse contributions to Austronesian historical linguistics and culture history. An earlier version of this paper was presented at the 7th International Conference on Oceanic Linguistics, Noumea, July 2007. The paper has benefited from discussions with Roger Green, Stuart Bedford, Bethwyn Evans, Frank Lichtenberk, Malcolm Ross, Matthew Spriggs and Darrell Tryon.

[2] Whereas 'Near Oceania' consists of New Guinea, the Bismarck Archipelago and the main Solomons Archipelago (ending at Makira), which form a chain of largely intervisible islands, 'Remote Oceania' consists of the remaining, much more widely dispersed islands and island groups of the SW and Central Pacific (chiefly those of Vanuatu, New Caledonia, Fiji, Micronesia and Polynesia).

Alexander Adelaar and Andrew Pawley, eds
Austronesian historical linguistics and culture history: a festschrift for Robert Blust, 515–540.
Pacific Linguistics, 2009.
© This edition vested with Pacific Linguistics.

I will address three questions concerning the history of the Oceanic languages of the Solomons that have a bearing on this issue:

(1) Given that there is no major geographic barrier that would account for an early and sharp separation of these subgroups, what circumstances created the major subgroup boundary that runs through the centre of the Solomons archipelago, separating Northwest Solomonic from Southeast Solomonic?

(2) How long have the Northwest and Southeast Solomonic groups been in their present locations?

(3) Why have the Northwest Solomonic languages replaced a much higher percentage of Proto Oceanic core basic vocabulary items than Southeast Solomonic languages?

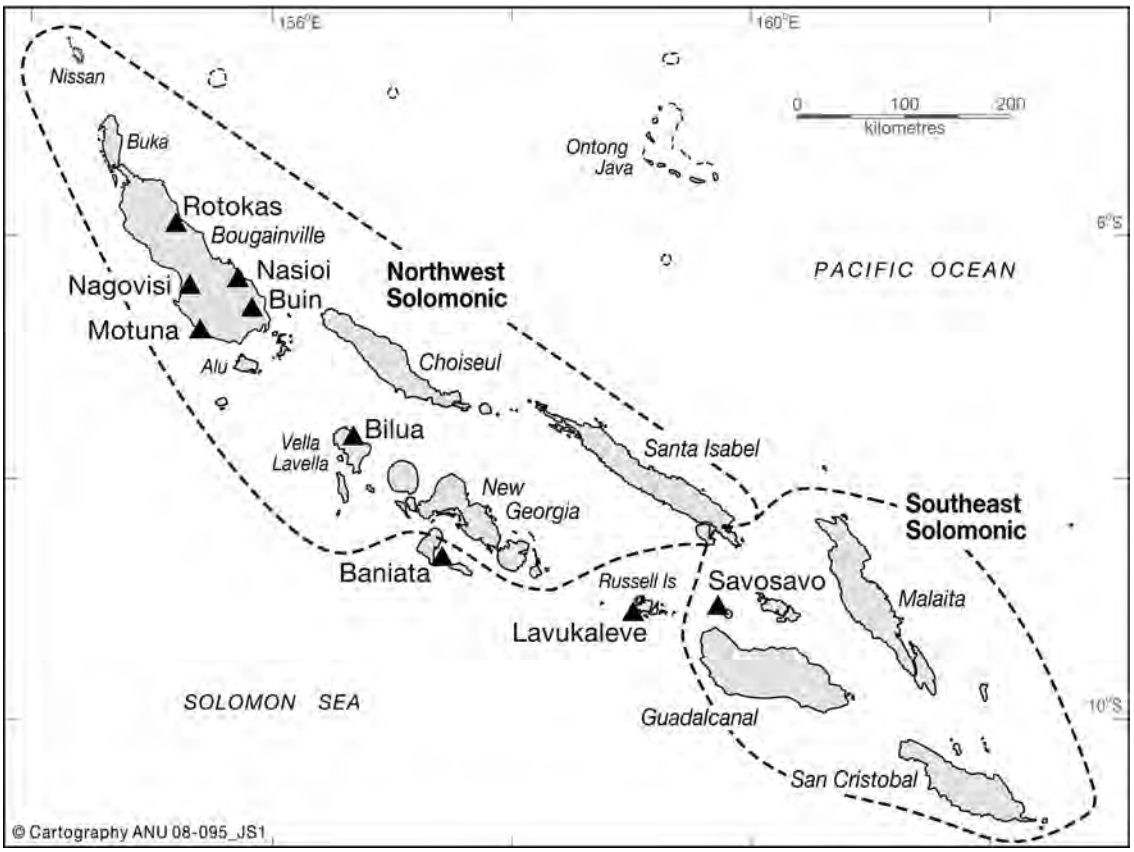


Figure 1: Boundaries of Northwest Solomonic and Southeast Solomonic and locations of non-Austronesian languages

2 The Solomons archipelago

Because the geographic span of the main group of Solomon Islands differs markedly from that of the nation called 'the Solomon Islands' I will refer to the former as 'the Solomons archipelago' or 'the main Solomons group'. The archipelago consists of a chain of closely spaced large islands that extends for about 1000 km from northwest to southeast (see Figure 1). The main islands are quite large: Bougainville is about 10,000 square

kilometres, Guadalcanal 6500, Makira 4600, Malaita 3900, Choiseul 3000, and New Georgia 2100. All the large islands are mountainous and heavily forested. Typically there is a narrow coastal strip of strand forest of sandy soil with light forest of salt-resistant trees and patches of mangrove and sago swamp. Man-made grasslands occur in some areas, most extensively in the plains of northern Guadalcanal. In several regions there are extensive fringing coral reefs and lagoons carrying a rich biota.

It will be convenient to distinguish between a Northwest Solomons region, including Buka, Bougainville, Choiseul, the New Georgia group and Santa Isabel, and a Southeast Solomons region, including Guadalcanal, Florida, Malaita and Makira. Buka and Bougainville are separated from New Ireland by an ocean gap of 180 km, with only the small island group of Nissan (aka Nehan or the Green Is.) in between. Some 400 km of open sea separate Makira from the small Santa Cruz-Reef Is. group to the east.

Humans reached New Guinea, New Britain and New Ireland by 40,000 years ago and by about 30,000 years ago had settled Greater Bougainville in the NW Solomons (Kirch 2000; Spriggs 1997; Specht 2005) at a time of lower sea levels, when the island of Bougainville extended from what is now Buka almost to Guadalcanal. However, the Solomons archipelago remained the limit of human expansion into the Southwest Pacific until just over 3000 years ago. Until then it appears that people lacked sailing craft capable of making the long crossings to islands further east, against the prevailing trade winds and currents.

3 The spread of Lapita as a marker of the dispersal of Oceanic languages

In the second half of the 2nd millennium BC people bearing a new language and technology entered Northwest Melanesia. These were fishermen–farmers from Southeast Asia, who by 3400–3300 BP had settled in various parts of the Bismarck Archipelago, chiefly on small islands, where they established the first nucleated villages known in Melanesia (Green 2003; Kirch 2000; Specht 2005; Spriggs 1997; Summerhayes 2000, 2001). The most visible archaeological marker of this Neolithic culture is its highly distinctive decorated pottery, with elaborated motifs impressed by dentate-stamping. In sites representing permanent habitations the decorated pottery is part of a cluster of distinctive elements: settlement patterns, rectangular houses raised in stilts, an array of ceramic vessel forms, mainly undecorated, fishing gear, adze/axe kit, shell ornaments and evidence of long distance exchange of obsidian. The pottery tradition is known as Lapita, after which the cultural complex as a whole is named. Changes in the styles and proportions of decorated pots lend themselves to the construction of a fine-grained seriation chronology which can supplement C^{14} dating of Lapita assemblages. Many elements of the Lapita complex have close parallels in Neolithic cultures that appear in Taiwan, the Philippines and the Marianas and parts of Indonesia in the early to mid 2nd millennium BC (Bellwood 1997; Bellwood and Dixon 2005; Green 2003; Kirch 1997, 2000).

The sudden appearance of this distinctive cultural complex in the Bismarck Archipelago can be strongly associated with the arrival there of Austronesian languages, and specifically with the separation of the large Oceanic branch from its nearest relatives, spoken in the Cenderawasih Bay area at the western end of New Guinea, and in South Halmahera (Blust 1978a). Oceanic is a well-defined subgroup which contains all the Austronesian languages of Melanesia except the western end of New Guinea, plus those of Polynesia and (with two exceptions) Micronesia. The lexicon of Proto Oceanic has been

reconstructed in considerable detail (Ross et al. 1998–2008, in prep.) and, when compared with the lexicon reconstructed for Proto Malayo-Polynesian (Blust 1995) shows a fairly high degree of continuity in terminologies for various domains of material culture and social organization (Green 2003; Pawley 2007).

The earliest attested phase of Lapita in the Bismarcks is known as Early Western Lapita, which appears between 3400 and 3300 BP. Around 3200 BP or soon after bearers of the Early Western Lapita culture moved east of the Bismarck Archipelago into Remote Oceania. The Reefs/Santa Cruz group, some 400 km east of Makira, contains one of the earliest and most extensively excavated Lapita sites in Remote Oceania. Site SZ–8R, with initial occupation dated to between 3200–3100 BP (Green 1991, 2003, pers. comm.) is among 19 Lapita sites in Reefs/Santa Cruz. For some time the Lapita occupants of this group kept importing considerable quantities of obsidian from Talasea in New Britain, an indication that initially they maintained trade links with the homeland. Some Talasea obsidian appears in early Northern Vanuatu Lapita sites, a strong indication that this region was settled at about the same time as Reefs/Santa Cruz (Bedford 2003, Bedford et al. 2006). By 3050 BP, Lapita people had occupied New Caledonia (Sand 2001) and Fiji (Nunn et al. 2004). By 2950 BP they were in Tonga (Burley et al. 2007) and by 2800–2700 BP they were in Samoa and some of the other islands in the Tonga–Samoa voyaging corridor (Kirch 1997; Green 2003). In each of these island groups in Remote Oceania the distinctive Lapita decorated ware disappeared within a few centuries of first settlement but in most regions some other features of the Lapita cultural complex including, as a rule, the plain ware ceramic vessel forms, continued for much longer.

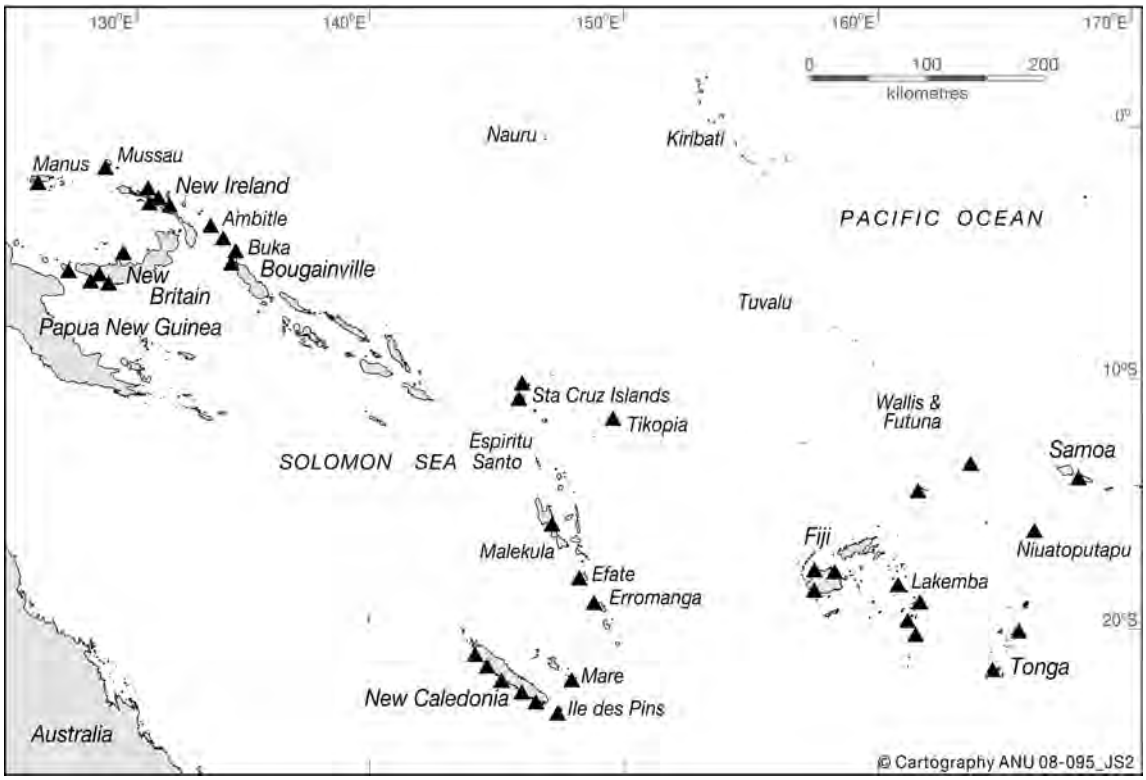


Figure 2: The distribution of important Lapita sites (after Spriggs 1995:113)

It appears that there was a pause of about 200–300 years in the Bismarcks before bearers of the Lapita culture moved eastwards into Remote Oceania. The final stages in the development of Proto Oceanic (POc) can be associated with this pause (Blust 1998; Pawley 2003a, 2008). The initial eastward migrations of Lapita people mark the spread of Oceanic languages into Remote Oceania. All but two of the 180–190 indigenous languages spoken in Remote Oceania at time of first European contact belong to the Oceanic subgroup. The two exceptions are two languages on the western margin of Micronesia, Chamorro and Palauan; both are Austronesian languages that probably stem from movements out of the Philippines or Indonesia before 3000 BP

4 Archaeological debates over Lapita settlement of the Solomons archipelago

Given the position and size of the main Solomons group one would expect colonies to have been established there during the first Lapita movements eastward from the Bismarcks. However, although Early Western Lapita sites have been found immediately to the west of the Solomons, on Nissan (3200 BP, Summerhayes 2000, 2001), and slightly later sites on Buka (3000 BP, Wickler 2001), no Lapita sites associated with the initial Lapita expansion of 3200–3000 BP have so far been identified in the main Solomons group east of Buka. The nearest approximations are various sites in the New Georgia group, chiefly in the Roviana Lagoon, which contain the remnants of stilt-house settlements built over the intertidal zone. These are evidenced by residues of potsherds and some stone tools in shallow water, one to two metres below the surface (Felgate 2001, 2003, 2007). The Roviana Lagoon sites are dated by seriation chronologies of ceramic styles as being late Lapita, around 2700–2400 BP.

The absence of Early Western Lapita pottery from the NW Solomons, and the almost complete absence of any pottery finds in the SE Solomons, has led to a lively debate among archaeologists about the role of the Solomons archipelago in the early Lapita settlement of Remote Oceania. Two competing sets of proposals have emerged, which I will refer to as the ‘early settlement’ and ‘late settlement’ hypotheses.

In a recent review of Solomons archaeology Sheppard and Walter (2006) put forward the following proposals:

(i) The early Lapita colonists leapfrogged the main Solomons group, moving directly to the Reefs/Santa Cruz Is. about 3200–3100 BP. (A similar proposal had been adumbrated by Roe 1993.) For a time the Reefs/Santa Cruz settlers maintained long distance obsidian trade connections with the Bismarck archipelago, as well as obtaining chert from Malaita or Ulawa and basalt for adzes from southeast Guadalcanal.

(ii) Several centuries later, ca 2700 BP, the NW Solomons were settled by Austronesian-speaking, farming, pottery-making populations who moved from the west (the Bismarcks) and whose languages in time became dominant over the non-Austronesian autochthonous languages.

(iii) More tentatively, they propose that Austronesian speakers did not settle the southeastern islands in the main Solomons chain (Guadalcanal, Malaita and Makira) until some 800–1000 years after the initial Lapita dispersal into Remote Oceania. Around 2300–2200 BP, these islands were settled by an a-ceramic, farming population coming from the Reefs/Santa Cruz group and/or Utupua and Vanikoro, where manufacture of pottery ceased about 2100 BP.

This scenario would of course explain the sharp linguistic boundary between the NW and SE Solomonic groups.

Felgate (2001, 2003, 2007) takes a more cautious view regarding the absence of early Lapita sites in the NW Solomons. He suggests that early Lapita occupation of the NW Solomons is likely to have been low density, because of the presence there of established non-Austronesian populations and perhaps because of malaria. He points out that archaeological surveys there have been mainly terrestrial, whereas Lapita settlements are likely to have consisted of stilt houses built over the edge of the lagoon, a pattern attested for late Lapita sites in the New Georgia region, as it is for a number of regions further west (Kirch 2000; Spriggs 1997). Felgate (2001:57) favours the view that:

> a pattern of intertidal settlement [in the Lapita period] has created the dual conditions of low site preservation/visibility and unexpected site location. Implicit in this proposition is a suggestion that early Lapita may have been continuously distributed across the Near Oceanic Solomon Islands in the past, as a shifting network of interacting settlements, located exclusively over the tidal zone, of which we are likely to find only rare traces in settings favourable to their preservation.

Felgate's critics feel that he overstates the domination of intertidal sites in the Lapita settlement of the New Georgia group. Sheppard has recently reanalysed the geomorphic context of inter-tidal sites there and concludes that it is unlikely that an Early Lapita record has been obliterated by submersion (Sheppard pers. comm.). Insofar as there is a consensus on this matter, it is that the earliest material in the Roviana Lagoon dates to around 2700 BP and represents the late end of dentate-stamped pottery, after which decorations on pots were made using a different technique.

Archaeological surveys of the SE Solomons from Guadalcanal to Makira have so far found almost no ceramics. This stands in sharp contrast with the NW Solomons, where pot sherds are highly visible on all the main islands, and it is clear that pots continued to be made long after the Lapita period. The pollen record for Guadacanal gives evidence of intensive slash and burn horticulture there beginning around 2300–2200 BP (Haberle 1996; Roe 1993) and the faunal record also points to increased predation and extinction of larger species about that time (Spriggs 1997). Comparing these indicators of the first appearance of large scale shifting argriculture in Guadacanal with earlier dates for similar signs in Aneityum and New Caledonia, Spriggs (1997:149) comments '[t]he nearly 800 year time lag on Guadalcanal and the lack of pottery in any of the sites so far investigated suggests that Austronesian settlement here was delayed until pottery was no longer in use in the region'.

However, there is reason to think this suggestion is premature. The best surveyed of the main islands in the SE Solomons is Guadalcanal but even there the archaeological record is poor. Malaita remains virtually an archaeological blank. A few small excavations have been carried out on Makira, Uki and and Ulawa, yielding no pottery or early dates. The solitary exception is a rock shelter on Santa Ana which contained plain (undecorated) ware ceramics of late Lapita type, dating to about 2900 BP (Green pers. comm.).

While it seems clear that the inhabitants of the SE Solomons have not made pottery during the past 2000 years, the scarcity of Lapita pottery in a region with a poor archaeological record should not necessarily be taken to indicate that the rest of the Lapita cultural complex was also absent. While pottery is an invaluable aid in finding sites and in dating assemblages, it was just one component in a rich Lapita cultural tradition. Phases 2 and 3 of Vatulama Posovi, a cave site in the Poha Valley, near Honiara on Guadalcanal,

have yielded an assemblage of artefacts dated to around 3250–2900 BP and 2750–2550 BP which has been described as 'Lapita without pots' (Roe 1993). Around 3000 BP the Lapita settlers of Reefs/Santa Cruz were importing basalt for adzes from Marau Sound on SE Guadalcanal, chert for blades from Ulawa and/or Malaita and temper for pots from part of the Florida group, off N. Guadalcanal, and it would be strange if they did not establish settlements or interact with sister Lapita colonists in these places. The Santa Ana rock shelter site is presumably the byproduct of one such settlement.

In the sections that follow I will discuss some linguistic evidence that bears on these archaeological issues.

5 The language groups of the Solomons Archipelago

5.1 Overview

In many cases dialect chaining makes it hard to draw language boundaries without some degree of arbitrariness, but on a conservative estimate there are 60 or so mutually-unintelligible languages spoken in the Solomons archipelago. Some 50 of these languages belong to the large Oceanic branch of Austronesian. Another 12 or so are non-Austronesian ('Papuan') and fall into at least four different families that cannot, on present evidence, be convincingly shown to share a common origin (Ross 2001; Dunn et al. 2002, 2005).

Except on Bougainville, where they occur in coastal pockets, the Oceanic languages in the Solomons have a continuous distribution over all the habitable parts of the larger islands. Two major subgroups of Oceanic are represented there: Northwest Solomonic and Southeast Solomonic. The boundary between them runs roughly north–south between Santa Isabel in the west, and Guadalcanal and Malaita in the east. SE Solomonic languages are spoken on Guadalcanal and the Florida group, Makira, and Malaita. A single SE Solomonic language, Bugotu, is spoken on the south-eastern tip of Santa Isabel, where it is clearly represents an intrusive settlement from the Florida group or Guadalcanal within the last 1000 years. NW Solomonic comprises the Oceanic languages of Santa Isabel (other than Bugotu), the New Georgia group, Choiseul, Bougainville, Buka and the small Nissan island group which lies between New Ireland and Buka.

The few surviving non-Austronesian languages in the Solomons Archipelago are plainly the residue of a larger number that were present in this region when speakers of Oceanic Austronesian arrived. The surviving languages are genetically very diverse (Ross 2001; Dunn et al. 2002). According to Ross (2001), Bougainville contains two families of non-Austronesian languages with four members each. There are two non-Austronesian languages in the New Georgia group and two occupying the small islands of Russell and Savo to the northwest of Guadalcanal.[3]

Presumably, non-Austronesian languages were once spoken on all the main Solomon islands at least as far east as Guadalcanal, and possibly on Malaita and Makira as well. The pre-Austronesian populations were probably mobile foragers and this mode of life, in combination with the rugged and densely forested nature of the islands, and the lack of large terrestrial animals to hunt and, in some islands, the scarcity of fringing reefs, would have severely limited their numbers and distribution.

---

[3] The non-Polynesian languages of Santa Cruz, and Äiwoo of the Reefs, have sometimes been classified as non-Austronesian but recent work has strengthened the case made in Lincoln (1978) that they are Oceanic languages that have undergone an unusual amount of phonological and morphological change. It is likely that they fall together in a single first-order subgroup of Nuclear Oceanic (Ross and Næss 2007).

5.2 Southeast Solomonic and its subgroups

5.2.1 Southeast Solomonic

The existence of a SE Solomonic (SES) subgroup is uncontroversial. Milke (1958) and Grace (1959) observed that this group is defined by the merger of POc *l and *R, an unusual merger in the Austronesian family. A larger body of morphological innovations defining SE Solomonic was set forth in Pawley (1972), e.g. development of a special suffix marking inanimate 3rd person plural pronouns: Proto SES *-ki (direct object), *-ni possessor; replacement of POc preverbal subject markers *ku '1SG', *ko '2SG', *na '3SG' by Proto SES *u, *o and *e; replacement of the POc possessive pronoun *-da '1inc.pl.' by the independent form *-kita, used as a possessive.

However, the quantity of shared innovations defining SES is quite small. This indicates that the period of unified development of SES after it diverged from other Oceanic languages was no more than a few centuries, after which its two primary subgroups, Makira-Malaitan and Guadalcanal-Gelic, began to diverge.

5.2.2 Makira-Malaitan

Makira-Malaitan (MkMl) consists of some 13 languages. Seven are spoken on Malaita and its satellites (including Ulawa and Ugi, lying between Malaita and Makira), four on Makira, and two at the eastern end of Guadalcanal (the latter are both clearly intruders from Malaita or Makira). This subgroup is marked by a number of changes to the Proto SES sound system (Lichtenberk 1988, 1994; Pawley 1972; Tryon and Hackman 1983): *t was lost in Proto MkMl, *s > *t except before high vowels, *k > glottal stop in most cases and there was accretion of a prothetic consonant *y- before initial *a. There are also a few irregular changes in particular grammatical forms. POc *-kita '1inc.pl.' in Proto MkMl reduced to *-ka (presumably via *kia, after regular loss of *t). The Pre MkMl 1st inclusive trial form *kita-tolu reduced to *kaolu, and the 1st exclusive trial form reduced from *kami-tolu to Proto MkMl *ʔamelu.

From the pattern of overlapping isoglosses it is pretty clear that Proto MkMl persisted for many centuries as a chain of dialects extending over both Malaita and Makira (Lichtenberk 1988, 1994). While the geographic extremes in this chain began to diverge very early they remained connected by intermediate dialects. (See §8 for further discussion.)

5.2.3 Guadalalcanal-Gelic

Guadalalcanal-Gelic (GG) contains about seven languages. On Guadalcanal (where dialect chaining complicates the count) there are perhaps five languages. There is one (Gela) in the Florida group and another, Bugotu, is spoken at the eastern end of Santa Isabel.

Two phonological innovations mark GG: POc *w is lost in word initial position; *m and *mw merge as *m*. There are a few morphophonemic or irregular phonological changes, e.g. when certain disyllabic roots are reduplicated the second consonant drops out in the first root, e.g. Gela *taitahi* 'salt' instead of **tahitahi*. Proto SES *no- 'marker of general possessive relation' irregularly became Proto GG *ni-. It is clear that Proto GG was spoken on Guadalcanal and probably also on Florida.

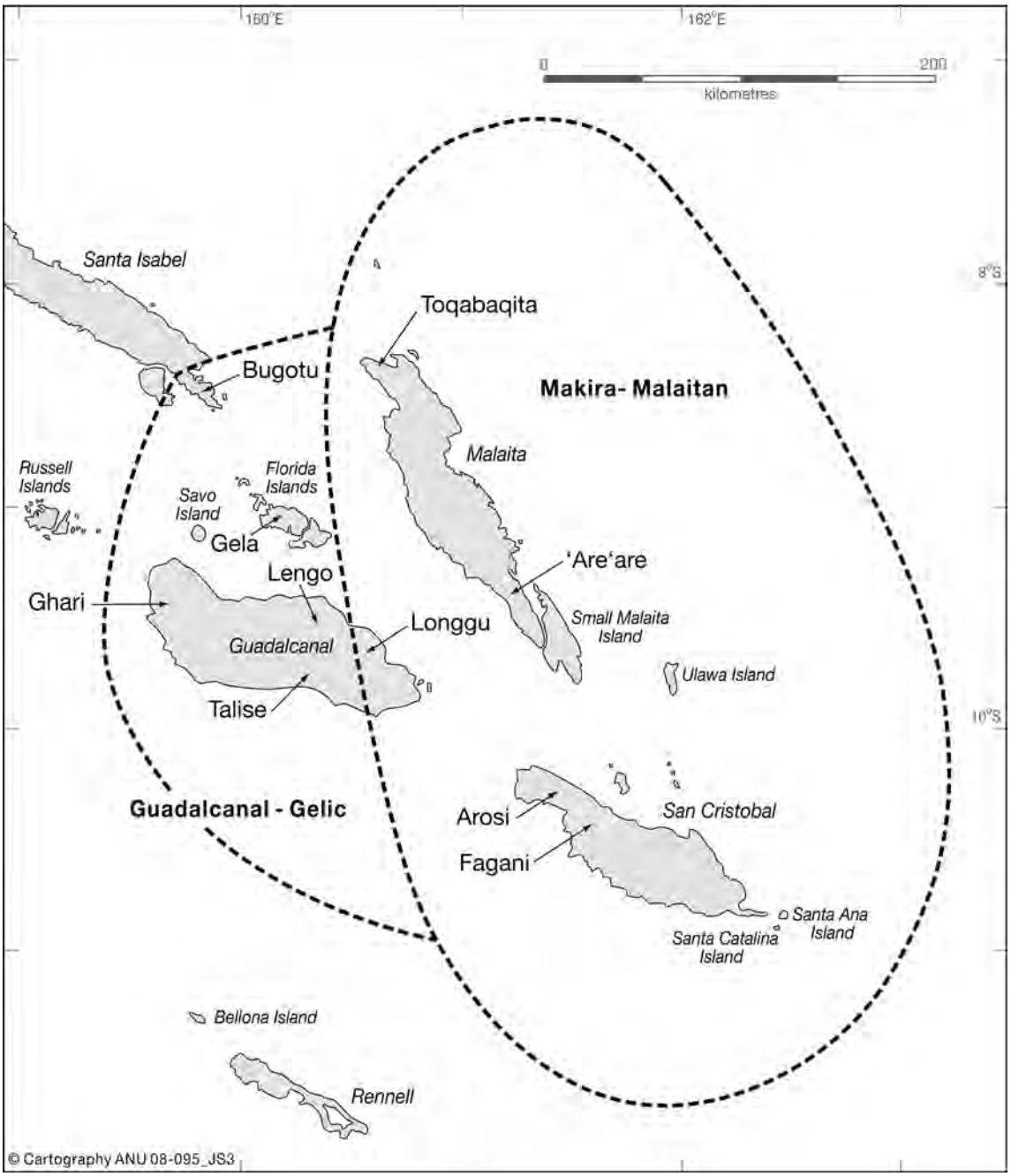


Figure 3: The primary subgroups of SE Solomonic, with languages mentioned in text

5.3 Northwest Solomonic and its subgroups

5.3.1 Northwest Solomonic

The Northwest Solomonic group was not recognised until the early 1980s. Tryon and Hackmann (1983) showed that all the languages from the Shortland Islands to Santa Isabel share a few innovations defining them as a single, though very heterogenous subgroup which they called 'Western Solomons'. Ross (1986, 1988) showed that this group also includes the languages of Bougainville, Buka and Nissan.

Three regular sound changes are attributed to Proto NWS: (i) POc *w is lost in all positions, (ii) an 'echo' vowel is added after word-final consonants, e.g. *onom 'six' > PNWS *onomo, (iii) POc word-final *q becomes PNWS *k, whereas initial and medial *q was either lost or merged with *γ. The POc 1st person singular independent pronoun *au was replaced in PNWS by *(a)rau. The relatively small number of innovations defining NWS indicates that the period of unified development was quite short.

5.3.2 Subgroups of Northwest Solomonic

Ross distinguished five primary branches of NWS: (1) Nissan-Buka-North Bougainville (10 languages), (2) Piva-Banoni (W. Bougainville) (two languages), (3) S. Bougainville-Shortlands (three languages), (5) Choiseul (four languages), and, more tentatively, (5) New Georgia-Santa Isabel (16 languages). Although New Georgia and Santa Isabel are each well-defined groups the evidence for uniting them is slender and any period of common development must have been very brief. For our purposes it is more useful to treat New Georgia (nine languages) and Santa Isabel (seven) as separate primary groups.

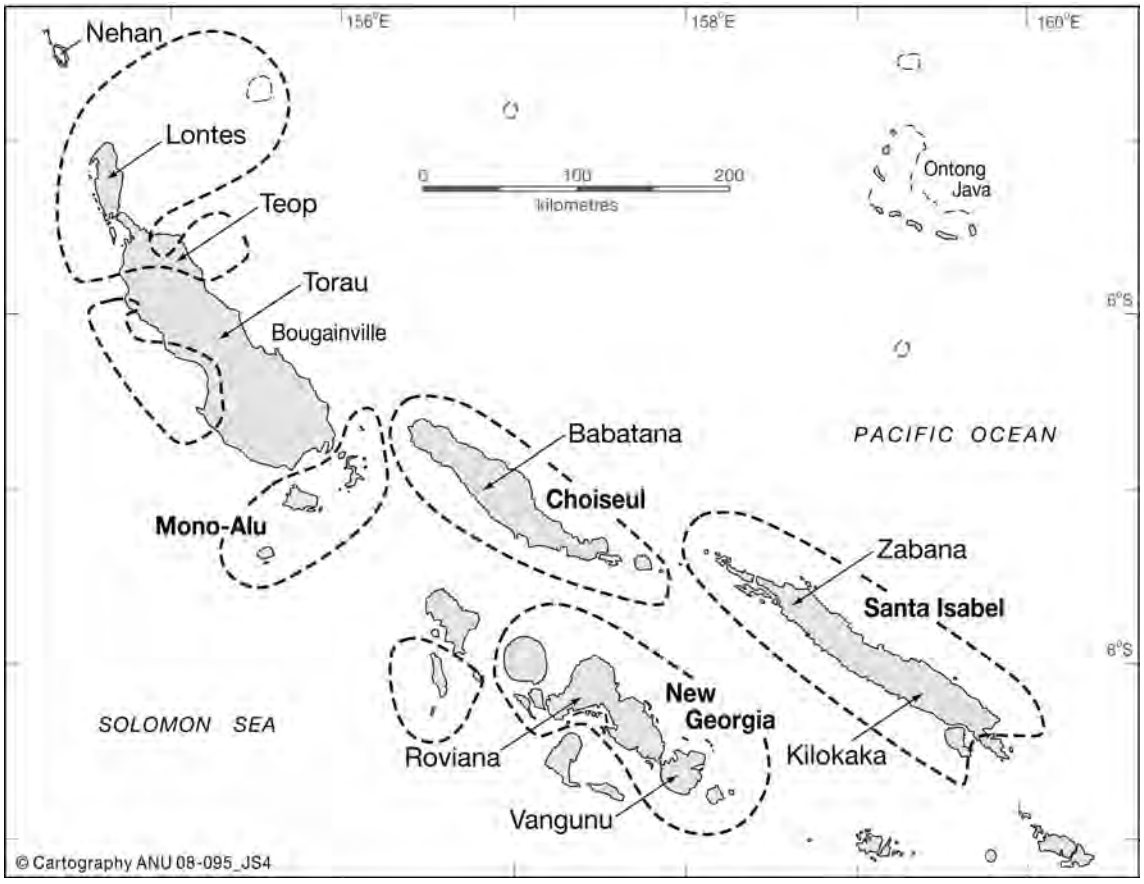


Figure 4: The primary subgroups of NW Solomonic, with languages mentioned in text.

6 Why is there a deep boundary between NW Solomonic and SE Solomonic?

Let us return now to the question of why there is a major subgroup boundary between Northwest Solomonic and Southeast Solomonic. There is no major geographic barrier that would account for this boundary. Ocean gaps between Santa Isabel and Malaita, and between Santa Isabel and the Florida group are on the order of 50 km—i.e. no greater than some of the distances separating islands within the NW Solomonic or the SE Solomonic regions. Oceanic speakers who settled the Solomons certainly had the sailing capacity to maintain regular communication across such ocean gaps. Nor are there other obvious environmental factors, such as periods of explosive vulcanism or the absence of key natural resources, which might account for the boundary

It seems, then, that we must look for an explanation of this boundary in terms of historical and social factors. An obvious question is: do NWS and SES belong to different branches of Oceanic, each with members elsewhere?

Our understanding of the high-order subgrouping of the Oceanic languages of western Melanesia rests largely on two important studies. Blust (1978b) showed that the 20 or so languages of the Admiralty and Western Is. form a closed subgroup. He also pointed to a single phonological change undergone by all other Oceanic languages except the Admiralities, namely the merger of Proto Austronesian *j and *s, and on this basis assigns all non-Admiralties languages to a single subgroup of Oceanic (Blust 1978b, 1998), which I will refer to here as 'Nuclear Oceanic'. Ross's (1988) monumental study encompassed all the Oceanic languages in 'western Melanesia' (defined as extending as far east as the boundary between NW Solomonic and SE Solomonic). He found evidence indicating that, within the Bismarck archipelago, there was an early two-way split between two primary branches of Oceanic: (i) an Admiralties subgroup, well defined by shared innovations, and (ii) a Western Oceanic (WOc) 'linkage', which includes all or almost other Oceanic languages of the Bismarcks and those of Papua New Guinea. A linkage is an imperfect subgroup, defined by innovations that link different sections of the chain, rather than by innovations shared by all members. A linkage derives from a well-differentiated dialect chain rather than a relatively homogeneous ancestor. Ross (1988) also noted the possibility that there was a third primary branch of Oceanic in western Melanesia, consisting of the small Mussau subgroup. He said little about Oceanic languages of the SE Solomons and Remote Oceania. However, he inclined to the view that these languages separated very early from Oceanic languages spoken in the Bismarcks, as the result of a single eastward movement from the Bismarcks through the Solomons and beyond into Remote Oceania.

Ross concluded that the Western Oceanic languages remained confined to the Bismarcks for some time, initially as a complex of dialects represented in parts of coastal north New Britain east of the Willaumez Peninsula, in Bali-Vitu (the French Is.) off the coast of New Britain, and in New Ireland and its offshore islands. At some point Western Oceanic dialects spread beyond this region in two directions: to the New Guinea mainland and to the NW Solomons. He found that the NW Solomonic languages share some innovations with Western Oceanic languages found in the Bismarcks that are not present in the Oceanic languages of the New Guinea mainland. These innovations define an imperfect subgroup that he called the Meso-Melanesian (MM) linkage. The diagnostic innovations are (i) merger of POc *r and *R as *r, (ii) merger of *d and dr as *d, (iv) merger of POc *c and *s as *s, (iv) the split of *k into *k and *γ, and (vi) the split of *p into *p and *v.

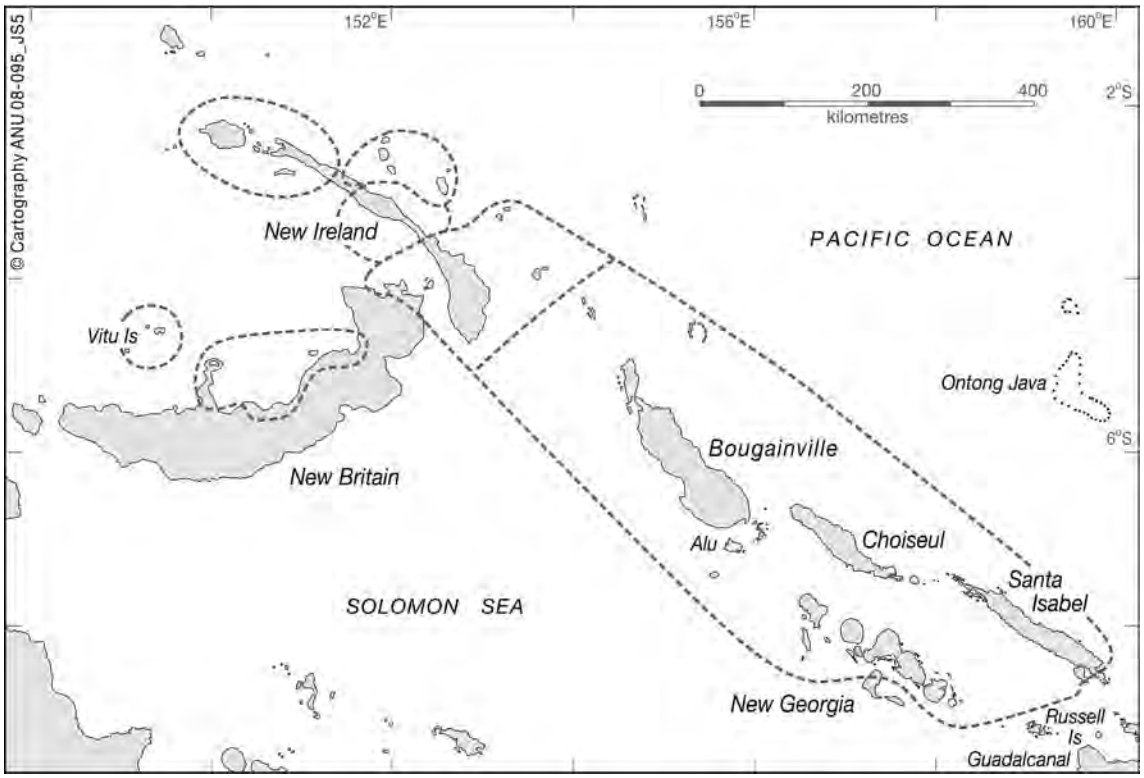


Figure 5: The Meso-Melanesian linkage and its subgroups (after Ross 1988)

There are fragments of evidence indicating that NW Solomonic stemmed from a particular area in the Meso-Melanesian linkage, namely a dialect network centred in southern New Ireland and perhaps extending to the nearby Tanga and Feni groups and to Nissan (Ross 1982, 1988). The evidence consists of a few innovations common to languages of that region and to the North Bougainville members of NW Solomonic. NW Solomonic then developed separately from the S. New Ireland/Tanga/Feni languages. The likely dispersal centre of NW Solomonic is the area consisting of Buka, N. Bougainville and Nissan.

Although Ross' work indicates that the ancestral NW Solomonic language arrived in the western Solomons some centuries after the breakup of POc, it does not explain why the expansion of NW Solomonic stopped at New Georgia and Santa Isabel. As part of the groundwork for tackling this question, I turn now to another vexing question: How intensive were interactions between incoming speakers of Oceanic languages and autochthonous speakers of non-Austronesian languages in different parts of the Solomons archipelago? Some evidence bearing in this question can be found in patterns of lexical replacements.

7 Evidence that NW Solomonic languages have replaced basic lexicon faster than SE Solomonic languages

7.1 Identifying the most stable 60 POc words

It has long been the impression of Oceanicists that SE Solomonic languages are among the most conservative members of the Oceanic group in respect of lexicon and that their sister languages in the NW Solomons have been more innovative. The usual explanation

for this difference is that the NW Solomonic languages have been strongly influenced by contact with non-Austronesian languages whereas SE Solomonic languages have not.[4] However, as far as I know no one has tried to measure rates of lexical replacement in the languages in question, or to pinpoint the periods when particular lexical changes took place. In order to achieve these two objectives, the rates of replacement in 60 highly stable words were investigated for a sample of SES, NWS and other Oceanic languages.

The following procedure was used to identify the 60 most stable POc words, i.e. the words with the highest retention rates in the daughter languages. (i) A first approximation was made by examining a table in Dyen et al. (1967) that ranks *word meanings* (not forms) on the Swadesh list of 200 basic lexical concepts according to how often pairs of languages had cognate forms for these meanings, using a sample of some 200 Austronesian languages. (ii) The 65 meanings yielding the highest percentages of cognate pairs were then extracted and the POc lexical form(s) reconstructable for each of these meanings were listed. (In five cases it was necessary to reconstruct pairs of synonymous forms and to count a retention of either etymon as a plus). (iii) A few problematic meanings were eliminated from the list, reducing it to 60. (iv) Retentions and losses for these etyma were recorded in 40 contemporary Oceanic languages drawn from various major subgroups.[5] (v) From these comparisons an average retention rate for each POc etymon was computed.

This procedure proved to have some flaws. It turned out that at least two of the lexical items that are among the most 20 stable items in Oceanic languages were missing from the variant of the Swadesh 200 word list used by Dyen et al. (1967), namely '(woman's) breast' and 'excrement'. In addition, several other etyma that are among the 60 most highly stable items in our Oceanic comparisons have meanings that do not appear in the top-ranked 65 items in Dyen et al.'s list. These included 'cry', 'night', 'tail', 'moon', 'star'

---

[4] I have found no works specifically addressing the differences between NW Solomonic and SE Solomonic but there is a large literature on the effects of contact between Austronesian and Papuan languages in various parts of Melanesia. See Dutton and Tryon (1994), Pawley (2006), Blust (2005, 2008), Donohue and Denham (2008) for recent discussions.

[5] The 40 languages in the sample used to calculate retention rates were:

SE Solomonic:	Guadalcanal-Gelic:	Bugotu, Gela, Talise
	Makira-Malaitan:	Arosi, Toqabaqita (To'aba'ita)
NW Solomonic:	Bougainville-Mono:	Mono, Teop, Torau, Lontes (Halia)
	Nehan:	Nehan
	Choiseul:	Babatana
	New Georgia:	Roviana, Vangunu
	Santa Isabel:	Kilokaka
Polynesian:	Samoan, Niuean	
Fijian:	Bauan (E. Fijian)	
Micronesian:	Marshallese, Woleai	
S. Vanuatu:	Erromangan (= Sye)	
N. Vanuatu:	Mota, Nguna, Raga	
Eastern Outer Is.:	Malo, Vano, Asumboa	
N. New Ireland:	Lihir, Tigak	
S. New Ireland:	Sursurunga, Kuanua	
New Ireland islands:	Anir	
W. New Britain:	Bali, Nakanai	
Manus:	Kele, Titan	
N. New Guinea:	Manam, Takia, Lote (= Pomio)	
Papuan Tip:	Motu, Galea	

and 'where?'[6] No doubt these discrepancies arise in part from the different language samples used in the two studies but they are likely to be due mainly to the fact that Dyen et al. dealt with cognate percentages for meanings whereas my study deals with the retention rate of individual word forms. The discrepancies were not noticed until the analysis was well advanced and time constraints have prevented me from redoing the calculations. However, the fact that a few highly stable words were omitted from the list of 60 used in this study does not matter—given a list of highly stable items the important thing is how different languages behave with regard to these.

Average retention rates for the 60 items in a sample of 40 Oceanic languages are shown in Table 1.

Table 1: Retention rates for POc reconstructions for 60 highly stable items on the basic vocabulary list, based on 40 languages

		POc	% retained			POc	% retained
1	eye	mata	97	31	fruit	puaq	60
2	we excl	kami	97	32	new	paqoRu	57
3	we incl.	kita	95	33	dig	kali-, keli-	56
4	two	rua	92	34	bird	manuk	56
5	father	tama-	90	35	inside	lalom	56
6	you pl.	kam(i)u	90	36	path	jalan	53
7	they	ira	90	37	name	[ŋ,q]ajan	52
8	mother	tina-	82	38	head	qulu, p^{w}atu	50
9	louse	kutu	82	39	tooth	nipon	50
10	die	mate	82	40	woman	papine	50
11	five	lima	82	41	to fear	matakut	50
12	thou	iko, koe	80	42	root	wakaR, lamut	50
13	three	tolu	80	43	one	tasa, sakai	50
14	hear	roŋoR	77	44	liver	qate	50
15	four	pati	75	45	blood	draRaq	47
16	tongue	maya	73	46	water	waiR	46
17	I	[i]au	73	47	far	sauq	46
18	come	(lako) mai	73	48	skin	kulit	43
19	ear	taliŋa	72	49	feather	pulu	43
20	nose	ŋicu	70	50	rain	qusan	42
21	eat	kani	70	51	fire	api	42
22	drink	inum	70	52	leaf	draun	40
23	vomit	luaq, mumutaq	70	53	sky	laŋit	40
24	tree	kayu	70	54	thin	manipis	40
25	he/she	ia	67	55	ashes	drapu(R)	37
26	stone	patu	67	56	egg	katoluR	36
27	hand	lima	66	57	day	qaco, raqani	36
28	fish	ikan	66	58	right(hd)	mataqu	36
29	what	sapa	66	59	bone	suRi	23
30	who	sai	60	60	heavy	(ma)mapat	23

[6] Retention rates (in percentages) for some additional stable POc etyma in the 40 language sample: *susu 'breast' 85, *taŋis 'cry' 80, *taqe 'excrement' 75, *pai, *pea 'where?' 57, *pituqun 'star' 55, *boŋi 'night '52'. Percentages for 'breast' and 'excrement', are based on samples of 34 and 22 languages, respectively, as some wordlists do not include these items.

7.2 Results

Retention rates for the 60 POc etyma were then calculated for each of the 40 languages in the sample plus a further dozen or so languages.[7] Table 2 shows retention rates for the NW and SE Solomonic languages in the sample.

Table 2: Retention rates for 60 highly stable words in some SE Solomonic and NW Solomonic languages

		Items retained	Percentage retained
SE Solomonic	**Guadalcanal-Gelic**		
	Gela	52	86
	Lengo	48	80
	Ghari	47	78
	Talise	45	75
	Bugotu	41	68
	Makira-Malaitan		
	Fagani	48	80
	Longgu	46	77
	Arosi	43	71
	‘Are‘are	41	68
	Toqabaqita	39	65
NW Solomonic	**Nehan-Buka-N. Bougainville**		
	Nehan	31	52
	Teop	24	40
	Lontes	32	53
	S. Bougainville		
	Mono-Alu	27	45
	Torau	34	57
	Choiseul		
	Babatana	25	42
	New Georgia		
	Roviana	35	59
	Vangunu	33	55
	Santa Isabel		
	Kilokaka	27	45
	Zabana (Kia)	26	44

[7] It is noteworthy that Proto Central Pacific (PCP) retained all 60 of the POc items we are concerned with here. Put another way, the forms for meanings 1-60 reconstructed by comparing just Fijian, Rotuman and Polynesian are the same as those reconstructed by comparing the full range of Oceanic groups. For two POc etyma, *katoluR ‘egg’ and *sauq ‘far’, there are reflexes only in Rotuman, not in Polynesian or Fijian. I do not suggest that such a high level of retention would hold for the PCP lexicon as a whole, but this is evidence that the early Oceanic language(s) that reached the Central Pacific region had changed rather little from POc itself. It indicates that the interval between the breakup of POc and the breakup of PCP was at most a few centuries. In PPn six items have been replaced: *draRaq ‘blood > *toto, *qajan ‘name’ > *hiŋoa, *maya ‘tongue’ > *qalelo, *katoluR ‘egg’ > *fua, *lalom ‘inside’ > *loto.

Among SES Solomonic languages the average percentage of retentions is 73, the highest being 86 (Gela) and the lowest 65 (Toqabaqita). Bugotu scores much lower (68) than other GG languages. This is mainly because it has borrowed some basic lexical items from Santa Isabel neighbours which show high replacement rates. It is also noteworthy that the languages of Makira (represented here by Arosi and Fagani) are in general somewhat more conservative than the Malaitan languages (represented by 'Are'are and Toqabaqita).

Among NW Solomonic languages the average percentage of retentions is 49, the highest being 59 (Roviana) and the lowest 40 (Teop).

It can be seen that all the NW Solomonic languages have replaced more of the POc basic lexicon than any of the SE Solomonic languages. However, there is considerable variation within each group and the most conservative NW Solomonic language, Roviana (59 per cent) scores only a few per cent less than the most innovative SE Solomonic language, Toqabaqita (65).

7.3 Determining when lexical replacements occurred in SE Solomonic and NW Solomonic

It is clear that NW Solomonic languages have replaced much more basic vocabulary than SE Solomonic. But can we determine when the changes occurred? To answer this question it is necessary to reconstruct particular interstages (intermediate protolanguages) in order to see which items were replaced between earlier and later stages. This has been done for some interstages.

7.3.1 Lexical changes in Proto SE Solomonic and Proto NW Solomonic

The proto-languages of the SE Solomonic and NW Solomonic groups were both lexically quite conservative. Proto SE Solomonic (PSES) replaced just three of POc items 1-60: *draRaq 'blood' > *kabu; *matakut 'be afraid' > *matolo; *laŋit 'sky' was replaced, probably by *masawa(ŋ).[8] (In POc the primary sense of *masawa(ŋ) was apparently 'the open sea, far from land', with a secondary sense 'vast open space(s)'.) Proto NW Solomonic (PNWS) replaced just four of items 1–60: *draun 'leaf', *api 'fire', *papine 'woman' (retained only in the sense of 'man's sister') and *waiR 'water'.[9]

It is noteworthy that no replacements of POc reconstructions for items 1–60 are shared by PSES and PNWS. This is strong evidence that the two protolanguages had independent histories after they diverged at the level of Proto Nuclear Oceanic. However, in later times some borrowing occurred between certain neighbouring languages across the NWS/SES boundary, and this occurred even in a few items of basic vocabulary.[10]

---

8 The same replacement, *laŋit > masawa(ŋ), is found also in some Vanuatu languages.

9 For each of the four items replaced it is hard to reconstruct a Proto NW Solomonic etymon because the replacements differ across subgroups.

10 Evidence for borrowing between Guadalcan-Gelic languages and the nearer NW Solomonic languages is suggested by the following comparisons, among others. In most S. Isabel/Guadalcanal-Gelic languages, POc *talinŋa 'ear' is replaced by the type of Gela *kuli* and *pisiko 'flesh' by the type of Gela *vinasi*, Poc *tam^{w}ata and *tau 'person' are replaced by the type of Gela *tinoni* in most New Georgia/SE Solomonic languages. See Blust (2007:411–412) for a fuller list.

7.3.2 Lexical change in subgroups of SE Solomonic

The proto-languages of the major subgroups of SE Solomonic remained lexically conservative. In addition to the three replaced in Proto SE Solomonic, Proto Makira-Malaitan replaced four to five items: *draun ‘leaf’ > *ʔa[f,p]a, ʔapa [Toqabaqita has *rau* ‘leaf, leaflet’]; *api ‘fire’ > *kiu or *[d,t]una; *mataqu ‘right hand’ > *matolo or *katolo; *pulu ‘feather’ > *(wara)ifu. POc *mapat ‘heavy is lost but a reflex of POc *(b,p)ita ‘heavy’ is retained in a few languages.

Besides the three items replaced in Proto SE Solomonic, Proto Guadalcanal-Gelic replaced four items: *taliŋa ‘ear’ > *kuli; *maya ‘tongue’ > *lapi, *api ‘fire’ > *lake, *waiR ‘water’ > *kolo. In addition, Proto GG lost *wakaR, the most general term for ‘root’ but retained POc *lamut ‘root’, a term that probably referred specifically to fibrous roots and root hairs.

The contemporary languages in both SE Solomonic groups, as we saw in §7.2, also remain lexically more conservative than NW Solomonic languages. This relatively small number of lexical replacements strongly suggests that neither Proto GG nor Proto MkMl nor their descendants were much influenced by contact with non-Austronesian languages. Evidently at the time Oceanic speakers arrived in the Southeast Solomons non-Austronesian speaking populations in this region were small and were easily absorbed or displaced.

7.3.3 Lexical change in subgroups of NW Solomonic

Once speakers of early varieties of NW Solomonic dispersed across the NWS region each local variety underwent rapid lexical change. Thus, of the 56 POc items retained by Proto NW Solomonic in the 60 item list, Proto Choiseul, as we reconstruct it, retains only 30. That is to say, in the period between Proto NW Solomonic and Proto Choiseul almost half of the highly stable lexicon was replaced, Proto S. Isabel retains 36/56, having replaced more than a third. Proto New Georgia retains 47/56 but this is still a loss of almost 20 percent. I have not calculated percentages for the other NWS subgroups.

This very high rate of replacement in the most stable part of the lexicon indicates extensive borrowing from non-Austronesian sources. A reasonable inference is that in each of these regions the speakers of incoming NW Solomonic languages encountered substantial populations of non-Austronesian languages and that sustained bilingualism, especially in Choiseul and Santa Isabel but also in the New Georgia group, led to many non-Austronesian loanwords entering the basic vocabulary of the NW Solomonic languages. It remains to be seen to what extent putative borrowings from non-Austronesian sources can be associated with particular surviving non-Austronesian languages of the Solomons group.

7.4 Comparison with other Oceanic languages

Comparison of replacement rates in Oceanic languages spoken outside of the Solomons Archipelago reveals a pattern consistent with the hypothesis that higher rates correlate with more intensive contact between Oceanic and non-Oceanic languages. Table 3 gives retention rates for the 60 most stable items in a sample of languages from Polynesia, Fiji, Micronesia and Vanuatu. All are spoken on islands in Remote Oceania and probably had no direct contact with non-Austronesian languages after these islands were settled.

Table 3: Retention rates for the 60 most persistent words items in some languages of Remote Oceania

		Items retained	Percentage retained
Polynesian			
	Tikopia	53	88
	Tongan	51	85
	Samoan	50	83
	Niuean	50	83
	Maori	48	80
Fijian			
	Bauan (E Fijian)	49	81
	Wayan (W Fijian)	49	81
Central and Northern Vanuatu			
	Raga (Pentecost Is.)	47	78
	Nguna (Efate)	45	75
Southern Vanuatu			
	Erromangan	37	61
Nuclear Micronesian			
	Woleai	43	71
	Marshallese	39	65

The range of retention rates in these particular languages is similar to that found in SE Solomonic. All have retained more of the POc basic lexicon than any of the NW Solomonic languages.

Next is a set of languages also spoken in Near Oceania which, at certain periods in their history, are likely to have had sustained contact with non-Austronesian languages. It can be seen that scores for these languages fall within the range of the NW Solomonic languages.

Table 4: Retention rates for some languages of Near Oceania likely to have had fairly high contact with non-Austronesian languages

	Items retained	Percentage retained
North New Guinea subgroup		
Takia	29	48
Sengseng	19	31
Southern New Ireland subgroup		
Kuanua	30	50

8 How long have SE Solomonic languages been in the Solomons archipelago?

Let us now return to the hypothesis (Sheppard and Walter 2006, Spriggs 1997) that when the Oceanic-speaking Lapita people first colonized Remote Oceania just over three millennia ago they bypassed the Solomons Archipelago, and that it was another 800 years or so before speakers of Oceanic languages established permanent settlements in the

Guadalcana-Malita-Makira region. The archaeological evidence bearing on this proposal is equivocal, as was noted in §2.

Historical linguistics could throw light on this matter if a way could be found of dating the nodes on the SE Solomonic branch of the Oceanic family tree. The chief absolute dating method developed in linguistics is the much-maligned '(lexicostatistical) glottochronology', which uses cognate percentages in basic lexicon to date the length of time since particular related languages diverged.[11] In the foundation research on glottochronology the mean replacement rate for items in the 200 list was initially calculated to be about 19.5% per millennium. Rounding this to 20% yields the following predictions for a single language: 80% of the original 200 items will be retained after 1000 years, 64% after 2000, 51% after 3000, 41% after 4000. When estimating separation dates from cognate percentages between contemporary languages, the equations based on 20% replacement per millennium are: 64% cognates = 1000 years separation, 41% = 2000, 28% = 3000, 17% = 4000.

In the case of Austronesian languages, these estimates can be tested against an independent chronology that can be established for particular intermediate proto-languages (the ancestors of particular subgroups) by correlating linguistic and archaeological events. Austronesianists have a valuable external point of reference when estimating the dates at which particular subgroups broke up, namely, several cases where archaeological dates for the settlement of a particular can, with high confidence, be matched with the arrival of a particular language in that region, a language ancestral to a large subgroup. Thus, one can date the breakup of Proto Malayo-Polynesian to about 4000 BP, because the emergence of the Malayo-Polynesian branch of Austronesian can be connected with the movement of people from Taiwan across the Bashi Channel into the Batanes Is. and Luzon at about that time (Bellwood pers. comm.; Bellwood and Dizon 2005; Ross 2005). The breakup of POc can be placed at between 3400 and 3100 BP (see §3). We can be confident that the Central Pacific languages (Fijian, Rotuman and Polynesian) diverged from both the NW Solomonic and SE Solomonic groups no later than about 3000 years ago. This is because the foundation settlement of Fiji and Tonga is rather securely dated to about 3050–2950 years ago. An earliest possible date for the split is that assigned to the breakup of POc itself.

Although it has been shown that Malayo-Polynesian languages vary greatly in their retention rates (Blust 1981, 1999), there is reason to think that the standard glottochronological estimates are about right for lexically conservative Oceanic languages. Assuming that Proto Malayo-Polynesian broke up about 4000 BP, we get results close to the mark for the most conservative Oceanic languages, such as Gela, Samoan and Fijian. Each is known to retain about 40% of the reconstructed Proto Malayo-Polynesian items for 200 item basic lexicon. And although the calculations have not been done for the full range of languages, we can be reasonably sure that quite similar results will be obtained for almost all the SE Solomonic languages, all the Fijian languages and many of the Polynesian languages.

Given this method, it is possible to assign approximate dates to the breakup of Proto SE Solomonic and its daughter subgroups, Guadalcanal–Gelic and Makira-Malaitan. The following account of lexical diversity exhibited by languages in the SE Solomons and neighbouring areas draws on the percentages given in Tryon and Hackman (1983) for the Swadesh 200 item basic lexicon.

---

[11] Russell Gray and his associates have in recent years been developing an alternative dating method (Gray 2005; Gray and Atkinson 2003; Greenhill and Gray this volume).

Let us first consider how SE Solomonic languages score with other Oceanic languages that are known to be fairly conservative and compare these agreements with those between Guadalcanal-Gelic and Makira-Malaitan languages.[12] Recall that the split between SE Solomonic and Polynesian is dated to no later than 3000 years ago. Cognate percentages between SE Solomonic languages and five Polynesian Outlier languages in the Solomons region (Rennellese, Tikopia, Sikaiana, Luangiua and Pileni) fall between 25 and 36, with a median of 29.[13] Percentages between Guadalcanal-Gelic and Makira-Malaitan languages fall between 28 and 43, with a median of 36.

The differences between the SE Solomonic-Polynesian agreements and the agreement between Guadalcanal-Gelic and Makira-Malaitan are thus on the order of 7 percent. This is consistent with about 500 years elapsing between the SE Solomonic-Polynesian split, and the breakup of SE Solomonic into incipient Guadalcanal-Gelic and Makira-Malaitan branches.

Next let us consider agreements within the Makira-Malaitan group. The Makira–Malaitan languages are clearly descended from a dialect chain that extended over most of the Makira-Malaitan region. Today the lexical diversity of languages from opposite ends of this region is almost as great as the divergence between Makira-Malaitan and Guadalcanal-Gelic. The most differentiated Makira-Malaitan languages show percentages in the 34–40% range, e.g. Toqabaqita of N. Malaita has the following percentages with Makira languages: 34 with Santa Ana, 35 with Kahua and Bauro, 40 with Arosi. These are about the same as Toqabaqita shares with Guadalcanal-Gelic (32–36%). All this suggests that the opposite ends of the Proto Makira–Malaitan region began to diverge into dialects soon after Makira–Malaitan split off from Guadalcanal-Gelic but that the divergence proceeded gradually because the central dialects of Makira-Malaitan remained in close contact with the extremes.

Guadalcanal-Gelic is more homogeneous than Makira-Malaitan. Excluding Bugotu, the most differentiated GG languages show cognate percentages in the range 50–55% and some pairs of languages score 60–70%. This strongly suggests that the ancestral GG dialect chain remained fairly cohesive for much longer than Makira-Malaitan, with most dialects remaining mutually intelligible until about 1000 years ago. Table 5 gives approximate divergence dates for pairs of groups based on the median percentage, using the standard glottochronological equations.

---

[12] Excluded from the intra-SE Solomonic comparisons are Marau and Longgu, two MkMl languages spoken on Guadalcanal, whose percentages are inflated by loans from GG neighbours. Also excluded is one GG language, Bugotu, whose percentages with MkMl and with other GG languages are much lower owing to sustained contact with Santa Isabel languages. Bugotu's agreements with MkMl are in the range 26–32%, i.e. almost 10% lower than other GG languages.

[13] For example, the lexically most conservative GG language, Gela, scores 31–36% with Polynesian Outliers. It scores just a bit higher, 34–43%, with MkMl languages. Its sister language Ghari scores 28–32% with Polynesian Outliers, compared to 33–40% with MkMl languages. The most conservative MkMl language, Fagani (of Makira), scores 28–33% with Polynesian Outliers compared to 36–43% with GG. The least conservative MkMl language, Toqabaqita (of Malaita), scores 25–27% with Polynesian Outliers, compared with 32–36% with GG. The most conservative MkMl language, Fagani (of Makira), scores 28–33% with Polynesian Outliers compared to 36–43% with GG.

Table 5: Cognate percentages for inter-group comparisons with approx. divergence dates

	percentages	median	approx. divergence date for median
SES-Polynesian	25–36	29	2900 BP
MkMl-GG	28–43	36	2400 BP
extremes of MkMl	34–40	37	2300 BP

These figures do not, of course, tell us how long the ancestral SE Solomonic language was in the SE Solomons before it diverged into GG and MkMl. However, it is reasonable to assume that the innovations defining SES were accumulated over a few centuries when pre-SES was a single language—no doubt with dialect variants—spoken in a string of mainly coastal and small island settlements in parts of Makira, Malaita and Guadalcanal.[14]

But where was pre SE Solomonic spoken before it was carried to the SE Solomons? Does this group fall into a subgroup with any other branch of Nuclear Oceanic?

From time to time it has been argued that SES falls into an Eastern Oceanic group together with most or all of the Oceanic languages of Remote Oceania, especially those of Vanuatu, New Caledonia and the Loyalties, Fiji, Polynesia and possibly Micronesia. There are a few scraps of evidence supporting such a group but the hypothesis remains highly problematic and this is not the place to review the evidence.[15]

9 Conclusions

We are led to the following conclusions.

1. The sharp boundary between NW and SE Solomonic is not the product of *in situ* divergence. The NW and SE Solomons regions were settled independently by two different populations of Oceanic speakers.

2. The position of the NW Solomonic languages on the Oceanic family tree is consistent with Sheppard and Walter's proposal that that the NW Solomons was bypassed in the initial movement of Lapita people into Remote Oceania. NW Solomonic is a division of the Meso-Melanesian branch of Oceanic. The centre of diversity within Meso-Melanesian, and its original site is clearly in the New Britain-New Ireland area. At some point speakers of a Meso-Melanesian language moved to the Nissan-Buka-N. Bougainville region. There the language developed the few innovations that define the NW Solomonic subgroup. After a short period of unified development Proto NW Solomonic spread to the Shortlands, Choiseul, New Georgia and Santa Isabel. Linguistic methods do not allow us to date precisely the spread of NW Solomonic. However, it is clear, from the archaeological

[14] Recently Lynch et al. (2002:110ff.) have suggested that Proto SE Solomonic was confined to the Bugotu-Gela-North Guadalcanal region and that its descendants later moved from Guadalcanal into Makira and Malaita. However, this scenario rests on a very flimsy argument.

[15] Re the Eastern Oceanic hypothesis, see Grace (1976), Pawley (1972), Lynch and Tryon (1985). The Oceanic languages of the Eastern Outer Islands region are not known to share any innovations with Southeast Solomonic. Although their histories are still poorly understood it seems likely that the better known languages of Utupua and Vanikoro form a first-order subgroup of Nuclear Oceanic, to which Aiwoo of the Reef Is. may also belong (Ross and Næss 2007). In that case, they are likely to be relics of the first Lapita movement into Remote Oceania. All this does not rule out Greater Reefs/Santa Cruz as a source for pre-SE Solomonic. It simply implies that if it was, pre-SE Solomonic speakers left Reefs/Santa Cruz quite soon after Oceanic speakers first arrived there.

evidence, that the breakup of Proto Oceanic must have occurred between about 3350 BP, by which time Lapita settlements had been established in various parts of the Bismarck Archipelago, and 3100 BP, by which time Lapita settlements had been established in Remote Oceania. The innovations marking off Meso-Melanesian from the rest of Oceanic, and those marking off NW Solomonic from the rest of Meso-Meso-Melanesian are relatively few, and in all, probably took no more than three or four centuries to accumulate. This estimate would place the breakup of NW Solomonic as occurring between about 3000 and 2700 BP.

3. Subsequently, in the course of dispersing across the NW Solomons, the ancestral NW Solomonic language developed regional variants that underwent very rapid lexical change. Many words not known to have Austronesian antecedents entered their core lexicons. A reasonable explanation is that in each locality small populations of immigrant Oceanic speakers came into contact with established populations of non-Austronesian speakers, leading to extensive intermarriage, bilingualism and lexical borrowing from non-Austronesian languages.

4. Over the next couple of millennia Austronesian languages replaced non-Austronesian languages over most of the NW Solomons. An exception is Bougainville, where non-Austronesian languages remain dominant over most of the island.

5. The scenario sketched in 2–4 above does not preclude the possibility that speakers of NW Solomonic were not the first speakers of an Oceanic language to settle in the NW Solomons. However, if there were earlier Oceanic-speaking colonists, they left no surviving daughter languages in the region. This fact suggests that, at best, any earlier Oceanic-speaking populations must have been small.

6. The SE Solomonic languages show few signs of influence from non-Austronesian languages, an indication that the pre-Austronesian populations were sparse in the SE Solomons. However, non-Austronesian languages survive on two small islands near Guadalacanal: Savosavo and Russell.

7. The linguistic evidence weighs strongly against Sheppard and Walter's suggestion that the islands from Guadalcanal to Makira were not settled until around 2300–2200 BP, around the time when the making of ceramics had ceased in the Reefs/Santa Cruz area. Southeast Solomonic is a fairly well defined subgroup of Oceanic, without obvious close relatives elsewhere and it must have separated from the language ancestral to the Fijian and Polynesian groups no later than 3000 BP. The set of phonological, morphological and lexical innovations that define Southeast Solomonic indicate several centuries of unified development in the Southeast Solomons region. The internal diversity of Southeast Solomonic is also considerable. In comparisons of a 200 item basic lexicon the two primary subgroups of SE Solomonic (Guadalcanal-Gelic and Makira-Malaitan) diverge from each other almost as sharply as they diverge from Fijian and Polynesian. This degree of difference points to the two subgroups as having followed separate paths since about the middle of the first millennium BC. Furthermore, the languages at opposite ends of the Makira-Malaitan subgroup differ from each other, lexically, almost as sharply as they do from Guadalcanal-Gelic languages, indicating that internal differentiation within Makira-Malaitan began around the same time (although the rate was slowed by the persistence of a dialect chain). I conclude that the SE Solomonic languages have been present in Makira, Malaita and Guadalcanal for well over 2500 years and probably for around 3000 years.

8. It is uncertain where the immediate ancestor of SE Solomonic came from. There is no decisive evidence to subgroup SE Solomonic with any other branch of Nuclear Oceanic. On archaeological grounds an immediate origin from the east, from the Eastern Outer Islands of the Solomons, or from Vanuatu, is perhaps more likely than direct settlement directly from the Bismarcks. Over the years a number of linguists have pointed to scraps of evidence suggesting a brief shared history with certain other languages of Remote Oceania, especially those of Vanuatu, Fiji, Polynesia and Micronesia but the evidence is far from decisive.

9. If SE Solomonic speakers dispersed over the coasts and offshore islands of Makira, Malaita and Guadalcanal in the first half of the 1st millennium BC one may ask why did they not also settle the nearest parts of the Western and Central Solomons, such as Santa Isabel and New Georgia. I think a good part of the answer is that at that time the latter islands were populated exclusively, or almost exclusively by non-Austronesian speakers and that they remained largely non-Austronesian speaking for many centuries after that. In Santa Isabel and New Georgia, as well as on the small islands of Russell and Savo, non-Austronesian speaking areas for a time formed a buffer between NW Solomonic and SES Solomonic languages. However, once the two subgroups came into contact there was a good deal of borrowing between the languages closest to each other.

References

Bedford, Stuart. 2003. The timing and nature of Lapita colonization in Vanuatu: the haze begins to clear. In Sand, ed. 147–158.

Bedford, Stuart, Matthew Spriggs and Ralph Reggenvanu. 2006. The Teouma site and the early human settlement of the Pacific Islands. *Antiquity* 80:812–828.

Bedford, Stuart, Shaun Connaughton and G. Clark, eds. 2007. Oceanic explorations: Lapita and western Pacific settlement. Terra Australis 26. Canberra: ANU E–Press.

Bellwood, Peter. 1997. *Prehistory of the Indo–Malaysian Archipelago.* (Revised edition.) Honolulu: University of Hawaii Press.

Bellwood, Peter and Eusebio Dixon. 2005. The Batanes archaeological project and the 'out of Taiwan' hypothesis for Austronesian dispersal. *J. of Austronesian Studies* 1(1):1–32.

Blust, Robert A. 1978a. Eastern Malayo-Polynesian: a subgrouping argument. In S.A. Wurm and L. Carrington, eds *Second international conference on Austronesian linguistics: Proceedings,* 181–234. Canberra: Pacific Linguistics.

—— 1978b. *The Proto-Oceanic palatals.* Polynesian Society monograph No.43. Auckland: Polynesian Society.

—— 1981. Variation in retention rate among Austronesian languages. Paper presented at 3rd International Conference on Austronesian Linguistics, Bali, January 1981.

—— 1995. Austronesian comparative dictionary. Computer file. Department of Linguistics, University of Hawaii.

—— 1998. A note on higher-order subgroups in Oceanic. *Oceanic Linguistics* 37:182–188.

—— 1999. Why lexicostatistics doesn't work: the 'universal constant' hypothesis and the Austronesian languages. In Colin Renfrew, A. McMahon and L. Trask, eds *Time depth in historical linguistics*, 311–332. Cambridge: MacDonald Institute for Archaeological Research.

—— 2005. Review of Lynch, Ross and Crowley 2002. *Oceanic Linguistics* 34(1):191–202.

—— 2007. Proto-Oceanic *mana revisited. *Oceanic Linguistics* 46:404–423.

—— 2008. Remote Oceanic: one history to two? An addendum to Donohue and Denham. *Oceanic Linguistics* 47:433–444.

Burley, David and S.P. Connaughton. 2007. First Lapita settlement and its chronology in Vava'u, Kingdom of Tonga: New data on old problems in the kingdom of Tonga. *Radiocarbon* 49(1):131–137.

Donohue, Mark and tim Denham. 2008. The language of Lapita: Vanuatu and an early Papuan presence in the Pacific. *Oceanic Linguistics* 47:433–444.

Dunn, Michael, Ger Reesink and Angela Terrill. 2002. The East Papuan languages: a preliminary typological appraisal. *Oceanic Linguistics* 41:28–62.

Dunn, Michael, A. Terrill, G. Reesink, R. Foley, S. Levinson. 2005. Structural phylogenetics and the reconstruction of ancient language history. *Science* 22 September 2005, vol. 309:2072–2075.

Dutton, T.E. and D.T. Tryon, eds. 1994. *Language change and contact in the Austronesian world.* Berlin: Mouton de Gruyter.

Dyen I., A.T. James and J.W.L. Cole. 1967. Language divergence and estimated retention rate. *Language* 43(1):150–171.

Felgate, Mathew. 2001. A Roviana ceramic sequence and the prehistory of Near Oceania: work in progress. In G.R. Clark, A.J. Anderson and T. Vunidoilo, eds *The archaeology of Lapita dispersal in Oceania. Papers from the 4th Lapita conference, June 2000, Canberra, Australi*a, 39–60. Terra Australis 17. Canberra: Pandanus Books.

___ 2003. Reading Lapita in Near–Oceania: intertidal and shallow–water pottery scatters, Roviana Lagoon, New Georgia, Solomon Islands. PhD thesis, University of Auckland.

—— 2007. Leap–frogging or limping? Recent evidence from the Lapita littoral fringe, New Georgia, Solomon Islands. In S. Bedford, S. Connaughton and G. Clark, eds, 121–140.

Grace, George W. 1959. *The position of the Polynesian languages within the Austronesian (Malayo–Polynesian) language family.* Bloomington: International Journal of American Linguistics, Memoir 16.

—— 1966. Austronesian lexicostatistical classification: a review article. *Oceanic Linguistics*, 5(1):13–31.

—— 1976. Review of R.C. Green and Marion Kelly, eds, *Studies in Oceanic culture history*, vol. 3. *J. Polynesian Society* 85(1):103–112.

Green, Roger C. 1991. A reappraisal of the dating of some Lapita sites in the Reef/Santa Cruz Group of the Southeast Solomon Islands. *J. Polynesian Society* 100:197–207.

—— 2003. The Lapita horizon and traditions — signature for one set of Oceanic migrations. In Sand, ed. 95–120.

Haberle, Simon. 1996. Explanations for palaeoecological change on the northern plains of Guadalcanal, Solomon Islands: the last 3200 years. *The Holocene* 6:333–338.

Kirch, Patrick V. 2000. *On the road of the winds: an archaeological history of the Pacific Islands before European contact.* Berkeley: University of California Press.

Lichtenberk, Frantisek. 1988. The Cristobal–Malaitan subgroup of Southeast Solomonic. *Oceanic Linguistics* 27:24–62.

—— 1994. Reconstructing heterogeneity. *Oceanic Linguistics* 33:1–36.

Lincoln, Peter. 1978. Reef-Santa Cruz as Austronesian. In Stephen Wurm and Lois Carrington, eds *Second international conference on Austronesian linguistics*: *Proceedings*, 929–967. Canberra: Pacific Linguistics.

Lynch, John, Malcolm Ross and Terry Crowley. 2002. *The Oceanic languages*. Richmond, Surrey: Curzon.

Milke, Wilhelm. 1958. Zur inneren Gliederung und geschichtlichen Stellung der ozeanisch-austronesischen Sprachen. *Zeitschrift für Ethnologie* 83:58–62.

Nunn, Patrick D., Roselyn Kumar, Sepeti Matararaba and 10 others. 2004. Early Lapita settlement at Bourewa, southwest Viti Levu. *Archaeology in Oceania* 39:139–143.

Pawley, Andrew. 1972. On the internal relationships of Eastern Oceanic languages. In Roger C. Green and M. Kelly, eds *Studies in Oceanic culture history*, vol. 3. *Pacific Anthropological Records* No.13, 1–142. Honolulu: Bishop Museum.

—— 2003a. Locating Proto Oceanic. In Malcolm Ross, Andrew Pawley and Meredith Osmond, eds *The lexicon of Proto Oceanic. The culture and environment of Ancestral Oceanic Society*: vol. 2. *The physical environment*, 17–34. Canberra: Pacific Linguistics.

—— 2003b. The Austronesian dispersal: languages, technologies, peoples. In P. Bellwood and C. Renfrew, eds *Examining the farming/language dispersals hypothesis*, 251–273. Cambridge: McDonald Institute of Archaeological Research, Cambridge University.

—— 2006. Explaining the aberrant Austronesian languages of Southeast Melanesia: 150 years of debate. *Journal of the Polynesian Society* 116(3):213–256.

—— 2007.The origins of early Lapita culture: the testimony of historical linguistics. In Bedford, Sand and Connaughton, eds, 17–49.

—— 2008. Where and when was proto Oceanic spoken? Linguistic and archaeological evidence. In Yury Lander and Alexander Ogloblin, eds *Language and text in the Austronesian world. Studies in honour of Ulo Sirk*, 47–71. Münich: Lingcom Europa.

Roe, David. 1993. Prehistory without pots: Prehistoric settlement and economy in northwest Guadalcanal, Solomon Islands. PhD thesis, The Australian National University, Canberra.

Ross, Malcolm. 1986. Towards a classification of the Oceanic languages of Bougainville and the western Solomons. In P. Geraghty, L. Carrington and S.A. Wurm, eds *FOCAL II: papers from the Fourth International Conference on Austronesian Linguistics*, 175–200. Canberra: Pacific Linguistics.

—— 1988. *Proto Oceanic and the Oceanic languages of western Melanesia.* Canberra: Pacific Linguistics.

—— 2001. Is there an East Papuan phylum? Evidence from pronouns. In A. Pawley, M. Ross and D. Tryon, eds *The boy from Bundaberg: essays in Melanesian linguistics in honour of Tom Dutton*, 301–321. Canberra: Pacific Linguistics.

—— 2005. The Batanic languages in relation to the early history of the Malayo-Polynesian subgroup of Austronesian. *J. Austronesian Studies* 1(2):1–23.

Ross, Malcolm and Ashild Næss. 2007. An Oceanic origin for Äiwoo, the language of the Reef Islands? *Oceanic Linguistics* 46.

Ross, Malcolm, Andrew Pawley and Meredith Osmond, eds. 1998–2008. *The lexicon of Proto Oceanic. The culture and environment of ancestral Oceanic society*: vol. 1 *Material culture.* vol. 2, *The physical environment.* vol. 3. *Plants.* Canberra: Pacific Linguistics.

Ross, Malcolm, Andrew Pawley and Meredith Osmond, eds. In prep. *The lexicon of Proto Oceanic. The culture and environment of ancestral Oceanic society*: vol. 4. *Animals.* vol. 5. *People and society.* Canberra: Pacific Linguistics.

Sand, Christophe. 2001. Evolutions in the Lapita Cultural Complex: a view from the Southern Lapita Province. *Archaeology in Oceania* 36:65–76.

—— ed. 2003. *Pacific archaeology: Assessments and prospects. Proceedings of the international conference for the 50th anniversary of the first Lapita conference (July 1952).* Les cahiers de l'archéologie en Nouvelle-Caledonie, vol. 15. Noumea: New Caledonia Museum.

Sheppard, Peter and Richard Walter. 2006. A revised model of Solomon Islands culture history. *J. Polynesian Society* 115(1):47–76.

Specht, Jim. 2005. Revisiting the Bismarcks: some alternative views. In A. Pawley, R. Attenborough, J. Golson and R. Hide, eds *Papuan pasts: cultural, linguistic and biological histories of Papuan–speaking peoples*, 255–288. Canberra: Pacific Linguistics.

Spriggs, Matthew. 1997. *The Island Melanesians.* Blackwell: Oxford and Cambridge, Mass.

Summerhayes, Glenn. 2000. *Lapita interaction.* Terra Australis 15. Canberra: Archaeology and Natural History Publications and the Centre for Archaeological Research, Australian National University.

—— 2001. Lapita in the far west: recent developments. *Archaeology in Oceania* 36:53–63.

Tryon, Darrell and Brian Hackman. 1983. *Solomon Islands languages: an internal classification.* Canberra: Pacific Linguistics.

Wickler, S.H. 2001. *The prehistory of Buka, a stepping stone island in the Northern Solomons.* Terra Australis 14. Canberra: Dept of Archaeology and Natural History, Research School of Pacific and Asian Studies, The Australian National University.